RTO-MP-26

RTO-MP-26
AC/323(IST)TP/2

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 26

Tactical Mobile Communications

(Communications tactiques mobiles)

Papers presented at the RTO Information Systems Technology (IST) Symposium held in Lillehammer, Norway, 14-16 June 1999.

Published November 1999

Distribution and Availability on Back Cover

RTO-MP-26
AC/323(IST)TP/2

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 26

Tactical Mobile Communications

(Communications tactiques mobiles)

Papers presented at the RTO Information Systems Technology (IST) Symposium held in Lillehammer, Norway, 14-16 June 1999.

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 7 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine
- MSG Modelling and Simulation

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

The content of this publication has been reproduced directly from material supplied by RTO or the authors.

Printed on recycled paper

Published November 1999

ISBN 92-837-1022-3

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Tactical Mobile Communications

(RTO MP-26)

Executive Summary

The concept of Ubiquitous Communication and Convergence Technology is a global trend, which will influence tactical military communications and make impact on the military acquisition programs throughout the world. During recent operations in the Gulf War and in Bosnia, there was a need to support the military units with more information than ever before and during an international military operation. The security risk involved in using civilian telecommunications services was continuously traded against the use of slow and reliable military services. In the future, these differences will be smaller. During the cold war military products were developed first and were technologically advanced and expensive. Now the situation is different since military products will need to adapt to civilian standards due to budgetary constraints. Mobile civilian systems will influence the future of tactical military communication systems.

Progress in command and communication systems interconnected with new sensors calls for more and better communications links. There is a need to support higher data rates. This is a problem in terms of available frequencies and possible area of coverage. Higher frequencies are used to permit higher bandwidths to be used. Military systems will use both spread spectrum techniques to achieve robustness to jamming, and a low-density transmitted spectrum to make it difficult to detect.

Higher frequencies in particular, along with the need for higher bit-rates, will limit the maximum range between the communicating parties. Multipath propagation, and the desire to use higher bit-rates, represent engineering challenges.

Military effectiveness requires the ability to acquire and assimilate intelligence in real time and to communicate this effectively on a wide front in the command chain. This is also a very important requirement of operations as demonstrated by the recent conflicts. With increasing NATO responsibilities in joint military operations involving many different national Communications and Information System environments, the need for a unified approach to support information / data transfer services becomes more crucial.

This symposium covers this large but very important operational area from all aspects; system characteristics, internetworking, spectral efficiency, propagation, security, and other relevant factors.

The symposium

The event attracted 150 participants. 19 Partners for Peace nations were represented. The technical program was focused on Personal communications & COTS, Protocols and Networks, Propagation, Speech & Signal processing, and H.F. and comprised oral presentations as well as a poster session. The topics of the poster sessions could quite easily have been included in the following sessions and are therefore not addressed separately.

Personal Communication Systems & COTS

There are several papers, which cover the possible military applications of existing and emerging PCS technologies. The area is very interesting and there is a clear trend in military industry, as well as in civil, for using COTS products in system design.

Protocols and Networks

Protocols and networks are gaining more interest within the communications community. The impact of the Internet, together with continuously increasing needs for high-speed data communication, has made networking one of the key areas in communication technology.

Propagation

Wave propagation is an important subject of never-ending interest. The usage of higher frequencies and the demands for higher data rates call for continuing research on, and measurements of, wave propagation.

Speech & Signal processing

Speech & signal processing is a topical subject, not least regarding the ongoing selection work in NATO on new speech coding algorithms. This was also the focus for the majority of papers in the session, although results on image coding also were presented. In total, three papers were presented. The papers summarise the algorithms of interest and indicated the increased performance available offered by new, improved algorithms.

HF

The H.F. spectrum is of high importance for military applications. In this session, principles and applications for data communication at HF frequencies were presented in four papers. The papers are mainly of tutorial character, and do not focus on scientific research results.

Communications tactiques mobiles

(RTO MP-26)

Synthèse

Le concept de la technologie des communications omniprésentes et de la convergence, qui est universellement accepté, aura une influence appréciable sur les communications tactiques militaires et se répercutera sur les programmes d'approvisionnement militaires dans le monde entier.

Lors des opérations récentes de la guerre du Golfe et en Bosnie, le besoin s'est fait sentir de soutenir les unités militaires en leur fournissant plus de renseignements que jamais, tant avant que pendant les opérations militaires internationales. Les risques au niveau de la sécurité liés à l'utilisation des services de télécommunications ont dû être en permanence contrebalancés par le recours aux services militaires moins rapides mais plus fiables. A l'avenir, ces différences diminueront. A l'époque de la guerre froide, les produits militaires étaient développés en priorité. Il s'agissait de produits technologiquement avancés et chers. Aujourd'hui la situation est différente, car les produits militaires doivent se conformer aux normes civiles à cause des contraintes budgétaires actuelles. Les systèmes mobiles civils auront ainsi un effet sur l'évolution des systèmes de communications militaires tactiques.

Le développement des systèmes de commandement et communications interconnectés à de nouveaux capteurs nécessite d'envisager des liaisons de communications meilleures et plus nombreuses. Il va aussi falloir supporter des débits de plus en plus grands, ce qui pose un problème de disponibilité de fréquences et de couverture possible. Des fréquences plus élevées sont utilisées pour disposer de bandes passantes plus larges. Les systèmes militaires tireront avantage à la fois des techniques d'étalement du spectre pour assurer la protection contre le brouillage, et d'un spectre émis de faible densité pour rendre leur détection plus difficile. En particulier, des fréquences plus élevées, associées à la demande de débits plus grands, auront pour effet de limiter la portée maximale entre deux interlocuteurs. Aussi, la propagation multitrajets et le souhait d'utiliser des débits plus grands, représentent des défis techniques importants.

L'efficacité militaire dépend de la capacité d'acquérir et d'assimiler le renseignement en temps réel et de le communiquer avec efficacité à une grande partie de la chaîne de commandement. Cette capacité est aussi l'une des principales conditions requises pour les opérations, comme en témoignent les conflits récents. Avec l'implication croissante de l'OTAN dans des opérations interarmées mettant en jeu de nombreux environnements nationaux de communications et d'information différents, le besoin d'une approche unifiée pour appuyer les services de transfert de données et de renseignements se fait de plus en plus sentir.

Ce symposium a couvert tous les aspects de ce large et très important domaine opérationnel, en prenant en compte les caractéristiques opérationnelles, l'efficacité spectrale, la propagation, la sécurité, ainsi que d'autres facteurs pertinents.

Le Symposium

Cent cinquante personnes ont participé à cette manifestation. Dix-neuf pays partenaires ont été représentés. Le programme technique a porté sur les communications personnelles et les produits sur étagère (COTS), les protocoles et les réseaux, la propagation, le traitement du signal et de la parole et la HF. Ces sujets ont été traités par le biais de présentations orales et d'une exposition d'affiches. Les sujets de cette exposition auraient très bien pu être présentés lors des sessions présentées ci-après. Par conséquent, nous n'en donnons pas ici une description distincte.

Les systèmes de communications personnelles et les produits COTS

Les applications militaires possibles des technologies des systèmes de communications personnelles existantes et émergentes ont fait l'objet de plusieurs présentations. Ce domaine est d'un grand intérêt, car une tendance nette se dessine tant dans l'industrie militaire que civile, en faveur des produits COTS pour la conception des protocoles, des réseaux et des systèmes.

Les protocoles et les réseaux sollicitent de plus en plus d'intérêt de la part des spécialistes en communications. L'impact de l'Internet, associé à la demande de plus en plus pressante de systèmes de transmission de données à grande vitesse, a fait du travail en réseau l'un des domaines clés des technologies de communication.

Propagation

La propagation des ondes est un sujet important d'intérêt permanent. L'utilisation de fréquences plus élevées et la demande de débits plus grands nécessitent d'engager des travaux de recherche continus et des mesures sur la propagation des ondes.

Le traitement du signal et de la parole

Le traitement du signal et de la parole est un sujet d'actualité, en particulier en ce qui concerne les travaux en cours à l'OTAN sur le choix de nouveaux algorithmes de codage de la parole. La majorité des communications présentées lors de cette session ont traité de ce sujet, même si des résultats de travaux sur le codage de l'image ont également été présentés. En tout, trois communications ont été présentées. Elles ont résumé les algorithmes intéressants et ont souligné l'amélioration des performances autorisée par de nouveaux algorithmes améliorés.

HF

Le spectre HF est d'une très grande importance pour les opérations militaires. Les principes et les applications de la transmission de données à des fréquences HF ont été présentés dans quatre communications. Il s'agissait de présentations de nature plutôt pédagogique, n'ayant pas pour objet d'exposer des résultats de travaux de recherche scientifiques.

Click inside the blue boxes or on the titles to view the corresponding section

Contents

	Page
Executive Summary	iii
Synthèse	v
Theme	x
Thème	xi
Information Systems Technology Panel	xii

	Reference
Technical Evaluation Report by B.-A. Molin	T
SESSION I: PERSONAL COMMUNICATIONS SYSTEMS & COTS I **Chairman: Dr Maj E. TORUN (TU)**	
Emerging Personal Communications for Military Applications by P. Wells and P. Thorlby	1
Applicability of TETRA for Use as a Military Tactical Radio by C.M. Dearlove, J.C. Currie and D.J. Taylor	2
Supporting User & Infrastructure Mobility in the Tactical Environment by R. Barfoot and J. Daniell	3
Tactical Radio Access Networks Based on Future Civilian Wireless Communication Architectures by J. Nilsson, E. Englund, J. Grönkvist, A. Hansson, M. Sköld, U. Sterner, I. Söderquist and J. Zander	4
SESSION II: PROTOCOLS AND NETWORKS I **Chairman: Dr H. WICHANSKY (US)**	
Energy-Efficient Wireless Networking Techniques: Fundamental Issues and an Application to Multicasting in Ad Hoc Networks by J.E. Wieselthier, G.D. Nguyen and A. Ephremides	5
Extending the Littoral Battlespace (ELB): Advanced Concept Technology Demonstration (ACTD) by E.L. Althouse	6
POSTER SESSION	
NATO C3 Agency – PCS Study Plan by N. Young and D. Yavuz	7
Tactical Mobile Communication using Civilian Standards by D. Müller	8

Management System for Mobile Communication Network Used in the Polish Army 9
by K. Mamcarz, M. Małowidzki, R. Matyszkiel and K. Szczygielski

Paper 10 withdrawn

Present and Near Future Tactical Air Defence Operations Requirements 11
by K.N. Nesterov and N.L. Georgiev

Multi-Mode Radios – The Way Forward to Flexible Mobile Communications on the Battlefield 12
by G. Hering

Experimental Studies for a Future Tactical Multi Mode Radio 13†
by O. Künl and E. Kölble

Modeling and Simulation of the ZODIAC Tactical Communication System 14
by B. Boltjes and R.J.G.M. Langeveld

Modeling Terrestrial Mobile Networks in Real Terrain 15
by C. Jönsson, J. Grönkvist, U. Sterner, E. Englund, A. Hansson and I. Söderquist

Performance Analysis of Direct Sequence Spread Spectrum and Non-Spread Spectrum Techniques in Frequency Selective Fading Channels 16
by E. Torun and H.O. Zorba

High Latitude Off-Great Circle Propagation Effects with Reference to HF Communication Systems and Radiolocation 17
by E.M. Warrington, T.B. Jones, N.C. Rogers and A.J. Stocker

UHF MILSATCOM Systems with Emphasis on Changes Made by the Recent Introduction of Automatic Control (AC) Mode Demand Assigned Multiple Access (DAMA) 18
by G. Huckell and J. Parsons

SESSION III: PROTOCOLS AND NETWORKS II
Chairman: Dr H. WICHANSKY (US)

Simulation of a Combat-Net Mobile Radio System with ARQ Feedback Errors 19
by P.J. Vigneron and K.W. Moreland

Testbed for the Evaluation of Battlefield Information Management Techniques Applied to a Low Bandwidth Tactical Wireless Communications Environment 20
by J.-C. St-Jacques and A.W. Gibb

SESSION IV: PROTOCOLS & NETWORKS III
Chairman: Mrs G. GRANT (US)

iSTAR Radio Network for Tactical Use 21
by F. Eken and Ş. Uzun

A Wireless Ad Hoc Multihop Broadbank Network with Quality of Service Support 22
by M. Lott and B. Walke

A Connection Level Priority/Pre-emption Service for ATM Communication Networks 23
by T.D. Johnson, G. Nourry, E.J. Mellaney and H. Rahman

† This paper has not been included because of its classification. Copies may be available to certain members of the NATO nations on request to RTA/IST.

Mobility Management for Tactical ATM Networks 24
by P. Walker

SESSION V: PROPAGATION I
Chairman: Prof Dr T.S. RAPPAPORT (US)

Proposed Concept for a Non-LOS Tactical Wireless LAN 25
by S. Haavik

Spread Spectrum and Mobile/Tactical - Radio/Wireless Communications 26
by D. Yavuz

SESSION VI: PROPAGATION II
Chairman: Prof Dr T.S. RAPPAPORT (US)

Tactical Mobile Radiocommunications - A Challenge for Military Frequency Management 27
by P. Bingel, E. Trautmann and T.J. Taylor

Digital Communication and Multipath Propagation 28
by C. Zehntner, C. Patry and M. Maier

EMC Analysis on VHF Combat Net Radio (CNR) in Airborne Applications 29
by K.S. Kho

SESSION VII: SPEECH & SIGNAL PROCESSING
Chairman: Dr D. YAVUZ (TU)

Noise Pre-Processing for Tactical Secure Voice Communications 30
by J.S. Collura

Enhanced Error Correction of the US Federal Standard MELP Vocoder Employing Residual Redundancy for Harsh Tactical Applications 31
by D.J. Rahikka, T.E. Fuja and T. Fazel

Still Image Transmission in a Link 16 Network Using Embedded Zerotree Wavelet (EZW) Algorithm 32
by M. Martinez-Ruiz and A. Artés-Rodriguez

SESSION VIII: HIGH FREQUENCY
Chairman: Dr P.S. CANNON (UK)

HF Serial-Tone Waveform Design 33
by M.B. Jorgenson and K.W. Moreland

Beyond 9600 bps at HF 34
by M.B. Jorgenson, R. Johnson and K.W. Moreland

The High Latitude Performance and Availability of STANAG 4415 Using Multiple Frequencies 35
by T. Bergsvik, V. Jodalen, P.S. Cannon and P. Arthur

Mobile Internet/Intranet Access 36
by D. Yavuz and P. Vollebregt

Theme

Mobile communications are an important military requirement. Recent operations in the Gulf War and in Bosnia have made this requirement even more obvious. Such communications are naturally also a very important part of civil sector needs from the viewpoint of value added "wireless access solutions" to connect itinerant users to the proliferating fixed, primarily fiber-optic/photonic, networks. As a result many COTS products are available and under development, exemplified particularly by such systems as GSM, LEO/MEO satellite systems, TETRA, TDAB, wireless LANs, LAN bridges, low power SS microwave links, etc. The emerging PSC, UMTS concepts and systems will embody integrated mobile communications in the coming decade.

At the same time there is a need to support higher data rates. This is a problem in terms of available frequencies, transmitted power needed especially in non line of sight conditions in non-flat and vegetated terrain. Higher frequencies are used to permit higher bandwidths to be used. Military systems will use some spread spectrum technique to achieve robustness to jamming and a low density transmitted spectrum to make it difficult to detect. Higher frequencies, in particular along with the need for higher bit rates, will limit the maximum range between the communicating parties. Multipath propagation and the desire to use higher bit rates, represent engineering challenges.

This Symposium will cover the following topics:

- use of present PCS systems for military crisis management operations,
- characteristics and use of emerging PCS, UMTS systems,
- wireless radio, wide area communication,
- mobile networks for land, air and maritime applications,
- wireless LANs,
- terrestrial and satellite network services,
- adaptive modem, antenna techniques.

with emphasis on:

- system characteristics,
- system management,
- internetworking and interoperability,
- spectral efficiency, efficient modulation schemes, propagation & antenna issues,
- security.

Thème

Les télécommunications mobiles sont un besoin militaire important. Les récentes opérations de la guerre du Golfe et en Bosnie n'ont fait que souligner l'utilité de ces moyens. Naturellement, de telles télécommunications répondent en grande partie aux besoins du secteur civil en matière de « solutions d'accès sans fil », représentent une valeur ajoutée pour les connexions et pour les très nombreux réseaux fixes photoniques où à fibres optiques des abonnés itinérants. En conséquence, plusieurs produits sur étagères (COTS) sont disponibles ou en cours de développement, comme par exemple, le système global pour téléphones mobiles (GSM), les systèmes utilisant les satellites à orbite basse terrestre et à orbite terrestre moyenne (LEO/MEO), le système radio terrestre à commutation automatique de canaux (TETRA), la radiodiffusion numérique terrestre (TDAB) les réseaux locaux (LAN) sans fil, les passerelles LAN, les liaisons hyperfréquences à étalement de spectre (SS) de faible puissance etc. Les concepts et systèmes de communications personnelles par satellite (PSC), et systèmes universels de télécommunications mobiles (UMT) naissants seront dotés de télécommunications mobiles intégrées au cours de la prochaine décennie.

En même temps, apparaît un besoin de débits de plus en plus grands. Il existe un problème en ce qui concerne le nombre de fréquences disponibles et la puissance émise demandée, en particulier dans des conditions autres que celles de visibilité directe en terrain accidenté et/ou recouvert de végétation. Des fréquences plus élevées permettent d'utiliser des bandes supérieures. Les systèmes militaires tireront avantage de certaines techniques d'étalement du spectre pour assurer la protection contre le brouillage, avec un spectre émis de faible densité pour rendre ces communications plus difficiles à détecter. Les fréquences plus élevées, associées à la demande de débits plus grands, auront pour effet de limiter la distance maximale entre deux interlocuteurs. La propagation multitrajets et la décision d'utiliser des débits plus grands, représentent des défis techniques à relever.

Ce symposium traitera des sujets suivants :

- utilisation des systèmes PSC actuels adaptés à la gestion d'opérations militaires en temps de crise,
- caractéristiques et utilisation des systèmes PSC,UMTS naissants,
- radio sans fil, télécommunications en réseau étendu,
- réseaux mobiles pour applications terre, air et mer,
- services de réseaux terrestres et par satellite,
- modems adaptatifs, techniques d'antenne.

l'accent étant mis sur :

- les caractéristiques des systèmes,
- la gestion des systèmes,
- l'interconnexion des réseaux et leur interopérabilité,
- le rendement spectral, les schémas de modulation performants, la propagation et les types d'antennes,
- la sécurité.

Information Systems Technology Panel

CHAIRMAN
Dr M VANT
Deputy Director General
Defence Research Establishment Ottawa
Dept of National Defence
3701 Carling Ave
OTTAWA, ON, K1A 0K2, CANADA

DEPUTY CHAIRMAN
Dr R JACQUART
ONERA/CERT/DTIM
Directeur du DTIM
DP 4025
31055 TOULOUSE CEDEX 4, FRANCE

TECHNICAL PROGRAMME COMMITTEE

CO-CHAIRMAN:	**Prof T MASENG**	**NO**
MEMBERS:	**Prof C GOUTELARD**	**FR**
	Ing R de LEO	**IT**
	Maj E TORUN	**TU**
	Dr D YAVUZ	**TU**
	Dr P CANNON	**UK**

PANEL EXECUTIVE

From Europe:

RTA-OTAN
Lt-Col A GOUAY, FAF
IST Executive
7 Rue Ancelle
F-92200 NEUILLY SUR SEINE,
FRANCE

From the USA or CANADA:

RTA-NATO
Attention: IST Executive
PSC 116
APO AE 09777

Telephone: 33-1-5561 2280/82 - Telefax: 33-1-5561 2298/99

HOST NATION LOCAL COORDINATOR

Prof T MASENG
Director of Research
Norwegian Defence Research Establishment (FFI)
PO Box 25
N-2007 KJELLER
Tel: (47) 63 807 204
FAX: (45) 63 807 212

ACKNOWLEDGEMENTS/REMERCIEMENTS

The Panel wishes to express its thanks to the Norwegian members to RTA for the invitation to hold this Symposium in Lillehammer and for the facilities and personnel which made the Symposium possible.

Le Panel tient à remercier les membres du RTB de la Norvège auprès de la RTA de leur invitation à tenir cette réunion à Lillehammer, ainsi que pour les installations et le personnel mis à sa disposition.

Technical Evaluation Report
RTO Symposium on Tactical Mobile Communications
Lillehammer, Norway, June 14-16, 1999

Dr. Bengt-Arne Molin
Lund University
Dept. of Applied Electronics
Box 118, SE-221 00 Lund
Sweden
bam@tde.lth.se

SUMMARY

In this report I briefly summarize the activities at the RTO symposium on Tactical Mobile Communications in Lillehammer, Norway, June 14-16, 1999. Furthermore, I will discuss my view of the event and present some possible ways for improvements.

The technical program at the symposium was focused on Personal communications & COTS, Protocols and Networks, Propagation, Speech & Signal processing, and HF and comprised oral presentations as well as a poster session. My general impression was that the conference was well organized and of high quality. The oral presentations were in general good but the poster session could have been better. I also believe that the symposium would gain on having a more clear focus, for example through a somewhat more narrow scope of the research areas presented.

1 INTRODUCTION

NATO's Research and Technology Organization (RTO) is the single focus in NATO for Defense Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defense research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision-makers. RTO's scientific and technological work is carried out by Technical Teams, operating for six panels covering different areas of research and technology. The symposium on Tactical Mobile Communication in Lillehammer was organized by the Information and Systems Technology Panel (IST).

Mobile communication is an important military requirement, not least shown by recent operations in the Gulf War and former Yugoslavia. There are many Commercial Off The Shelf (COTS) products available and under development, that might be suitable for military applications. The emerging 3rd-generation mobile communication systems may also offer services for tactical and strategic systems. Furthermore, the need to support higher data rates will create engineering challenges within the military industry as well as within the civil.

The symposium covered the following topics:

- use of present PCS systems for military, crises management operations,
- characteristics and use of emerging PCS systems,
- wireless radio, wide-area communication,
- mobile networks for land, air, and maritime applications,
- wireless Local Area Networks (LANs),
- terrestrial and satellite network services, and
- adaptive modem and antenna techniques.

This report will summarize my findings from attending the conference. I have briefly read the papers sent to me in advance (some of them in detail). Furthermore, I have attended all sessions at the conference. I have also talked with several attendees during the event, regarding their views on the symposium. Finally, I have studied the evaluation forms filled in by the attendees.

I start with presenting an overview of the technical program, followed by some comments on the organization of the event. Then I summarize the views put forward in the evaluation forms. Finally, my conclusions from evaluating the symposium will be given, together with some recommendations for future meetings.

2 TECHNICAL PROGRAM

The technical program consisted of five oral sessions and one poster session, all described in more detail below. The papers could mainly be categorized as

- tutorials,
- scientific research results, and
- reports from system design activities and field trials.

Although the topic for the conference was Tactical Mobile Communications, many of the presented papers were more generic and could have been presented at any communications conference. This was probably due to the dominance of papers presenting research results.

In general, the papers were of high quality and clearly comparable to an international scientific conference. However, the quality of the papers varied a lot and a few of them, I would judge, were not suitable for publication. Furthermore, from the content of the papers it was in many cases hard to identify the overall theme of the conference.

Personal communication systems & COTS

In this session four papers were presented, covering possible military applications of existing and emerging PCS technologies. The area is very interesting and there is a clear trend in military industry, as well as in civil, for using COTS products in system design.

The presentations gave a good overview of the governing ideas for the various projects started. However, it was hard to extract any substantial information regarding the suitability of the presented technologies; this might be due to the early stage the projects seemed to be in.

Protocols and Networks

Protocols and networks are gaining more interest within the communications community. The impact of the Internet, together with continuously increasing needs for high-speed data communication, has made networking one of the key areas in communication technology. The session comprised eight papers, presenting system descriptions as well as scientific results.

The session was very heterogeneous and the papers had hardly any connections with each other. Still, some of the papers were of very high quality scientifically and the presentations on system design were also of high standards.

Propagation

Wave propagation is an important subject of never-ending interest. The usage of higher frequencies and the demands for higher data rates call for continuing research on and measurements of wave propagation. In this session five papers were presented, covering various aspects of the topic.

The papers gave a good overview of problems in the area. However, no concrete results could be extracted from the presentations.

Speech & Signal processing

Speech & signal processing is a topical subject, not least regarding the ongoing selection work in NATO on new speech coding algorithms. This was also the focus for the majority of papers in the session, although results on image coding also were presented. In total, three papers were presented.

The presentations summarized the algorithms of interest and indicated the increased performance available offered by new, improved algorithms.

HF

The HF spectrum is of high importance for military applications. In this session, principles and applications for data communication at HF frequencies were presented in four papers.

The presentations were mainly of tutorial character, and did not focus on scientific research results.

Poster Session

The poster session comprised twelve papers, the topics of which covered a wide range of research areas and applications. The poster session was organized as two different events where the authors were available for comments and discussions.

In general, the quality of the posters was low. Although some of them were good the majority was A4 pages pinned on a wall, rather than real posters. Furthermore, it was not easy to get contact with the authors at the announced times.

3 ORGANIZATION

The symposium was held at Quality Hotel, Hafjell, just outside Lillehammer, where all activities took place. Information about the event was available in advance on the symposium web site. The presentations were given in plenary sessions with breaks for the poster sessions on Monday and Tuesday.

As a whole, the organization was of high quality. The facilities were well suited for this type of event and the practical arrangements were excellent. Although about 150 people were gathered at the symposium, the meeting rooms were never crowded. Also, there were plenty of possibilities for the attendees to get acquainted with each other, which I judge was one of the primary aims of the event.

4 EVALUATION FORMS

In total 66 evaluation forms were handed in, corresponding to approximately half of the attendees. The form prompted mainly for the following information:

- affiliation of the attendee,
- overall value of the event,
- the impact of the event with respect to some specific areas, and
- recommendations for future events.

There was an overwhelming representation of governmental employees. There were also some participants from industry, while only a few came from academia. The overall value of the event was judged in a scale from 1 to 6, representing "Not worth the effort" as 1 and "Extremely valuable" as 6. The average judgement was 4.1 with a standard deviation of 1.1.

Regarding the impact of the event, I followed up the judgements "Substantial" and "Highly significant" in more detail. The greatest impact of the event was, not surprisingly, described as "Expands personal contacts for potential future collaboration". However, two other areas were clearly of substantial and significant importance, namely "Provided valuable new perspective in your work" and "Provided new insights and considerations for organizational long term planning". The recommendations for future events, finally, didn't provide any coherent guidance.

5 DISCUSSION

My general impression of the symposium is very positive. I found the program both interesting and relevant. Also, the high quality of the conference facilities and the beautiful surroundings made the trip to Lillehammer well worth while.

There are some things that I believe can be improved at future events, of which the main issue is a more clear focus. Although targeted for Tactical Mobile Communications, I have a feeling that the symposium tried to cover too much. As far as I understand, the audience consisted mainly of three categories, namely

- military representatives (i.e., users),
- defense industry, and
- researchers.

Clearly, people from the different categories are interested in different aspects of tactical mobile communications. I have a feeling that the scientific presentations were too technical for the military representatives, while generally not technical enough for the researchers. Similarly, reports about system design activities were to some extent not concrete enough for satisfying the needs from users and researchers.

The solution would be to have a more clear focus for the symposium. This could be realized through a more narrow scope for the areas presented. Another possibility would be to select papers and sessions more thoroughly, thus creating a balanced mapping of the paper categories on the audience categories.

Regarding the selection of topics for the symposium, I believe that tactical mobile communications was a bit too restrictive. Mobile communications in the future will be part of a distributed information structure. Thus, it is necessary to look at information management as a whole, and not just the mobile communication part. For military applications, this implies that tactical communications will be coordinated with strategic information management, thus including for example information centers and communication backbones. Although this might seem contradictory to my request for focus, I believe that the focus should be stronger scientifically, while application areas could be covered more broadly.

Finally, I will give my comments on the concrete activities at the symposium, thus putting forward some possible ways for improvements.

Oral presentations

In general, the quality of the oral presentations was satisfactory. However, there was no clear line on how the papers were related to each other in a session. I would suggest that the sessions be more thoroughly planned. I also call for a more active roll of the session chairs. I would have appreciated a summary of each session topic (and why not the papers) given by the session chair, rather than the mandatory biographies.

Poster sessions

In principle, I think poster sessions are a good way for allowing people to get detailed information about a paper through talking with the authors. However, at the symposium I had a feeling that the poster session did not succeed very well. I have three suggestions for making the poster session more successful at the next RTO event (I do believe that you should keep on offering poster sessions):

- Be more precise on how a poster should be prepared, regarding available space, font sizes, etc.
- Let the poster session start with two minutes oral presentations of the authors and their work, so that the audience is aware of what is being presented.
- Don't mix the poster session with other activities and be sure to make the session a *real* event.

Organization

As a whole, the organization was very good. The use of the web is very efficient and I strongly recommend that the papers be published at the web site.

Regarding future improvement, I only have a minor comment: I would have appreciated a final program giving an overview of the activities at the symposium (as a complement to the detailed final program that was distributed at the conference).

Emerging Personal Communications for Military Applications

Paul Wells[1], Dr Paul Thorlby[2]

1 Introduction

The technology revolution in mobile communications combined with the widening of military roles during the 1990's continues to present new opportunities for civil systems to be used by the military.

The new global political context requires a military capability that can function in concert with other nations and respond rapidly to a wide variety of situations (ranging from humanitarian aid, through peace keeping and peace enforcement to coalition intervention). Many of these operations may occur 'out of area' (OOA).

The response to this situation is exemplified in the creation of the UK Joint Rapid Reaction Force (JRRF), the ACE Rapid Reaction Corps (ARRC), and the NATO Combined Joint Task Force (CJTF) concept. These forces require flexible, deployable, scalable communications and information systems (CIS) often to areas where the indigenous infrastructure cannot be assumed.

The significant new 'mobile communications' market sector is beginning to challenge the dominance of the fixed telecommunications infrastructure for the provision of services to the end user. This technology revolution is generally known as Personal Communications Services (PCS) and includes a number of system standards; terrestrial and satellite based.

Most notably in the provision of public service is GSM (65% global digital market). The emerging TETRA system (digital trunked mobile radio) is of particular interest to law enforcement and paramilitary users since it offers features specifically required by that user community. The emerging satellite PCS systems (e.g. Iridium, Globalstar and ICO) are also of interest due to the global geographical coverage they provide with a relatively sparse ground infrastructure. Future terrestrial PCS, so called '3rd Generation' (or 3G) systems (e.g. UMTS) and satellite (S-UMTS) are promising high data rates (up to 2Mbit/s) and advanced service concepts.

This paper presents an overview of the major system standards (current and emerging) and discusses the potential for military use of each. Regional standards (such as AMPS, ACeS or Thuraya) have not been considered.

PCS systems are the subject of ongoing study and test-bed activity by DERA (UK) and NC3A-NL. The latter is the subject of a separate paper at this conference.

2 Operational Roles

These PCS systems aim to offer global coverage whilst providing voice and data connectivity. This allows an ideal opportunity for NATO to provide relatively low cost global communications, achieve interoperability and in some cases provide communications from mobile/tactical deployments that have not been possible in the past. The opportunities for position reporting and alerting allow for better situational awareness.

It must be stressed that these civil systems have been developed for maximum throughput to earn revenue for their owners. This may be at odds with the military's need for assured levels of communications, non-reliance on fixed infrastructure, operation in stressed environments and the need to minimise the visibility of some operations.

The roles in which M-PCS may play a role include:

Strategic Rear Link. The initial liaison troops will require infrastructure free (i.e. not transported in by the troops) communications back to the permanent headquarters. This could use satellite systems or existing commercial terrestrial systems (e.g. GSM).

[1] Communications Department, DERA Defford, Worcester, WR8 9DU, UK. DERA is an Agency of the UK MoD
[2] Communications Systems Division, NATO C3 Agency, The Hague, The Netherlands. (Seconded from DERA UK)

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

Initial Mobile Communications. Troops deployed in theatre will require basic communications that can be rapidly established. This may include 'all informed' Combat Net Radio (CNR) communications that provide a simplex broadcast voice service to all network subscribers. CNR doesn't require infrastructure. The key advantage of wireless communications is that wide area coverage can be rapidly provided with relatively simple infrastructure requirements, to cover for example a port of entry

Mobile Sub-system. As the deployed force builds up there is a requirement for the commanders to be able to establish point to point duplex communications between themselves, the fixed HQ and (depending on the mission type) other organisations (such as embassies, non-governmental organisations, host nation civil administration and government bodies).

3 GSM

3.1 System Description

GSM is a terrestrial cellular PCS operated by over 240 network operators in over 118 countries. The number of subscribers to GSM systems worldwide is growing exponentially and passed the 150 Million mark during early 1999 [1]. It offers duplex voice, short messaging, fax, data and an extensive range of supplementary services between the personal handset and other mobile users, PSTN (Public Switched Telephone Network) users and ISDN (Integrated Services Digital Network) users. Recently standardised extensions to the system include 'all-informed' group calls (half duplex, press to talk), and enhanced multi-level priority and pre-emption to support the 'professional' market. At least one manufacturer is now offering equipment with these features (including improved specification handsets). The GSM system has an advanced (in the civil context) security architecture, including user authentication and air interface encryption.

The GSM network [2] comprises a number of elements – the mobile or personal subscriber terminals (MS); a cellular base station transceiver (BTS) and controller (BSC); mobile switching centre (MSC) and interfaces (interworking functions) to other systems (such as the PSTN). Key internal system interfaces are standardised, the most important being that between the MSC and BSC (the 'A' interface) and the mobile station and the BTS ('Um' interface). This allows the choice of different suppliers for different sub-systems.

A number of fixed base station transceivers provide radio coverage over a number of cells to allow direct radio links to the mobile terminal. These links are currently standardised in the 900, 1800 or 1900 MHz frequency bands and the air interface between the mobiles and base use both frequency division and time division multiple access (FDMA/TDMA). Each frequency channel occupies 200kHz of spectrum, is modulated at 270 kbps using Gaussian Minimum Shift Keying (GMSK) and supports up to eight time division full rate (approximately 13kbit/s protected) traffic channels. (Some channels may be allocated for signalling). Data is currently supported in a full rate channel at rates of 2.4, 4.8, 9.6 (and in some cases 14.4) kbps. Each cell has a maximum range of 35km which is determined by system timing constraints (this can be doubled to 70km at the expense of traffic capacity). Traditional cellular engineering is used to provide wide area coverage. Advanced cellular concepts (such as hierarchical cell structures) are used to achieve very high traffic density. The GSM system allows seamless handover between cells (including hierarchical cells) and frequency bands.

The 'heart' of a GSM system is the Mobile Switching Centre (MSC). This entity not only switches traffic calls, but also manages the supplementary services and is the host for the interfaces and interworking functions to other types of network. Attached to the MSC are the key databases; the home location register (HLR) and visitor location register (VLR). These databases hold all subscriber related information including current (or last known) location. The only exception to this is the subscribers secret key, which is the cornerstone of the authentication and encryption systems. This information is held in the authentication centre (AuC), normally a separate (secure) facility associated with the MSC. The GSM sub-system components would normally be connected using 2Mbit/s E1 links, using either leased PTT lines or microwave line of sight links.

The air interface encryption operates between the BTS and the MS only and is under the control of the network. Currently 3 algorithms are possible; no encryption (A5/0), A5/1 and A5/2. The latter was developed to overcome export restrictions which apply to A5/1. The system can support more (up to 7) algorithms, but the others are not defined at this time. The standard was recently modified to forbid A5/0 (no encryption), but to what extent this has been implemented is not known. User authentication is done by a network specific algorithm (since it needs only be implemented on the network operator provided SIM (subscriber identity module) and in the network operator owned AuC). In practice most operators choose the 'example' algorithm available to bone fide operators from the GSM Association. The authentication process is standardised in that the sequence and size of signalling messages is determined. The suitability of the A5 algorithms for NATO operations is subject to national COMSEC authority recommendations, and would be unlikely to be approved for more than (possibly sensitive) UNCLASSIFIED information. Given this situation however, the privacy provided is significantly enhanced over analogue radio systems. All communication links within the network are in the clear.

GSM continues to evolve in terms of new technical features. The most significant of which include high rate data (EDGE and HSCSD) potentially up to 384kbit/s in favourable radio link conditions and a packet data system (GPRS) (projected to be commercially launched late 1999) designed to support efficient internet access for the MS.

3.2 Military Utility of GSM

GSM offers the military a low cost handheld, land mobile or close to shore maritime mobile toll quality voice, short messaging, data and fax services. GSM provides a widely accepted and used standard that is readily available for NATO interoperability.

Wide land based geographic coverage is provided subject to limited availability in remote areas (coverage is focussed on high population density areas). GSM requires fixed infrastructure and there is very significant existing civil investment in this already. Military owned tactical transportable base transceivers and connection to a tactical MSC is possible. If this is planned where an existing civil infrastructure exists, frequency planning and co-ordination is required with the existing operators and this may impose unacceptable planning constraints.

A national COMSEC authority approved encryption mechanism is required which is compatible with GSM for classified information exchange. GSM provides a potentially strong privacy mechanism provided this is enabled and proper protection is made of the infrastructure links.

If military users take advantage of the civil infrastructure there are some potential drawbacks; monthly bills, no guaranteed access to the system (priority in the event that the system is busy), susceptible to interception (for example by the national authority). The system may also be used to geo-locate a user.

In the case of using a civil system, or a deployed NATO owned system, GSM is susceptible to denial of service, through for example being jammed.

GSM is not readily suitable for use on airborne platforms due to its cellular architecture (a mobile station may 'see' and hence interfere with a very large number of cells from an airborne position) and Doppler limits (eg. on fast moving aircraft - the system was designed for high speed trains, up to 250km/h).

4 TETRA

4.1 System Description

TETRA is an emerging ETSI standard for a digital trunked radio system, targeted at both the PMR (private/professional mobile radio) and PAMR (public access mobile radio) markets. Like GSM it provides wide area coverage using cellular technology.

The system is based around a radio switch and associated databases which form the heart of the SWMI (switch and management infrastructure). The radio interface uses 4 TDMA traffic channels in a 25kHz FDMA radio channel. The traffic channels can support speech and data. Various modes of operation enable simultaneous speech and data, full duplex speech, half duplex (press to talk) speech, group and broadcast calls. The data modes include circuit (from 2.4-7.2 kbps in a single timeslot with up to 28.8 kbps with four concatenated timeslots) and a packet mode. The system also supports short messages.

The TETRA radio interface has been specified to work in several frequency bands; including around 400MHz and 900MHz. Manufacturers have built systems in frequency bands where they have customers, which (in Europe) is principally the 380-400MHz band (portions of which have been released by NATO for emergency services use) and the frequency band allocated to Dolphin (which has PAMR licenses in many European countries).

TETRA has some system features which differentiate it from similar systems (e.g. GSM). These include fast call set up (less than 300ms), direct mode (mobile to mobile without infrastructure) and dual watch (operation in direct mode with the ability to receive calls from the normal infrastructure supported communications).

TETRA also has features which are judged essential for PMR markets; priority, pre-emption and dispatcher control of communications and user groups. The equipment is also manufactured (in particular the mobile stations) to withstand a harsher physical environment than the typical consumer GSM phone. This makes the terminals larger and heavier than the more recent mobile telephones. Small 'cellphone like' TETRA terminals are planned by some manufacturers to address the PAMR market.

TETRA has an advanced security architecture facilitating mutual authentication and air interface encryption. The system specification also provides a framework for a manufacturer to implement a proprietary end to end encryption system. The latter requires further specification work to guarantee interoperability. The system appears to allow over the air re-keying (OTAR) through different forms of implementation although this is not yet fully specified.

The internal structure and interfaces within the SWMI are beyond the scope of the standards. The air interface and external system interfaces are standardised.

The TETRA standard is almost complete. Trial systems have been in operation at various sites in the world for over a year, and the first truly operational systems are expected soon. It is noteworthy that as GSM has attempted to incorporate PMR features (embodied in the advanced call speech items-ASCI specifications) to address the PMR market (e.g. the recently launched Ericsson Pro GSM products), so TETRA is adopting cellular telephony style handsets to address the public access mobile market.

4.2 Military utility of TETRA

TETRA has obvious attractions to the military user; supporting from the outset (rather than as a late addition) essential features such as priority, pre-emption and fast call set-up. The 'ruggedised' handsets also appear more attractive to a military customer than regular consumer grade cell-phones. TETRA also offers a ready opportunity to integrate military grade end to end encryption systems to provide COMSEC.

The utility of direct mode, which it could be argued emulates some of the military combat net radio (CNR) functionality (squad/section radio), depends on the operational role TETRA is being considered for. It could also be argued that this provides a backup mode of communication in the absence of infrastructure (or in the event of infrastructure failure). Repeaters and gateways can be used to enhance the capability.

TETRA is susceptible to EMC in the same way as other commercial systems. It has no electromagnetic protection mechanisms (EPM) built in.

The smaller PMR market compared to the enormous consumer mobile telephony market tends to encourage manufacturers to consider customising a TETRA system in a way that would not be cost effective for (e.g.) GSM.

There currently is no global TETRA infrastructure to take advantage of. There are some small private networks for which access might be negotiated in the event that the coverage coincided with an operational requirement, however this seems unlikely (except possibly for trials). There will be 'public' TETRA networks in various part of Europe and elsewhere where a service level agreement could be negotiated. The military user community then shares the infrastructure with other user communities. The segregation of users (and information about users) is in the hands of the network operator.

TETRA could be procured and deployed by the military to provide mobile access. The size and scale of the infrastructure equipment supports this. PMR systems may be sold to relatively small user communities and so the equipment is typically designed to scale from a small base. This is not true for systems like GSM where systems are typically sold (and scaled) for nation-wide coverage, making the infrastructure large and heavy. Deployable versions of GSM infrastructure do however exist.

5 TETRAPOL

5.1 System Description

TETRAPOL has many similarities to TETRA. It specifically addresses the security market, law enforcement and emergency services and the PMR (rather than PAMR) markets.

Technically it has some fundamental differences to TETRA; it is based on a single channel FDMA (in a 10 or 12.5 kHz channel allocation) and end to end (rather than air interface) encryption.

The specification was developed by Matra-Nortel Communications for the French Authorities, but is now offered as a publicly available specification (PAS) by the TETRAPOL Forum and several manufacturers can supply the equipment. The TETRAPOL Forum recently failed to have the PAS endorsed as an ETSI standard, primarily because the ETSI membership considered it too close (and therefore in competition with) the TETRA standard. So whilst it remains an open standard, it is not an ETSI standard.

5.2 Military Utility of TETRAPOL

Much of the above discussion concerning TETRA applies equally to TETRAPOL. The TETRAPOL proponents cite advantages over TETRA for wide area coverage where there is low traffic density (due to the FDMA radio access technology), cheaper infrastructure (less of it required due to the inherently simpler system design) and lower protection measures required for the base stations (since there is no encryption technology placed here).

TETRAPOL networks have been operational in France for several years (the National Police and railway system) and more recently have been deployed elsewhere in Europe. There are no public access TETRAPOL networks planned in the authors' knowledge.

6 Iridium

6.1 System Description

Iridium [3] is based on a low earth orbit satellite constellation of 66 satellites and provides global low rate voice communications between handheld and fixed PSTN users as well as paging. The system has recently become operational and currently there are 74 active satellites in orbit (8 in-orbit spares). Each satellite provides a 48 beam cell coverage of the earth with each cell of nominal 700 km diameter. The satellites beam pattern is fixed and users are handed from beam to beam within the satellite and then between satellites as the satellites fly over.

The satellite constellation comprises 6 orbits of 11 satellites and the network connectivity is base on on-board switching of user circuits using inter-satellite links (four 23GHz links per satellite). User connection is achieved at L-band (1621.35 – 1626.5 MHz) in the spot beams between the handheld and the satellite. The FDMA/TDMA user link (4 uplink and downlink channels per TDM frame) is switched on board the satellite and routed via the inter-satellite links to either another handheld or via a gateway earth station (at Ka-band) for connection to the PSTN.

Full global coverage is provided, including the poles and oceans, and the system supports a variety of land, maritime and airborne platforms. Handsets and mobile terminals are currently manufactured by Motorola and Kyocera with airborne terminals by Allied Signal. The US DoD has its own gateway in Hawaii and is planning secure handsets later this year. A low rate data service was not available at the time of writing but initial testing is underway for a 2.4kbps async service.

Dual mode cellular /satellite handsets are available e.g. Iridium,/GSM, Iridium/AMPS etc. The system routinely carried out position location of the handsets to allow for accurate billing and prevent operation from countries not agreeing to service.

6.2 Military Utility of Iridium

Iridium offers a moderate cost solution for near global low rate handheld and mobile voice communication.

DERA has been actively involved in both Beta and full service evaluation of the voice system and is continuing to evaluate its performance and use for UK MoD. A clear advantage is the ability to offer an interoperable communications bearer for wide geographic displacements as has been demonstrated in recent TTCP (The Technical Co-operation Programme) interoperability trials between Australia, Canada, UK and US.

The voice quality offered is near toll quality and user voice recognition is not guaranteed but is good for a low rate voice coder. Operating from land, maritime and airborne mobile platforms has been demonstrated including use of the TDMA air interface through rotating helicopter blades. The paging or alerting service has proved very useful, but is limited by the need to specify three 'message delivery areas' (of country/state size) rather than global page.

Operation in buildings and built up areas is limited for handhelds with a reasonable line of sight radio path to the satellite being required. The pager service (with a higher 35 dB margin) has proved more effective within buildings and built up areas. Use under foliage is possible but use under dense, wet foliage has proved difficult.

Reliance on the Iridium satellite and fixed infrastructure is required and assured levels of access are not currently available. A higher precedence level is available for certain users (emergency service, airborne etc) and the system is not fully available in all countries. The system user links could easily be denied (eg. via jamming), is susceptible to interception and the user terminals location is known within the system.

A shared NATO Gateway (possibly even a fall back facility for the US Gateway) would be possible allowing direct termination of NATO users in a common terrestrial switching centres – with the possibility of lower costs by virtue of bulk user negotiations. A transportable gateway may be an option.

The use of separate encryption is essential and co-ordination of this across NATO would be required to allow interoperability. The US DoD are funding development of encryption for Iridium.

7 Globalstar

7.1 System Description

Globalstar [4][5] is a low earth orbit mobile satellite service that is currently being deployed with 20 satellites launched (May 99) out of the 48 constellation required. Initial service is planned for October 99. The system is based on user terminals (handheld, mobile and fixed phone booth) communicating via one or more satellites to a regional gateway where routing into PSTN is achieved. The satellites provide L-Band (1610 – 1621.35 MHz uplink and S-Band (2483.5 – 2500 MHz) down links via 19 beams per satellite. FDMA/CDMA is used as the air access with the combination of two or more signals via different satellites in the regional gateway. Transparent transponders are use on the satellite (no switching) with the fixed beams moving over terrestrial users and the signal switching being handled in the gateway. No inter-satellite links are used.

Globalstar is only offering service between 70°N and S and although ocean coverage is available this is not seen as their prime market and coverage to the nearest land base gateways is not always available.

Dual mode Globalstar/GSM handsets will be available. Globalstar have offered a transportable gateway facility for military use. Although no Globalstar secure handset is available the US is investigating an IS95 (regional cellular) version that may be modified to provide Globalstar capability. Use of Civil encryption similar to GSM will be used allowing a certain degree of privacy.

7.2 Military Use of Globalstar

Globalstar will offer low cost handheld and land mobile voice and short messaging service. Use as a common standard for NATO interoperability is possible but this will probably be limited to regional coverage. Operation from land-mobile and possibly helicopter platforms should be feasible. The system will probably have limited use in buildings, built up areas and wet dense foliage due to the requirement for a reasonably clear radio line of sight path to the satellite (see Iridium) and propagation at L and S bands. However this will need verifying on the operational system when available.

Reliance on the Globalstar satellite infrastructure and regional gateways is required and assured levels of access are not available.

The system user links could be easily denied (eg. by jamming), calls intercepted and the user terminals position will be known in the regional gateway and the Globalstar system.

NATO approved encryption will be required. There is potential for a NATO transportable gateway to allow regional use with direct connection to deployed NATO communications facility.

8 ICO

8.1 System Description

ICO [6][7] aim to provide a mobile satellite service based on a 10 satellite (2 planes of 5 satellites) medium earth orbit (~10,300 km) constellation. An initial capability is planned for early 2000 with full service in September 2000. Each satellite provides a fixed 163 beam/cell user coverage at S-band (using the UMTS mobile satellite service bands of 1980-2010 MHz uplink and 2170-2200 MHz downlink). A user terminal communicates via the nearest satellite (FDMA/TDMA) to a fixed Satellite Access Node (SAN's). The connection is routed via a 'terrestrial' backbone connecting the twelve SAN's to an appropriate connection into the PSTN. The cells formed by the multiple beams move over the user as the satellite orbit progresses and user handover between beams is achieved in the SAN. Each SAN is capable of tracking 4 satellites so is able to use diversity on the uplink signal to choose the strongest signal from a user terminal at any one time.

The service plans to offer handheld and land, air and maritime mobile services with near global coverage but with the satellite capacity being adaptively adjusted over the oceans and poles to meet the high demand from higher population density areas. ICO aim to use a privacy key similar to that used in GSM. The government sector is clearly defined by ICO as one of their target 'vertical' market segments.

Dual mode satellite/cellular handsets will be offered. A low rate data service will be introduced. The US DoD are believed to be funding the development of encryption to be used over ICO.

8.2 Military use of ICO

ICO offers the military a low/medium cost handheld and mobile voice and data capability. Although not yet evaluated the voice coder proposed (4.8 kbps) would offer toll quality speech and good potential for user recognition. Wide geographic coverage appears possible but the day to day allocation of satellite power may leave coverage holes – this will require further investigation for military use. ICO offers a standard for allowing NATO interoperability.

Reliance must be made on the ICO satellite and fixed infrastructure and assured levels of access would not be available.

The system user links could be easily denied (eg. jamming), calls intercepted and the position of the user terminals will be known within the ICO network.

Use within buildings, built up areas and under dense wet foliage would be limited (see Iridium) but use on tactical mobile platforms would prove valuable.

Civil encryption (commercial privacy) is provided but NATO level encryption would be needed – requiring co-ordination across NATO for interoperability. A dedicated SAN may be feasible.

9 ORBCOMM

9.1 System Description

ORBCOMM [8] has just entered service providing global store and forward data messaging capability. 24 out of the full 32 satellites have been launched and the service is operational.

The VHF 137-138/148-149.9 MHz based system allows a user terminal to send and receive short data message (128 chars) to the nearest satellite when in range. The messages are stored on-board until the satellite overflies the appropriate downlink gateway whereby the information is fed via PSTN connectivity typically into Internet sites/servers for onward delivery to the recipients e-mail address.

The majority of the user terminal systems provide a GPS reporting capability which is integrated with the messaging system to provide remote position reports.

9.2 Military use of ORBCOMM

The use of ORBCOMM for position reporting of military equipment/stores is possible with global coverage being available. There is little link margin available and a well sited VHF antenna is required for successful use – particularly for mobile platforms. Operation inside buildings and built up areas is severely limited.

This is a good opportunity for a tactical transportable NATO gateway for non-critical position reporting and situational awareness. However, the service is not under military control and assured access is not under military control and assured access is not provided. Encryption of data will be required and an acceptable file encryption system adopted for widespread use and protection of the data is needed. The system user links could easily be denied.

10 UMTS and S-UMTS

10.1 System Description

The UMTS [9] (Universal Mobile Telecommunications System) is an emerging standard which aims to provide the 3rd generation (3G) mobile telecommunication system. UMTS was originally being developed within ETSI (the European Telecommunications Standards Institute) but is now the subject of (an ETSI initiated) '3rd Generation Partnership Project' (3GPP). This is a collaboration of regional standards bodies which includes the major global markets (Europe, North America, the Far East) which is striving for a common standard, based on the original ETSI (GSM core network and CDMA radio access based) UMTS work. The regional UMTS proposals will also be submitted to the ITU as candidate members of its 'family' of IMT2000 standards.

Spectrum has been identified and allocated for 3G systems ('Cordless Home and in buildings use': 1900-1920 and 2010-2025 MHz, 'Terrestrial Cellular use': 1920-1980 and 2110-2170 MHz, 'Satellite use' 1980-2010 and 2170-2200 MHz). Some countries have already started issuing licenses for UMTS and other countries (e.g. the UK) are in the advanced stages of industry consultation over the license issuing process. The target initial service offering date is 2002 with wide-scale roll out by 2005. There is a satellite component planned called S-UMTS which is expected to be the next generation of mobile satellite systems.

The key features of UMTS are high data rates (terrestrial handheld services up to 2 Mbps and 384kbit/s for wide area mobile users) supporting internet services and electronic commerce. Advanced service concepts (such as the 'virtual home environment') and new network architectures are being considered. Security and government usage will be key market opportunities for valued added services to UMTS operators.

10.2 Military use of UMTS and S-UMTS

Clearly UMTS has the potential to offer great value to NATO for military use for handheld and tactical mobile deployments. The NATO co-ordinated customer base may have a sufficiently large niche market to be able to influence the standards or offer added value in terms of security and system protection to the commercial markets being focussed upon. Whilst the standards are in development the opportunity exists for NATO to support their formulation to enhance the subsequent systems performance. However it must be stressed that the military and governments are only a small market sector and the added value of may enhancements must provide improved service for the larger commercial users.

Use of the new UMTS frequency bands and higher data rates will cause some limited use in buildings and built up areas where no fixed infrastructure exists. The potential combination of UMTS and S-UMTS in a single handset or mobile will offer the military a go anywhere capability.

Compatibility with NATO accredited INFOSEC systems would significantly enhance the utility of this system to NATO commanders.

11 Conclusions

The emerging systems offer small, hand-portable or mobile equipment providing wide coverage voice and data communications. The systems offer the military a number of communications services allowing interoperability by use of a particular commercial standard and offer wide geographical coverage. In addition, the drive by commercial markets for improved performance and added selling advantage forces technology refresh at rates way in excess of classic military communications procurement.

The systems offer the potential for global reach voice and data for reporting roles into and from tactical locations. They also offer communications from mobile platforms and can be used in several roles where military communications may not be appropriate or affordable.

Being driven by commercial markets by necessity keeps the cost of use and ownership low when compared to military systems. The military user could also consider other models than outright procurement, for example leasing equipment.

Use of the systems for diverse applications from situation awareness, reach-back for global broadcast services, engineering links for military systems, easy interface to internet and email, alerting and many others in addition to the typical roles of reporting and command and control.

However these systems are designed for commercial use and operation in stressed environments would only be possible using a "use it until you lose it" approach. Certainly peace time operations and even use in low level conflicts may be possible but how the service providers would react to military use of their systems is unclear!

The level of assured access and control of the systems by third parties are of concern as are the ease of denial of use and limits on capacity in regions with some of the systems. The equipment is not ruggedised to meet the harshness of some military environments but should survive general military use.

Use of external or embedded encryption is required with external encryption being preferred to allow technology refresh. However this would require two packages to be carried with the resulting physical vulnerability of connecting cables in rugged environments. Encryption for both voice and data is required as well as a method for key distribution.

UMTS (with S-UMTS) offer the potential for a common handheld providing both short range cellular and long range global satellite access. Higher levels of priority use for NATO operations is desirable with the added advantage that if NATO is using the system there is less chance of the denial occurring! There is potential for a multi-system mobile (or even handheld) with a common antenna for use on mobile platforms. Using intelligent networking overlaid on this multi-system mobile/handset would allow adaptive use of the systems for lowest cost, best connection or best survivability.

There could be potential for a dedicated NATO Gateway to some of the satellite based systems – possibly with enhanced service capability.

There are good opportunities for NATO use of the emerging systems to provide added value and enhanced capability. This should clearly be explored with a series of pan-nation interoperability trials/demonstrations to show the benefits and identify the optimum systems for use.

12 Acknowledgements

The opinions expressed in this paper are those of the authors and do not necessarily reflect official UK MoD policy, nor the official position of NC3A. The authors would like to thank UK MoD for the support of the analysis of civil systems being carried out in DERA and their many colleagues who have provided useful comment on this paper.

13 References

[1] "GSM the advanced and proven basis for success in mobile markets". Friedhelm Hillebrand, ETSI Tdoc SMG P-99-007, Feb 99
[2] "GSM Cellular Radio Telephony" Joachim Tisal, Wiley
[3] Iridium http://www.iridium.com
[4] Globalstar http://www.globalstar.com
[5] "The Globalstar Cellular Satellite System", Fred Dietrich et al, IEEE Trans on Antennas and Propagation, Vol 46 Iss 6, Jun 1998
[6] ICO http://www.i-co.co.uk
[7] "ICO Global Satellite Communications System", Jai Singh, ICC97 Conference, Montreal
[8] "Orbcomm Low Earth Orbit Mobile Satellite Communications System" Todd Hara, IEEE Tactical Comms Conf 94
[9] UMTS http://www.umts-forum.org

APPLICABILITY OF TETRA FOR USE AS A MILITARY TACTICAL RADIO

C.M. Dearlove
J.C. Currie
Marconi Research Centre,
West Hanningfield Road, Great Baddow,
Chelmsford, Essex, CM2 8HN,
United Kingdom

D.J. Taylor
Marconi Communications,
New Street, Chelmsford, Essex, CM1 1PL
United Kingdom

ABSTRACT

The increasing need for military requirements to be met wherever possible by commercial off the shelf (COTS) equipment makes the use of modern civil mobile radio systems by the military of strong current interest. One system of particular note is TETRA (TErrestrial Trunked RAdio). This is a new digital Private Mobile Radio (PMR) system which has been standardised by the European Telecommunications Standards Institute (ETSI). TETRA has been designed to include the requirements of public safety users, and provides a number of features particularly suited for that purpose, some of which make the use of TETRA of particular interest for possible military use. This paper gives an outline of the TETRA system, and concludes that TETRA has a role to play in some military scenarios.

INTRODUCTION

Private Mobile Radio (PMR) systems have been developed to provide the communications for closed user groups such as the police and other public safety users, public utilities and by vehicle fleet operators. Major characteristics of typical PMR systems include:

- Dispatcher type operation
- Infrastructure sharing to aid spectrum efficiency
- Fast call set up
- Short messages
- Low cost (equipment and call charges)
- Reliable
- Primarily voice communications
- Short data messages

Provision of circuit mode data in older, usually analogue, PMR systems has often been limited.

TETRA (TErrestrial Trunked RAdio) is a new digital PMR system standardised by the European Telecommunications Standards Institute (ETSI). TETRA systems are now being deployed worldwide. A description of some aspects of TETRA may be found in [1], [2]; for full details the complete TETRA standards may be consulted; these can be found on the world wide web [3].

TETRA offers a number of advantages over the previous generation of PMR systems, including:

- Greater spectrum efficiency
- Single standard for voice and data
- Higher data rates giving more data services
- Simultaneous voice and data transmission
- Message security
- Higher quality and reliability
- Faster call set up times
- Wide range of supplementary services

COTS TECHNOLOGY

There are a number of benefits for the military by their procurement of commercial off the shelf (COTS) technology such as TETRA. Primarily there will be significant cost savings compared with the procurement of a bespoke system. It is likely that a COTS system will not meet all the military requirements but the cost savings may be so great that some lack of non-critical functionality may be acceptable. A COTS system that is an open standard (such as TETRA) will lead to the situation of the military being able to procure systems and equipment from multiple suppliers thus again leading to cost reductions. This can also ease the opportunity for interoperability between international forces and between different services. COTS systems by their nature are also likely to be available in shorter timescale than bespoke systems. Other advantages include compatibility with civil systems in public safety use and future proofing by following expected standards developments.

TETRA SYSTEM DESCRIPTION

TETRA is a system consisting of a number of interconnected elements with interfaces between them. There are two basic system architectures considered in this paper. The architecture of the first, the Voice plus

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

Data (V+D) system, is described below. The architecture of the second, Direct Mode Operation (DMO), including repeater and gateway operation, is also described below. A third architecture, the Packet Data Optimised (PDO) system is not discussed in this paper; note that the TETRA V+D system offers a number of packet data facilities.

Services

A TETRA system may offer a range of services to the user. These include:

- Circuit mode speech using an ACELP codec at a rate of 4.567 kbit/s uncoded (7.2 kbit/s with dedicated channel coding)
- Circuit mode data at 7.2 kbit/s (unprotected)
- Circuit mode data at 4.8 kbit/s (with low protection error control coding)
- Circuit mode data at 2.4 kbit/s (with high protection error control coding)
- Connection oriented and connectionless packet data with convergence to IP (these may be replaced by IP packet data only)

Higher data rates for the circuit mode data services are supported by the use of multiple timeslots. The maximum data rate which can be supported is 28.8 kbit/s by using four unprotected timeslots. Multiple timeslot protected circuit mode data is also supported, for example up to 9.6 kbit/s with high protection.

These services, voice and data, can be provided in the following ways:

- As an individual (one to one) call
- As a group call (one to many)
- As an acknowledged group call
- As a broadcast (one to many, one way only) call

Various supplementary services are provided by the TETRA standard, including:

- Telephony type supplementary services such as call forwarding, call barring and call holding
- Call authorised by dispatcher
- Area selection
- Access priority
- Priority call
- Late entry
- Pre-emptive priority call
- Discreet listening
- Ambience listening
- Dynamic group number assignment
- Remote temporary or permanent disabling of mobiles
- Remote re-enabling of temporarily disabled mobiles

Finally some key security features are also provided. These include air interface encryption and "hooks" for adding end to end encryption. In addition the capability for mutual authentication of mobile by network and network by mobile is also provided. Related functions include the option for Over The Air Rekeying (OTAR).

System Architecture – Voice plus Data

The TETRA V+D system is a cellular type system where communications between radios is via a base station. The TETRA system is designed to be economically scalable from a system using a single base station to, for example, a national system. Various types of handover are supported between base stations including "seamless" handover. The TETRA V+D system has a number of interfaces some of which are standardised, where ETSI has specified the form and detail of the interface and thus these interfaces are in the public domain. Other interfaces are proprietary to a manufacturer. The TETRA V+D system architecture is shown in Figure 1, indicating which interfaces are standardised.

The standardised interfaces are:

I1 Air Interface
I3 Inter-System Interface
I7 Gateway Interfaces

Interfaces which it is planned will have some standardisation, but at present are manufacturer specific are:

I2 Line Station Interface
I6 Network Management Interface

Interfaces which are proprietary to each manufacturer are:

I4 Interface between switches which comprise a single TETRA system
I5 Interface between a remote base station site and a central switch

An Example TETRA System

Owing to the existence of proprietary interfaces within a TETRA system, manufacturers are free to adopt a range of different system architectures for their TETRA systems. As an example the main features of the Marconi Communications ELETTRA TETRA system [4], [5] are presented. This shows an example of a scaleable architecture with decentralised switching capability. Other manufacturers may adopt a similar approach, or a more centralised hierarchical structure.

The ELETTRA system has an architecture based on a two level hierarchy (see Figure 4). The first level is the radio Base Station (BS) which provides the air interface to the Mobile Station (MS) as well as multiplexing and interfacing with the second level Switching and Control Node (SCN). The SCN is the main network node which manages call processing and switching of voice and data traffic, including circuit mode and packet data. The SCN thus interfaces to internal network elements such as BSs, other SCNs and any central databases. It also interfaces to Line Connected Terminals (LCTs), other TETRA networks and the Network Management Centre (NMC). It provides gateway facilities to other types of network. The architecture is flexible and scaleable, so that small to medium sized systems can be configured around switching base stations. These can operate either independently of, or in conjunction with, SCNs, which are the main switching centres of larger networks. A larger network may contain only a single SCN or else a network of SCNs, which may be interconnected in a resilient configuration (with a range of possible network topologies). In terms of the architecture presented above, the SCN provides all the interfaces other than the air interface (I1). The interface between SCNs is I4. The proprietary interfaces in the ELETTRA system are based on standard protocols; for details see below.

TETRA V+D Air Interface

The main air interface between the BS and the MS is that used to transmit both voice and data, which may include packet data. The TETRA air interface is specified, as is usual, by a protocol stack which, in this case, covers the lower three layers of the standard OSI reference model: physical, data link and network layers (see Figure 5). This protocol stack includes, as is also typical, subdivisions of these layers. The various layers are now briefly described from bottom to top (layer 1 to layer 3).

The physical layer (layer 1) is concerned with radio aspects such as frequency use, duplexing, power control implementation, burst building and multiplexing, modulation, and frequency and symbol synchronisation.

The data link layer (layer 2) is divided into a lower sublayer handling Medium Access Control (MAC) and a higher sublayer concerned with Logical Link Control (LLC). The MAC layer is also divided into two, a lower MAC layer and an upper MAC layer. The lower MAC layer includes channel coding (including interleaving and scrambling) and decoding (including descrambling and deinterleaving) and mapping logical channels to the appropriate physical channels. The upper MAC layer includes frame and multiframe building and synchronisation, random access control and organisation of logical channels. The LLC layer includes data transmission and retransmission and logical link handling.

The network layer (layer 3) is divided into a lower sublayer, the Mobile/base Link control Entity (MLE), and a number of separate upper layers, the Sub-Network Access Functions (SNAF), handling Mobility Management (MM), Circuit Mode Control Entity (CMCE) and Packet Data (PD). The CMCE is subdivided into three entities, Call Control (CC), Supplementary Services (SS) control and the Short Data Service (SDS). Packet data is provided by two services, the Connection Oriented Network Protocol (CONP) and the TETRA Specific ConnectionLess Network Protocol (SCLNP).

Some of the other key features of the air interface are as follows:

- The frequency bands used by a given TETRA system and their separation will be specified for that system and allocated by the appropriate National Regulatory Administration (NRA). The TETRA system has been designed to work in the frequency range from 150 MHz (currently standardised from 300 MHz) to 1 GHz. A number of frequency bands have been allocated within this range, including for public safety use, from 380 MHz to 390 MHz and 390 MHz to 400 MHz for the two link directions, with a 10 MHz duplex separation. TETRA equipment is available covering a number of frequency bands.
- Carriers have a separation of 25 kHz (in both uplink and downlink bands). There are four physical channels per carrier, using Time Division Multiple Access (TDMA). Timeslot length is approximately 14.17 ms. A TDMA frame thus consists of 4 timeslots, one per physical channel, lasting approximately 56.67 ms.
- To allow duplex use, with the MS both transmitting on the uplink and receiving on the downlink, but without doing both simultaneously, the start of an uplink TDMA frame is delayed by two timeslots after the start of the corresponding downlink TDMA frame. This also permits access information about an uplink timeslot to be transmitted on the corresponding downlink timeslot, for use by the mobile access mechanism.
- Transmission uses $\pi/4$-shifted Differential Quaternary Phase Shift Keying ($\pi/4$-DQPSK). The modulation rate is 36 kbit/s (18 ksymbol/s). $\pi/4$-DQPSK is a linear modulation method, used because of its reasonably high spectral efficiency, which requires a linear transmitter; the TETRA frame structure provides suitable opportunities for transmitter linearisation algorithms to be used.
- Every 18th TDMA frame is reserved for control information, in particular including information broadcast to all mobiles using, or attempting to use, a base station. In usual use one or more channels (timeslots on a single carrier) are also reserved for control signalling to individual mobiles. For fast signalling timeslots may be "stolen" from traffic channels; this mechanism may be used to implement synchronisation for end to end encryption.

Other Interfaces

As described above most of the interfaces in a TETRA V+D system other than the air interfaces are not standardised. Details of some of these interfaces are illustrated here by indicating the approach taken in the Marconi Communications ELETTRA system described above. It is expected that other manufacturers' TETRA systems will have some similarities, some differences are expected also.

Line Station Interface (I2)

This is a digital interface to the switch allowing the connection of a Line Station (LS), e.g. a dispatcher terminal, to the system. This interface will support voice and data calls to subscribers, usually at a higher priority than a radio user. The interface will also permit features such as regrouping and radio unit disabling. Manufacturers may include access into the subscriber management system at the line station interface.

This interface is manufacturer specific, but is likely to be 64 kbit/s G.703 or basic rate ISDN. The Marconi Communications ELETTRA system uses a 64 kbit/s circuit, with four 8 kbit/s sub-multiplexed traffic channels, and a higher speed control link. Four TETRA codecs are fitted at the dispatcher, on an interface card which is inside the dispatcher terminal. This avoids the need for transcoding, and would allow end to end encryption to be carried through to the dispatcher terminal, if required.

The dispatcher is able to access certain functions relating to subscribers, such as configuring and regrouping, and needs to have access security before the operator can log on to the system, and then to access the databases. It is possible to apply a form of the air interface encryption on the control link to the dispatcher.

Inter-System Interface (I3)

This is a standardised interface permitting TETRA systems from different manufacturers to be linked, either to provide a larger coverage area, or to link two disciplines. An example would be where different police authorities independently purchase TETRA systems; these systems could interoperate, but would need to do so via the interface I3 if the two authorities were not to be constrained to use the same manufacturer. The inter-system interface allows users to be able to roam onto the other system, and operate as though they were on the home system. This in turn means that database information and subscriber details will travel across the interface.

Physically the interface is built on the top of Private Signalling System 1 (PSS1), also known as Q-SIG, and will normally be implemented as a 2 Mbit/s G.703 circuit. The inter-system interface (ISI) supports authentication, allowing a mobile on a visited system to be authenticated without revealing the original authentication key of the MS. A migrated subscriber can also authenticate the visited infrastructure implicitly by authenticating the home system using the visited system as an agent. The ISI also includes a mechanism to support the synchronisation of synchronous stream ciphers, for end to end encryption over the ISI.

Inter-Switch Interface (I4)

This is an interface between switches of the same manufacturer, when those switches are being used to provide a system larger than that available with a single switch. The physical and logical protocol for this interface is proprietary, but provides as a minimum the functionality of interface I3, with the addition of full roaming and network management capabilities.

The Marconi Communications ELETTRA system uses a 2 Mbit/s circuit for this interface, similar to the ISI, although the traffic channels may be carried in a compressed form, such as 8 kbit/s submultiplexing.

Base Station to Switch Interface (I5)

This interface may be, for example, between a remote hilltop base station and a centrally located switch. The interface is proprietary to each manufacturer.

The Marconi Communications ELETTRA system implements this link as a number of 64 kbit/s circuits, typically one for each carrier on a site, and the links either being several discrete links or timeslots in an E1 2 Mbit/s circuit. The 64 kbit/s circuit has traffic (voice or data) channels sub-multiplexed at 8 kbit/s into the circuit, plus a higher speed control and supervisory circuit.

Network Management Interface (I6)

This provides an interface for access to the TETRA system for network management activities, including configuration management, subscriber management, alarm management and security management. This interface may be standardised by ETSI, but is currently proprietary for each manufacturer, but based on industry standard networking and management protocols.

The Marconi Communications ELETTRA system has a LAN/WAN connection with Ethernet between the Switching and Control Node (SCN) and the Network Management System (NMS). There will be a NMS server with a number of client PCs, all operating under Windows NT, and standard network management protocols.

PABX/PSTN and Other Gateway Interfaces (I7)

These standardised interfaces include normal telephony interfaces, which may be analogue or digital, and interfaces to data networks. These telephony interfaces will provide an alternative means of providing gateways to other systems, which need not be other TETRA systems. These could be 4 wire PABX trunk interfaces, with simple signalling and facilities, or digital Q-SIG or other high level interfaces.

The Marconi Communications ELETTRA system has both analogue and digital interfaces. While digital interfaces are preferred, evidence from customer enquiries shows that there is still demand for analogue interfaces. The system will have transcoders between the

TETRA signal and the analogue or PCM interfaces. If traffic on the system is to use end to end encryption, there will also be encryption gateways.

Direct Mode Operation

TETRA Direct Mode Operation (DMO) is defined in a separate set of standards to the TETRA V+D standards. In its simplest form direct mode will be directly from mobile station to mobile station, using a modified air interface I8 and without using a base station, as illustrated in Figure 2. Direct mode operation can be extended using a repeater, and enhanced using a gateway which provides connectivity from direct mode to a TETRA infrastructure. These two operational modes, and the interfaces they use, are illustrated in Figure 3. The main purpose of direct mode is for two or more MSs to be able to communicate directly with each other, or in a group, without using any additional infrastructure (i.e. BSs). A TETRA MS may be able to use either V+D or DMO, or both; in the latter case it may or may not be capable of operating dual watch. This is a feature where either an MS may monitor a BS whilst using direct mode but be able to switch to using the BS if a call is indicated by the BS for the MS, or an MS using a BS may switch to direct mode if a direct mode call is indicated.

At the lower levels of the TETRA protocol stack the direct mode air interface is similar to the V+D air interface. At the higher levels of the TETRA protocol stack direct mode is much simpler. Direct mode also supports a much simpler set of services, principally circuit mode speech or data at up to 7.2 kbit/s, and short data messages. Direct mode permits only simplex operation. Security is weaker than the V+D system. Direct mode can support air interface encryption, and includes the same "hooks" for end to end encryption. However owing to the absence of an infrastructure, and hence a user database, air interface encryption has to use static cipher keys, which are available to all appropriate users, rather than an individualised system. In addition authentication is principally that if the other party can decrypt messages it is authenticated.

SCENARIOS

There are a range of military scenarios where TETRA may be suitable. These could range from relatively benign scenarios where operations are being carried out in peacetime, to "harder" scenarios where the threats to the communications system could be greater and where the need for secure communications is greater.

At the lower end of the scale there are scenarios such as providing the communications at a static site, such as an army barracks, where in peacetime the threat to such a site is relatively low and thus a TETRA system is likely to provide most of the communication requirements and features required. There are also some roles, such as the military police, where the communications needs are so similar to their civil counterparts that again TETRA is likely to offer most of the requirements. Another scenario where TETRA could play a part is in Military Aid to the Civil Authorities (MACA) including times of disaster or when the military take the place of civil operatives, for example replacing unavailable civilian fire brigades. At the "harder" end of the scale of scenarios there will operations where the threat to communications could be significantly greater and therefore great care would be needed before deciding upon the use of a system such as TETRA, which might however be appropriate for specialised uses, in particular in rear support areas. Intermediate scenarios of interest could be evacuation and peacekeeping roles in other countries. TETRA direct mode is likely to be of particular application to a wide range of military requirements. Note that other civil radio systems are also likely to suffer from the same, and in some cases even greater, weaknesses to threats as TETRA.

MILITARY BENEFITS OF TETRA

There are a number of issues that can be identified as key to the consideration of TETRA (or any other radio system) being acceptable in a military role. These include the following:

- Voice service
- Data services
- Supplementary services
- "Architecture" issues
- Frequencies of operation
- Coverage and cell size issues
- Security issues
- ECM/ESM vulnerabilities
- EMC considerations
- Interoperability
- Deployment
- Environmental issues
- Standard and adapted products
- Development time
- Cost

These issues are now considered in turn.

Voice Service

As noted above TETRA uses an ACELP speech codec which with its dedicated channel coding uses a data rate of 7.2 kbit/s. This was designed and tested to operate under the high noise and interference conditions sometimes typical of public safety operations and thus may be expected to be suitable for many military uses.

This codec is not in military use, hence in many scenarios speech would have to be transcoded. Transcoding is a potential system performance limiting factor, and appropriate testing would be required.

Data Services

The range of TETRA data services (2.4 kbit/s up to 28.8 kbit/s, depending on bandwidth and error protection, plus short data messages and packet data) are described above. Such data rates may be used by civilian systems to offer features such as still image and slow scan video. Automatic position reporting is not included in the TETRA standard, but it is expected that some or most manufacturers will offer mobiles (including handheld units) incorporating a GPS receiver, with the facility to use this to send short messages at appropriate intervals.

The various data services (other than short messages, which can be directly displayed on a mobile unit) will typically be output from a standard data port on the mobile which may be connected to a standard laptop or palmtop computer.

Supplementary Services

TETRA offers a number of other supplementary services and features that may be either indirectly or directly applicable to military use.

- **Call Authorised by Dispatcher.** This restricts user access to certain facilities, and means that users can only perform these functions when authorised by the dispatcher (control station or operations room). This could be used by the military to, for example, restrict user access to the PSTN.
- **Area Selection.** This allows an authorised user to define areas and allows other users to select those areas as the basis for establishing calls. This effectively creates an all-informed group that could be used as the basis for the dissemination of information relevant to that area. This feature could be a particularly useful in military scenarios when information has to be transferred rapidly from headquarters to troops on the ground. The mechanism allows this function to be performed quickly and avoids subordinate stations in the chain of command having to relay the information.
- **Access Priority.** This enables a user to gain access to a TETRA system uplink in order to make a call during periods of heavy traffic. This is achieved by assigning an access priority level to individuals or groups. TETRA provides 8 priority levels. Individuals or groups of users with the highest priorities will be given preferential uplink access in the event of there being insufficient radio channels.
- **Priority Call.** This enables a user to gain preferential access to TETRA network infrastructure assets to make a call during periods of heavy traffic. Like access priority, this is achieved by assigning a priority level to individuals or groups. Individuals or groups with the highest priorities will be preferentially allocated network resources (i.e. everything except the uplink).
- **Pre-Emptive Priority Call.** This allows privileged users to have the ability to make individual or group calls, even if this results in users with lower priorities having their calls disrupted. This could be an important feature of a military system. It could be used by commanders and other key personnel to force transmissions through to other users when operationally required to do so.
- **Late Entry.** This allows users to join a group call after it has started. This will be useful in a military scenario since it allows users who were suffering from a poor signal or other problem at the beginning of the call to subsequently join it. This mimics the operation of a conventional radio net.
- **Dynamic Group Number Assignment.** This enables either authorised or served users to create, modify or delete groups even if the calls are in progress. This tool could be valuable during operational scenarios, for example when neighbouring units had to perform a common task along a boundary.
- **Ambience Listening.** This allows a control point to monitor the activity of a selected MS or LS. It does this by causing the MS/LS to transmit without the user's knowledge. This could be used by the military in situations were the user was unable to transmit by operating the press to talk button.
- **Discreet Listening.** This enables an authorised user to monitor a conversation between other users. This feature could be used by communications security (COMSEC) monitoring teams. It should be noted that this feature will probably not be compatible with end to end encryption.
- **Remote Disable.** This allows a control station or dispatcher to disable, either temporarily or permanently, a TETRA radio. This would be extremely useful in the event of a military radio being lost or captured.
- **Remote Enable.** This allows a control station or dispatcher to re-enable a temporarily disabled TETRA radio. A permanently disabled TETRA radio will require some other action to re-enable it, typically requiring physical access to the radio to reprogram erased data.

"Architecture" Issues

The TETRA system architecture has been described above. A number of features are particularly relevant to possible military use. Implementations which take full advantage of the network scaleability options of the TETRA standard to permit efficient single base station operation, as well as large network operation, will be most appropriate for military use where a range of network sizes is expected. TETRA direct mode, including repeaters and gateways, is of particular importance, since it may be used to match a number of military requirements, such as communications within small units. It is likely that direct mode mobiles will be handheld, with repeaters and gateways vehicle mounted.

Frequencies of Operation

The frequencies issue is more complex than purely a question of optimising coverage. There are several other frequency related issues that need to be considered prior to the procurement of a TETRA system. TETRA is currently specified from 300 MHz to 1 GHz, with the former figure intended to be reduced to 150 MHz. A specific TETRA system will operate in a band within that range, such as from 380 MHz to 430 MHz. Equipment operating in a number of such bands is already available, more will become available according to customer demand.

If the system is designated for use worldwide then effort would need to be applied in order to identify optimum frequency bands. It is unlikely that a single frequency band will fulfil all requirements. Implementation of a new frequency band by a manufacturer will be expensive and time consuming, although this additional cost will reduce as more bands are implemented. It is expected that multiple band radios will become available if there is sufficient demand.

Coverage and Cell Size Issues

Cellular networks are not entirely straightforward to plan and this can lead to a requirement for a significantly greater expenditure on infrastructure than perhaps would be expected. As terrain gets progressively rougher and/or more highly developed, the problem escalates in magnitude. In the urban area military users are probably more likely to be on foot using handheld terminals than the users of civil networks. This further complicates the coverage issue. Therefore, the cost of infrastructure to support a community of users will depend critically on the terrain in which the equipment will be used. In these circumstances the use of direct mode and repeaters may assist in reducing the number of base stations required to provide suitable coverage across an area. Use of alternate frequencies may also be of benefit, however the costs of rebanding equipment would need to be assessed.

Security Issues

The suitability of TETRA for military use depends on the grades of traffic that will be permitted to use the system, which will be the decision of appropriate national authorities. It is expected that a military system would use TETRA air interface encryption and mutual authentication between mobile and base station. Increased security, and hence permitted grades of traffic, may be possible if the appropriate TETRA "hooks" are employed to implement end to end encryption of circuit mode voice and data. This may be for some or all users; the infrastructure (base stations) would be unmodified.

TETRA offers the facility for over the air rekeying for air interface encryption, depending on appropriate key management infrastructure and procedures. This is likely to be useful for military use, if permitted. OTAR may also be implementable for end to end encryption.

ECM/ESM Vulnerabilities

COTS systems are not designed, in general, to handle deliberate ECM and ESM threats, although some consideration has been given to these for public safety use, which may make TETRA at least as capable as other COTS systems. Obvious threats include jamming and exploiting broadcast information, however it is not possible, in an unclassified paper, to describe details of the vulnerabilities of a TETRA system or what might be done to counter these.

EMC Considerations

The EMC performance of TETRA equipment is essentially that of a commercial specification and may not match well with existing military specifications. There is likely to be a requirement for the military to update these to cater for modern communications equipment. However regardless of the specifications, there will always be a requirement for testing. This is expensive and, due to the wide range of military vehicles, and other installations, complex. It may therefore be the case that the integration of TETRA mobile terminals into combat vehicles already fitted with other radio equipment will be the exception rather than the rule. This still leaves a large class of users for whom TETRA, particularly using handheld units, is appropriate, giving them realistic mobile functionality.

A particular issue would be the requirement for airborne use. Many military scenarios include interworking with helicopters in particular. Airborne use of a system such as TETRA is particularly difficult due to the greater interference that a mobile unit can cause (it is likely to be in line of sight of many base stations) and, especially, the cost of EMC testing of an airborne platform. The latter consideration would appear to make airborne use of TETRA unlikely. However it may be noted that similar issues apply to public safety users, and airborne use of TETRA, or at least airborne interworking with a TETRA network, is the subject of active consideration.

Interoperability

TETRA systems (possibly procured from different manufacturers) can be linked by means of the inter-system interface. This means that different elements of a national force could procure TETRA systems from different manufacturers which could then be linked by means of the ISI interface to form a "single" network; in addition different TETRA networks from different parts of a coalition force could also be linked together.

Standard TETRA interfaces will permit interfacing to other TETRA systems and various civilian systems (PSTN, ISDN). Interfacing to almost any system is possible using a TETRA interface to ISDN or to a PABX, to which the other system may be interfaced. The degree of interoperability will vary according to the other system. Two TETRA systems (without nonstandard enhancements) should be able to fully interoperate, however this is a configuration and management issue for both systems, particularly with regard to encryption. Nonstandard features (such as end to end encryption – note that the capability to add this is standard, the method of addition is not) will only interoperate if implemented compatibly by both systems. Between other systems interoperability of simple voice calls and possibly data connections should be possible; voice transcoding will probably be required at the interface (other systems are unlikely to be employing the TETRA speech codec) and hence a secure gateway may be needed. Within the UK the emergency services are expected to move to a TETRA system (PSRCP) and interfacing with this should be possible (including military TETRA radios being able to access the PSRCP network directly – this is principally a management issue).

Interfaces to many military systems can use the above method (via ISDN or a PABX). It may also be possible to use a trunked military system to carry TETRA network traffic. Similar interoperability issues apply as for civilian systems, with a more likely need for a secure gateway.

TETRA equipment should be able to interface to the PSTN, ISDN and PABXs and, via one or more of these, to other mobile networks. It is also possible that other PMR systems will be modified to comply with the TETRA ISI. This ability to interconnect can allow a degree of interoperability with other organisations even though the communications systems are different. There may be performance limitations but it is certainly an improvement on no communications at all. In addition, it permits calls to and from fixed public networks and also gives the potential for public bearer services to interconnect islands of TETRA coverage in a low density operational scenario.

Deployment

Many civilian TETRA systems will not require frequent physical configuration changes, other than (in some cases) addition of otherwise fixed base stations. Major reorganisation of such a network configuration is likely to be an infrequent event. However some civilian, particularly emergency service, systems require configuration changes, such as including transportable parts of the infrastructure (for example to cover major events or fill in gaps in propagation) and hence some or most manufacturers will support this facility, which may even include the possibility of a vehicle mounted moving base station (working in isolation, not connected to a network). A short timescale to deploy such equipment is essential and expected to be feasible. A base station capable of independent operation fitted in a vehicle could be deployed simply by driving to or air dropping at the required location, setting up an antenna and switching on. Establishing a link to the rest of the network would take a little longer, depending on the nature of the link and positional knowledge.

Reconfiguration within a network (such as a change of frequencies) is expected to be reasonably straightforward (at least by a manual management change, this could probably be made automatic if such a facility were required).

Environmental Issues

All communications equipment, civilian and military, has to handle a range of environmental conditions covering such aspects as temperature, water resistance and shock and vibration considerations. The range of conditions is usually wider for military equipment than it is for COTS civilian equipment, and this may limit the deployment of unmodified COTS equipment.

A typical COTS system will operate over a temperature range such as -10 °C to +55 °C, with limited water resistance. (Storage may be over a wider temperature range.) A wider temperature range requirement will have implications for the use of COTS equipment. Some civilian systems will require a greater degree of water resistance than will be available as standard and it is expected that some manufacturers will provide for this, however this will have some limitations.

Public safety use is likely to have one of the most demanding civilian environmental requirements.

Standard and Adapted Products

In a number of circumstances noted above there may be restrictions on the deployment of unmodified COTS products. However in many cases the advantages of a TETRA system may still be achieved by considering the possible adaption of otherwise standard COTS TETRA products. Potential adaptations include physical product packaging (for more difficult environmental and EMC requirements) and communications security (such as using TETRA “hooks” to implement end to end encryption).

Development Time

TETRA systems are currently being deployed worldwide, demonstrating that TETRA technology is already available. However the TETRA standards contain a wide range of options, not all of which may have yet been implemented by manufacturers. Development of such features is generally market driven, and the market (especially public safety use) is already driving the development of most of the options which are likely to be required, and in a timescale which is short compared to typical military procurement programme timescales. Development timescales are likely to be longest for any secure network required.

Cost

The primary motivating factor for the consideration of TETRA or other COTS systems is cost. TETRA would offer substantial cost savings compared to specialised military systems, particularly when the advantages of open procurement (both initially and for followup orders) are taken into consideration. The main exception to this would be that a TETRA radio would be more expensive (due to its greater functionality) than a simple personal role radio, although a cost saving would be possible if combined with other functions, such as radio access to a military trunked network.

CONCLUSIONS

COTS systems offer a number of advantages to military users. The primary advantage is cost compared to specialised military equipment, but other advantages include standardisation and independence from single manufacturers and expected upgrade paths, as well as potential advantages in development and procurement timescales.

TETRA is a COTS system developed for professional users, including public safety users, whose requirements were part of the TETRA design process. Consequently TETRA includes a number of features of potential military interest including a wide range of services and supplementary services, an architecture which is scaleable from a single base station to a national network, direct mode operation without infrastructure and the ability to select from a wide range of frequency bands.

The principal drawbacks to the use of TETRA are those inherent in COTS systems, that they are not designed to handle the full range of military requirements. This applies in particular to ECM/ESM considerations. TETRA is not however believed to be especially vulnerable. (Note that consideration of this topic in this paper is limited by its classification.)

The applicability of TETRA to specific military requirements depends on expected usage scenarios. TETRA is able to meet most, if not all, of the requirements of a peacetime, low threat, relatively low security scenario such as barracks security or use on training ranges. At the other end of the scale TETRA is not suitable for use in forward positions in a war-fighting operation, however it may be suitable for rear area support functions. Between these extremes there is no simple answer as to where TETRA is or is not applicable, although key factors in such a decision are noted above. In particular because of cost considerations a system such as TETRA may offer the option for a facility that could otherwise not be afforded.

Overall it is concluded that where military budgets are constrained, and the use of COTS communications equipment is being considered, that TETRA has a part to play.

ACKNOWLEDGEMENTS

The material presented in this paper is taken from a Pathfinder collaborative project on Security and the Military Use of TETRA undertaken by the Defence Evaluation Research Agency (DERA) Malvern, Marconi Communications Chelmsford, Marconi Communications Infosec and the Marconi Research Centre.

REFERENCES

[1] ETSI Technical Report, "Terrestrial Trunked Radio (TETRA); Voice plus Data (V+D); Designers' guide; Part 1: Overview, technical description and radio aspects", ETR 300-1: May 1997.

[2] ETSI Technical Report, "Terrestrial Trunked Radio (TETRA); Voice plus Data (V+D); Designers' guide; Part 2: Radio channels, network protocols and service performance", ETR 300-2: May 1997.

[3] http://www.etsi.org/TETRA/TETRA.htm

[4] Marconi Communications "TETRA: ELETTRA Digital Trunked Mobile Radio System", MS2/6171, SIMPDES2.

[5] http://www.marconicomms.com/products/mobile_net/

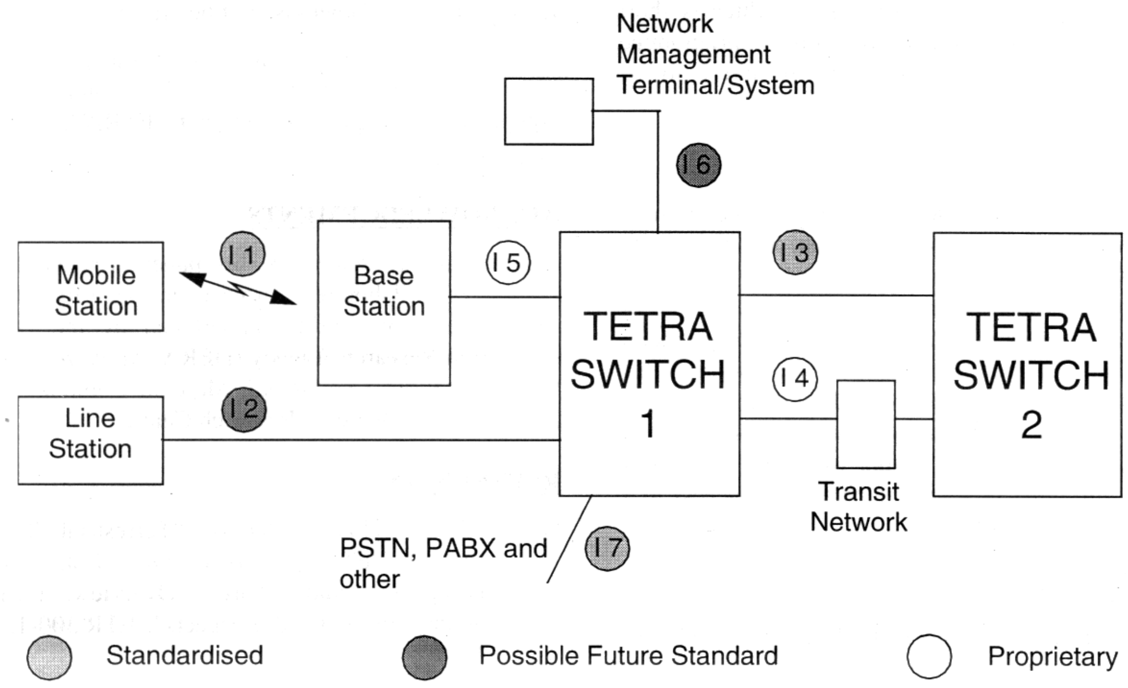


Figure 1: TETRA V+D Infrastructure Interfaces

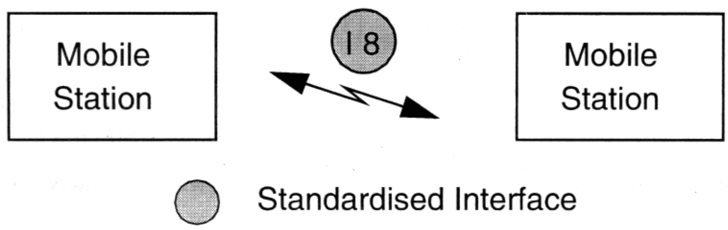


Figure 2: TETRA Direct Mode Operation

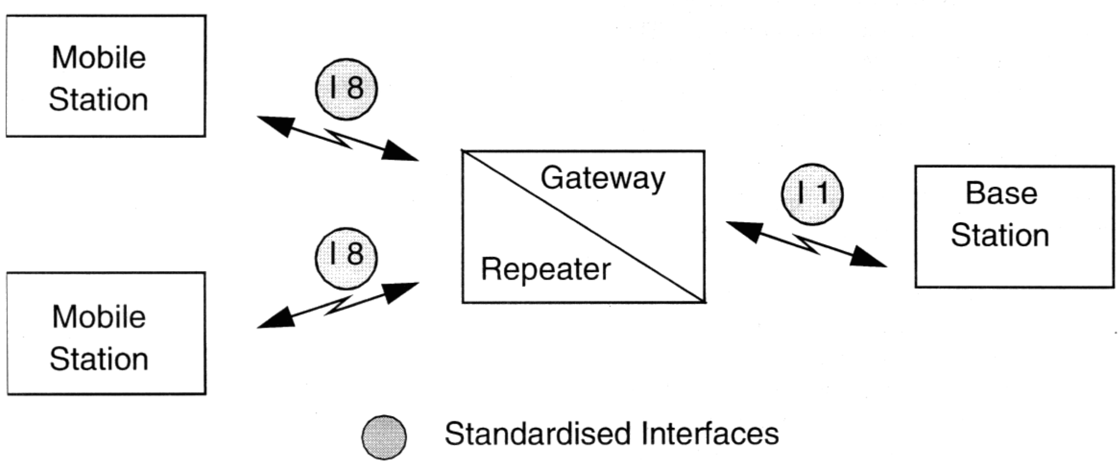


Figure 3: TETRA Direct Mode Repeater/Gateway

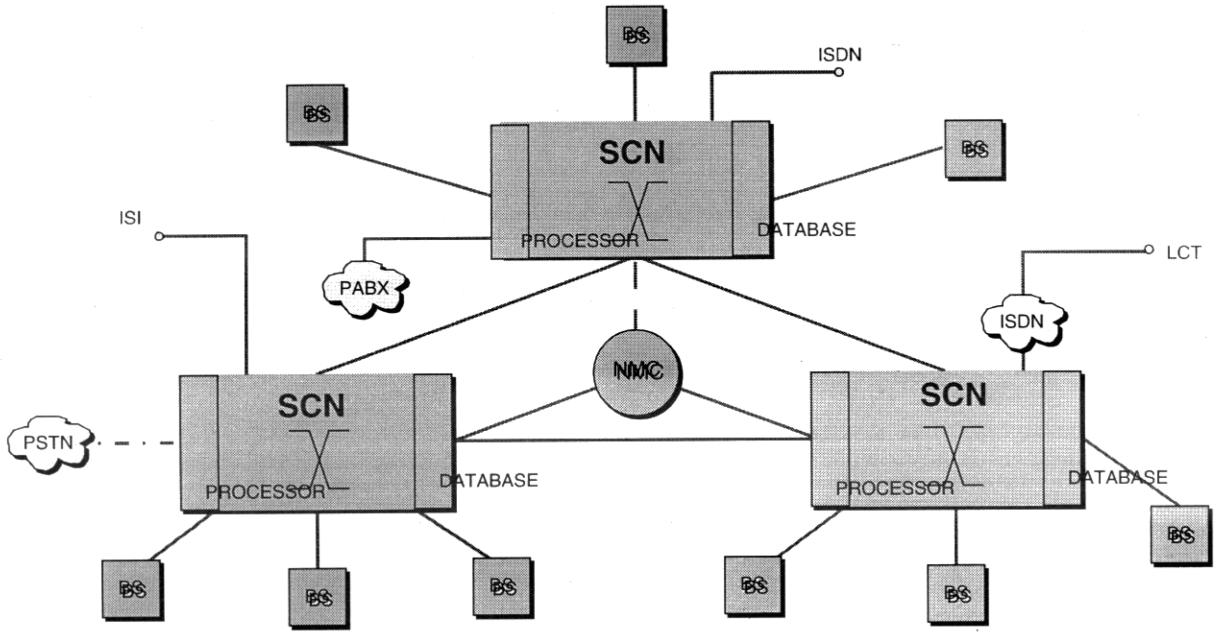


Figure 4: ELETTRA System Architecture

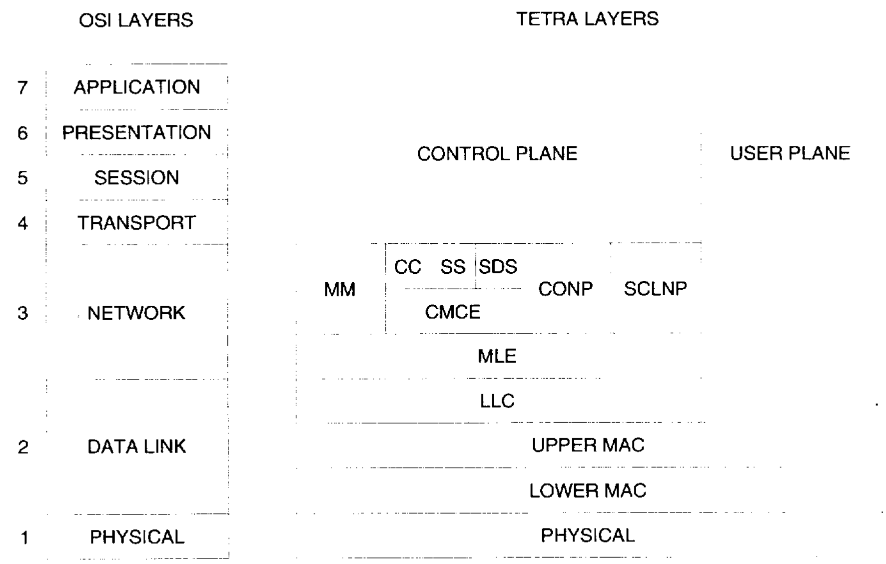


Figure 5: TETRA Air Interface Protocol Stack

Supporting User & Infrastructure Mobility in the Tactical Environment

(May 1999)

Rick Barfoot, Jason Daniell

Communications Department,
Defence Evaluation & Research Agency (DERA),
Saint Andrews Road, Malvern,
Worcestershire, WR14 3PS
United Kingdom

Summary

The military environment poses a challenge for the support of high levels of subscriber mobility. The requirement for an increasingly fluid infrastructure further compounds this problem. Additionally, these requirements have to be met by an infrastructure that is resilient to fault and loss, but constrained by limited link capacity. This paper indicates techniques that are being examined in DERA's research programme to help overcome these issues. These ideas centre on the use of prediction techniques to position appropriate information for resilience and responsiveness, and then slowly adapt the location to the changing infrastructure.

Introduction

This paper focuses on the efforts to close the gap between commercial capabilities and military aspirations for manageable, secure and robust tactical communications. To this end, we address the characteristics of military mobility and identify the disparity with commercial systems and standards. Finally, we describe the mechanisms that we are examining in an effort to close the differential.

The Military Environment

The conduct of warfare has evolved from the relatively static deployments of the First World War, through the mobile engagements of the Second World War to the highly manoeuvrable exercise of the Gulf War. The ability to rapidly move combat forces, by land, sea or air, is a central part of modern military doctrine. As well as the high degree of mobility, military communications infrastructure supporting these forces must be able to tolerate high levels of loss and stress, respond rapidly to changes in demand, be flexible in use and manageable by soldiers in adverse conditions.

Stress on the infrastructure may be caused by attack, at the physical, electronic or even information level. Stress may also result from equipment failure, RF interference and inadequate planning. The latter may be due to the paucity of information from which to plan.

This environment poses an exciting and significant challenge to the communications system designer. The system design must be able to support users that are constantly on the move and either require communications services whilst on the move, or have rapid access to services once stationary. These services not only have to be resilient to the stress mentioned, but must also protect against unintentionally revealing valuable information to an adversary. Therefore security is a key requirement in any military system.

To add to the designer's challenge there is no guarantee of the availability of a high bandwidth, low error and reliable optical infrastructure. This may not exist, have been damaged beyond use or simply be in the control of the wrong hands. Therefore there is a requirement for a transportable, secure and resilient communications infrastructure. Extensive use has to be made of radio frequency bearers, usually providing a mix of terrestrial and satellite capability.

To date, the mobile user has been supported by military bespoke systems based either on all informed netted radio, similar to PMR[1], or radio access to trunk systems, similar to PCS[2] (e.g. GSM[3]). Indeed the military had these systems before any wide scale commercial use of analogue or digital systems. Having got there first, military technology has been overtaken by rapid development in the civil world.

---

[1] Personal Mobile Radio
[2] Personal Communications System
[3] Global System for Mobile Communications

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

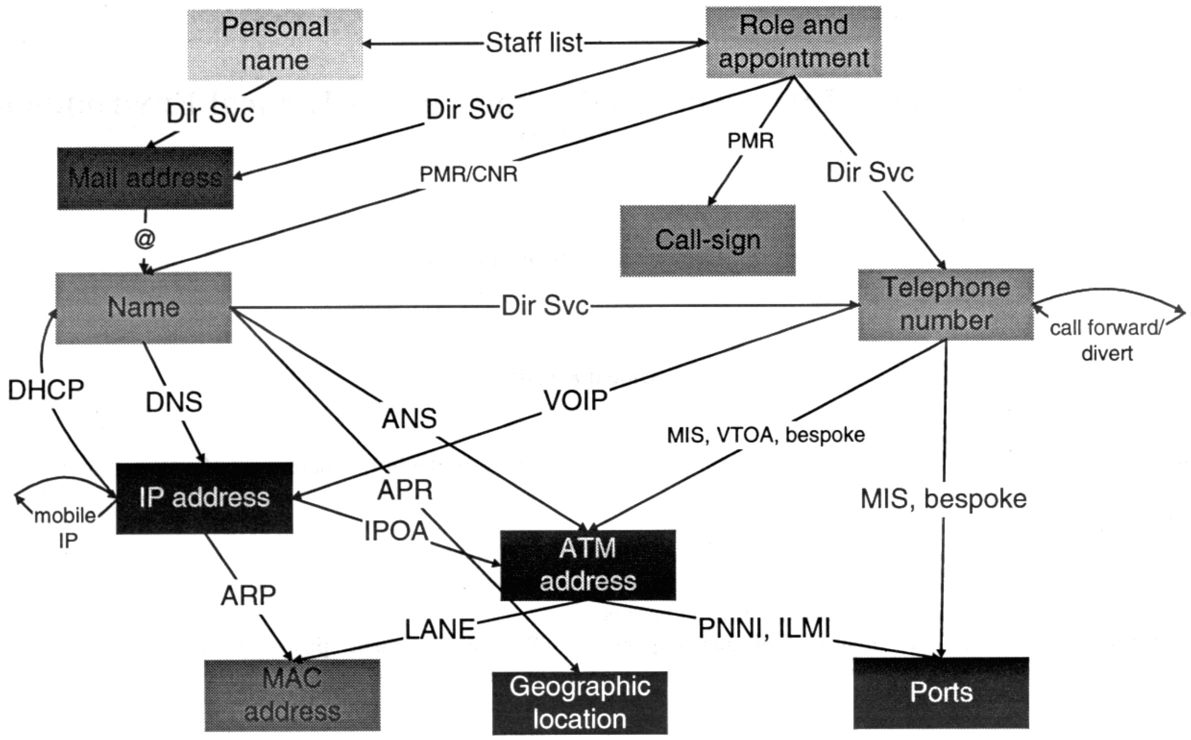


Figure 1; Example Mobility Model

Military system designers plan to, and are making, full use of the enormous civil investment. However, commercial systems are designed to cost constraints and functionality is determined by the minimum requirements to make the system commercially competitive, unnecessary functionality adds cost. Therefore essential military requirements are often not addressed at all or are implemented in insufficient depth.

Additionally, user expectations are high due to the increasing availability of flexible multimedia technology. These new applications make ever-greater demands for bandwidth on the supporting infrastructure. As well as differences between the military and civil environments there are still many technical challenges that both communities have to face.

Technical Challenges

User mobility is a rich problem. There are many mappings between layers in a communications system that can change resulting in mobility or apparent mobility. Figure 1 above illustrates the complexity that can exist and indicates some of the mechanisms used to solve the issue.

Although this illustration is taken from the military domain, most if not all, applies to the civil domain. At the top, the user may be referred to using their personal name, role or other label. The connection between these designations may be via a staff list or other mapping that is not expected to change frequently. In current military systems, role often equates to the name. Further, these initial labels (role, personal name, etc.) may be associated with an e-mail address, telephone and fax numbers, and other service oriented identifiers. These mapping functions are often implemented using directory service mechanisms and serve as a means for matching a user label to a service label. User mobility is typically not handled at this level.

To utilise a service, it is necessary to translate from the service label to a protocol identifier, e.g. an e-mail address to an IP address. The protocol identifier serves as a routeable address for the destination user or service (e.g. voice mail server). This translation function is often implemented using name servers. User mobility could therefore potentially be handled at this level, however this solution is not scaleable. The problem is that a large number of mobile users could generate an unacceptable load through significant update messages and service requests. Commercial technologies therefore attempt to hide the user mobility from the name server through a variety of techniques.

User mobility has the effect of changing some of the mappings between layers. There are circumstances where mappings can change without anything physically moving.

The overall effect can be the same and could be described as virtual mobility. A good example in the military domain is a change of command. Headquarters normally have a back up, ideally identical, that can take over if the need arises, either as part of a plan or in an emergency. This results in an exchange of roles for some subscribers between the two HQs whilst others remain unchanged.

The issues raised so far address locating the user, this is generally referred to as location management. Maintaining services to that subscriber once they have been established is the task of mobility management.

The infrastructure supporting the user can, itself, be mobile. With current military systems, communications nodes provide discontinuous service support, i.e. they disconnect from the operational system whilst they are on the move. For planned movement controlled engineering and de-engineering of links can be used, with the engineering process providing the stimulus to update the stored network topology. In the military scenario, the unplanned movement or loss of infrastructure is also likely to occur. This will require the network to react quickly and minimise the impact of any disruption.

An aspiration is to provide continuous services whilst infrastructure nodes move. Long range bearers, such as geo-stationary satellite, can accommodate user movement without a change in network topology. Shorter range systems based on lower elevation platforms will demand mobility management techniques capable of managing handover of many multi-channel links at once; all without any interruption to service. Example low elevation platforms include; terrestrial, airborne (UAV[4]) and LEO[5] satellite. This support to user and infrastructure mobility must be maintained even if parts of the network are lost or fragmentation occurs.

Existing Solutions

To date the research in DERA has focused on defining the requirements for mobility (i.e. the problem), the issues of location management and the user mobility aspects of mobility management. The demanding aspects of mobility management for mobile infrastructure are in the early stages of investigation. Many existing civil technologies have been examined and their shortfalls identified, an overview of the principle mechanisms employed follows.

<u>GSM</u>

GSM location management and mobility management is handled through the combination of a home location register (HLR) and a visitor location register (VLR). A subscriber will be allocated to an HLR, which stores all necessary administrative information for him as well as an indication of his location. This location indicator is usually the signalling address of the VLR associated with the subscriber's current whereabouts (local base-station). The VLR contains a selected subset of administrative information from the HLR to allow the VLR to provide service to the subscriber.

A subscriber is associated with a single base-station, adjacent base-stations may be grouped together to form a location area. A VLR may be responsible for one or more location areas. As a subscriber moves between location areas, the VLR responsible for the new area is informed. If the VLR responsible for the new area is a new VLR, then the HLR is informed. To locate a subscriber, all the base-stations within the location area associated with the subscriber will be paged to determine exactly which base-station the subscriber is attached.

The processing load on the HLR is significantly reduced through the usage of the VLRs to hide subscriber mobility from the HLR. Similarly, the processing load on the VLRs is reduced through the application of location areas and base-station paging.

The GSM infrastructure is typically geographically fixed with centralised management functions. The threat to commercial operators is low and resilience is achieved through the use of backup equipment for essential services, this is often co-located with the primary equipment. In addition, the infrastructure supporting GSM is not designed to be mobile and needs to be configured with care using a management system.

Unlike a GSM subscriber, the military subscriber does not necessarily have a 'home' or fixed point of reference. They could therefore stray and stay well away from where subscriber and location information is maintained. This may lead to poor call set up response times, particularly in a narrow band network, or call failure if the required information is inaccessible. For example, in Figure 2, if user A wishes to call user B then he cannot as although he is in the same fragment as user B, user A cannot contact the remote HLR to find this out. User A and user B may even have entries in the same VLR.

[4] Unmanned Airborne Vehicle
[5] Low Earth Orbit

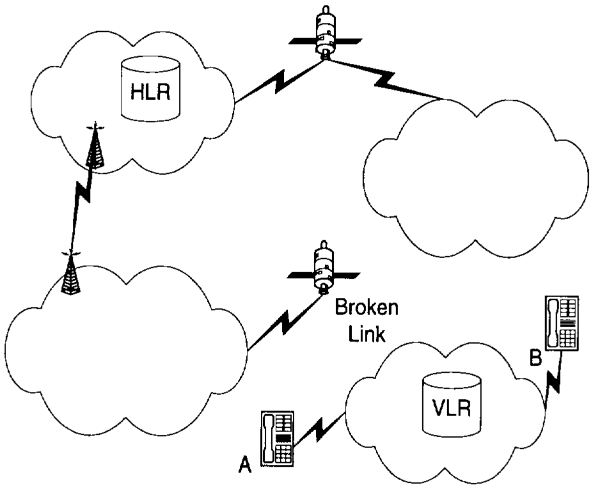


Figure 2; Network Fragmentation Example

The next generation of PCS, UMTS[6], will support both circuit switched and packet switched services, albeit as separate virtual networks. It is highly likely that the GSM HLR / VLR mechanisms will be retained and the approach applied independently .to both the circuit and packet switched virtual networks. This raises the possibility of separate HLR / VLR centres for each service and consequently separate paging / updating for each service.

Internet Protocol (IP)

As with the GSM telephony infrastructure, an IP network needs to be configured with care. This involves the complex business of defining sub-nets, address structures and sub-net masks. This issue is posing a considerable challenge in the use of IP in mobile military systems. The challenge is currently being met by exploiting the topological structure that is inherent in the military command hierarchy. Such a structure may not be as prominent in future, more fluid, operations.

Location management in the IP domain is achieved using the DNS[7] and DHCP[8] standards. Mobility management is handled through the use of the Mobile IP standard.

DNS was created to provide a mechanism for mapping a name to IP address, it was not designed to be updated frequently and often requires manual configuration. The use of DHCP enables the automated allocation of an IP address to a terminal but there are currently no standard mechanisms for informing a DNS server of a DHCP allocation. This is because the IP infrastructure was not designed to support mobile users. To engender routing to a mobile user, they must be allocated (by DHCP) a new IP address consistent with the point of attachment. Updating the DNS with every newly allocated address would be unscaleable as identified earlier.

A DNS entry therefore links a name to a fixed 'home' address. Mobile IP enables user mobility by using this home address as an anchor point. Whenever the user is away from home, he registers with a Foreign Agent (FA) entity on the visited network to obtain a new local address. The FA informs the Home Agent (HA) entity on the users home network that it is now acting on behalf of the user. The HA is then responsible for intercepting packets addressed to the user and 'tunneling' them to the FA where they are forwarded to the user. In GSM terms, the HA is analogous to the HLR and the FA is analogous to the VLR.

With this technique, a home address could be that of a HA service thereby creating a virtual home address. Similarly, the FA service could be hosted by the mobile user's terminal although this can lead to data loss when the mobile user moves, as there is no longer a FA at the original location to perform temporary route extension to the new FA.

The problem with this approach is that it can result in significantly extended non-optimal routes. The Mobile IP mechanisms (HA, FA, and the resulting information flow across a network) are illustrated in Figure 3.

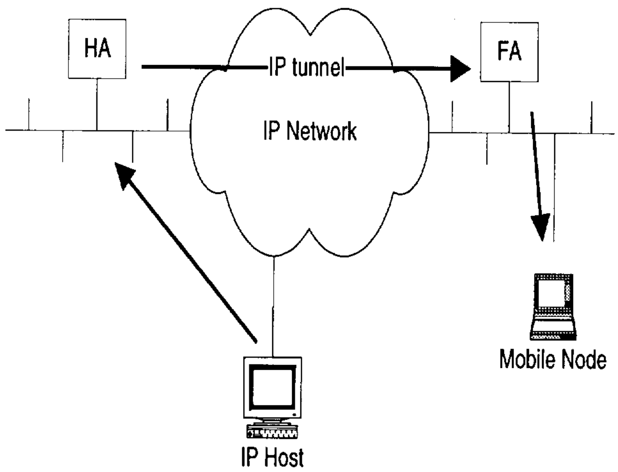


Figure 3; Mobile IP Illustration

[6] Universal Mobile Telecommunications System
[7] Domain Name System
[8] Dynamic Host Configuration Protocol.

Mobile infrastructure is being addressed by the IETF[9] Mobile Ad-hoc NETwork (MANET) working group. MANET is responsible for developing standards to support multimedia services over autonomous ad-hoc networks constructed from mobile routers (and associated hosts). The group is in its early stages but well supported. Much of its initial focus has been on the creation and assessment of new routing techniques. To date there are over ten proposals (Internet drafts). A common theme is the creation of routing communities; detailed routing information is shared within a community but summarised to outsiders to facilitate a scaleable architecture.

Current activity involves detailed modelling to promote better understanding of the proposed techniques and allow cross comparison. The group accepts that no single technique will be suitable for all situations and that a suite of mechanisms may be required.

Presently, each MANET technique proposed has used its own (non-IP) address mechanism. Where connectivity to a fixed backbone is required, this is likely to be via a gateway function that would perform address translation between the MANET deployment and the fixed backbone. The question of scaling has been raised but not addressed directly, the group charter specifies scaling for hundreds of routers.

MANET like systems are already present in the military domain, packet radio is an example. They tend to be limited to the support of non-real time traffic and current algorithms do not readily scale to large networks.

Asynchronous Transfer Mode (ATM)

The ATM PNNI[10] protocol defines mechanisms for exchanging topology and resource information between switches and clusters of switches. Support for large scale networks is possible because PNNI is a hierarchical protocol where the address structure can reflect the topology thereby allowing the topology information to be aggregated. PNNI also allows address exceptions to be advertised although excessive usage of exceptions significantly increases bandwidth overheads because address aggregation becomes difficult. Peer groups are an important aspect of PNNI, routing information is propagated in detail within a peer group and summarised to other peer groups.

To make a large network scaleable, a peer group structure is required analogous to IP sub-nets but with more flexibility. As with IP, the determination of this structure is a complex configuration task placed on the management system. A very fluid infrastructure therefore poses a significant challenge.

Mobility issues are also arising in the development of ATM. Like IP, ATM has location management mechanisms in the form of ANS[11] supported by automatic address allocation. Similarly to IP, the ANS architecture could be used to support the nomadic user but the same scaleability issues apply. The approach favoured by WATM[12] is to hide end user mobility from the ANS through the use of intermediate functional nodes akin to the VLRs within GSM.

Support for mobile infrastructure to allow ATM satellite switching services are included in the WATM remit but standards and means of supporting them have yet to be developed. This will not be a trivial task due to the complexity of handing over multiple calls simultaneously without service interruption. Aspirations to ad-hoc ATM networking exist but the work has not been started. The WATM group is scheduled to produce a draft specification by December 1999, currently no date is set for ratification of the specification.

Closing the Gap

One of the principle issues is to determine where, within the protocol layers, mobility is to be managed. A simple model of logical mappings is shown in Figure 4. So far, we have focused on location management, the mapping between name and address. No specific transport mechanism or set of protocols is assumed. Indeed the preferred approach is that of the Intelligent Network where mobility management and other core network functionality is protocol independent. The figure also shows a mapping between name and subscriber profile. This profile could contain security information for authentication, access privileges, contracted tariff schemes etc.

[9] Internet Engineering Task Force
[10] Private Network-Network Interface
[11] ATM Name Serving
[12] Wireless ATM, an ATM Forum Working Group

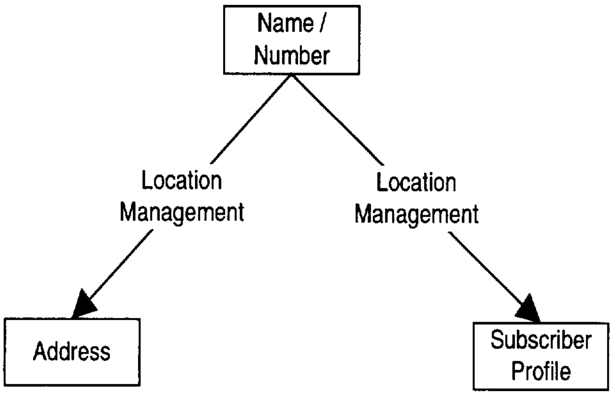


Figure 4; Simple Mobility Model

The subscriber profile is needed by the network to respond to requests for subscriber connection/affiliation. Placing this information where the subscriber is most likely to connect to the network could improve response time and the probability that the information is available even when the network is fragmented. Therefore this information should follow the subscriber around. The movement of subscriber profile A in Figure 5 illustrates this.

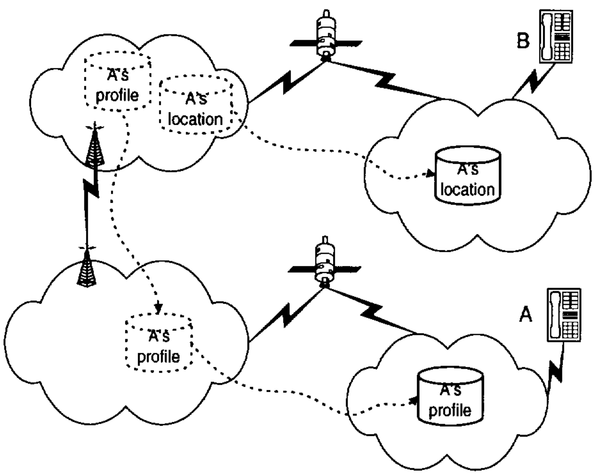


Figure 5; Distribution of Subscriber Information

The second issue is that fixed location registers cause a problem both in terms of resilience and users who have no home location. Positioning location management information where it is most likely to be needed may mitigate this. The movement of location management information is also shown in Figure 5. The term subscriber information shall be used to refer to the combination of address and profile details.

In a network where subscriber mobility and call behaviour is random, the positioning of subscriber information may as well be arbitrary. However, in the military and civil domains, subscriber behaviour often has patterns. Civil mobility and call patterns tend to follow cycles throughout the day and week. Military patterns can include regular movement of logistic assets along supply routes or the co-ordinated movement of many subscribers within a manoeuvre. The former may be predictable as a pattern develops, the latter may be predictable simply because it is a planned. Therefore predictive techniques may be based on a mix of historical behaviour and plans to aid the positioning of subscriber information.

The combination of these techniques therefore leads to subscriber information migrating around the network. The rate of location information migration may be derived from the network configuration and need not match the actual rate of movement of the subscribers. Using these techniques, locating the required information becomes an issue. Pointers can be used to direct requests to the current location of subscriber information. The update rate of these pointers may be another variable, and they themselves, can be located by yet more pointers. Scaleability will place limits on how elaborate this pointer system is.

Our research is focused on developing algorithms for distributing the subscriber information around the network. In addition, replication will be needed to achieve resilience requirements. The depth of replication will need to be balanced against the overhead produced by maintaining the information.

Intelligent caching of subscriber information can improve resilience and responsiveness. The trade-off is between the overheads saved by not performing the look-up and the consequences of finding that the information is invalid. By giving cached information a lifetime and ageing it, the retention of invalid information can be ameliorated to a degree. Additionally, the age of the cached information can be used to calculate a confidence likelihood for its validity.

These techniques could tolerate infrastructure mobility. Unplanned or unpredictable infrastructure mobility would need to rely on the resilience of the location management and mobility management systems. Planned infrastructure mobility could be used to notify the information distribution processes described above to move the information centres as required.

Finally, as in GSM, location areas provide a useful method for balancing the frequency of location update against paging attempts to search for a subscriber. These location areas, however, need to be carefully defined. In particular, network topology and logical structure will influence the capability of the network to efficiently route calls, or packets. Logical structure is used here to describe such attributes as peer grouping, sub-nets, masks, address prefixes, etc., depending on the protocols used. This logical structure will also need to be defined. For networks with some persistence in structure this definition and configuration task is within the managers capability. For fluid networks this may be a too large a burden and automated facilities may be required.

Therefore our research is also focused on those algorithms that can determine the most suitable logical network boundaries, and adapt these boundaries to the topology as it evolves. Effecting these changes will, in themselves, appear as mobility and will need to be tracked by the location management system. These algorithms will need to take into account numerous factors including the richness of network meshing and link characteristics such as total capacity, available capacity, error characteristics, etc.

Conclusions

We have examined the mechanisms used to manage mobility in some detail for three commercial communications standards. The initial conclusion is that whilst user mobility is supported well, mechanisms to support mobile infrastructures are in their infancy. Additionally, it can be seen that there is much commonality in the way that user mobility is supported, and the issue of scaleability has been comfortably demonstrated.

To close the gap between current capability and future aspirations to a fully mobile infrastructure, we have proposed a set of techniques, derived from current capabilities, to support the necessary resilience and responsiveness. Current work is concentrated on the analysis, development and demonstration of these techniques.

E-mail: rdbarfoot@dera.gov.uk
jmdaniell@dera.gov.uk

URL: www.dera.gov.uk

Tactical Radio Access Networks Based on Future Civilian Wireless Communication Architectures

Corresponding author: Jan Nilsson
Co-authors: Eva Englund, Jimmi Grönkvist, Anders Hansson, Mattias Sköld, Ulf Sterner, Ingrid Söderquist, Jens Zander
Sweden's Defence Research Establishment (FOA 72)
BOX 1165, S-581 11 Linköping, Sweden,
Tel: +46 13 378133, Fax: +46 13 378049
E-mail: jann@lin.foa.se

SUMMARY – We investigate architectures where a tactical radio subsystem can integrate in civilian systems. The idea is that civilian communication infrastructure should be used wherever it is available. One such possibility may be found in the proposed next generation wireless communication system UMTS, where the definition of a Radio Network Subsystem (RNS), encapsulating all the features of wireless mobile communication has been proposed. We adopt this design paradigm to see if it is possible to find a low cost military network (by maximal use of civilian network technology and communication applications). The idea is to within the framework of an RNS, encapsulate also features specific to tactical wireless communications, a Tactical Radio Access Network (TRAN). Based on UMTS we describe two system concepts and architectures for designing a TRAN. Furthermore, a feasibility investigation of the concepts is conducted.

1 Introduction

In the battlefield of the future, tremendous amounts of information will be transported and processed at an unprecedented speed. More and more combat information and sensor data is available at all levels of the military hierarchy, guiding vehicle commanders and even individual soldiers in combat in much more detail than before. In addition, the mobile information system projected in most studies is designed to allow low level units to make informed, autonomous decisions in fast moving combat situations. Lightweight, reliable and inexpensive equipment based on civilian mobile computing and communication technology, is expected to have wide spread military use. Although civilian technology will have a major impact on future military communications, there are crucial situations where military requirements conflict to those of civilian systems and where special solutions must be sought.

A natural strategy in designing tactical communication systems is to use civilian technology and architecture to as large extent as possible and to encapsulate all features specific to tactical wireless communications in a few, well defined modules. In Europe, the Universal Mobile Telecommunication systems (UMTS), is currently being defined and standardized.

In the UMTS architecture the concept of a Radio Network Subsystem (RNS) has been adopted. The purpose is to make different access networks, wireless or fixed, look indistinguishable from the fixed backbone, or core network. One example of a future RNS is the new radio access network known as UMTS Terrestrial Radio Access Network (UTRAN). Our aim is to suggest architectures where the RNS framework can be adopted to encapsulate also the features of tactical wireless communication. Such an architecture will open up the possibility to integrate a tactical radio subsystem in civilian communication infrastructure.

We describe two system concepts and architectures for a tactical radio access network (TRAN). The first, called a distributed TRAN, aims to fulfill tactical requirements to as large extent as possible within the RNS framework. The second concept, on the other hand, aims towards, with a minimum of modifications, extending the UTRAN concept to fulfill the most essential tactical requirements. The idea is to have two competing concepts, one representing a true tactical radio, still compatible with UMTS infrastructure, and one representing a cost efficient commercial of-the-shelf (COTS) solutions using as much of the UMTS technology as possible.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

2 Third generation wireless system

The UMTS architecture, in phase 1, is illustrated in Fig. 1. The radio specific functionalities are encompassed in the network subsystem. A network subsystem domain may consist of existing mobile radio technologies as well as future technologies. Examples of access networks are: GSM Base Station Sub-System (GSM BSS), Satellite Personal Communications Network (Satellite), Broadband Access Network (BRAN), UTRAN, and the proposed TRAN. As an example of a network subsystem consider the GSM BSS, it consists of many base stations and a base station controller, as well as user equipment.

The Core Network (CN), can of course also consist of several networks: GSM and GPRS Core network, ISDN, TCP/IP based network, and others as well as future CN technologies. In the first phase, the UTRAN will use a GSM and GPRS core network. The interface between a future UTRAN and CN, denoted Iu, is to be defined in the UMTS standardisation. The same interface Iu is also assumed for TRAN. The keep all the information about the users, and to route messages to the destination are main tasks of the CN.

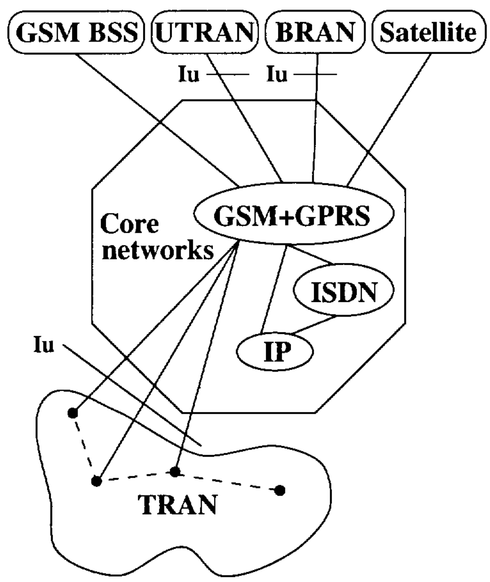


Figure 1: Simplified UMTS architecture, phase one

3 Requirements

Tactical requirements demand that the TRAN can be flexible to different types of command and control structures. Furthermore, a TRAN must be able to be autonomous, that is, it can be deployed in a remote area where no core network is available. To secure robust command and control the network administration should be distributed to as large extent as possible. If all infrastructure is destroyed (or simply lacking) communication between units within good radio coverage should be possible. Also, the TRAN should be able to have many and redundant access points, so called Gateways (GW), to the CN, see Fig.1.

Mobility is another critical issue, the network must be able to handle, and ensure communication with fast moving communication platforms. A flexible network should not only support point-to-point communication but also efficient broadcasting and multicasting. Furthermore, a TRAN should be able to establish high priority logical channels, for example by using several physical connections (macro diversity) and flexible routing methods. It is assumed that the enemy has full knowledge about the system, except cryptographic keys. Nodes can be destroyed, links can be jammed, and equipment can be taken by the enemy.

A TRAN could embrace everything from a small group of soldiers to some joint brigades. The first could be deployed in an area of less than 5 kilometers radius, and the second in an area of say, 100 x 100 kilometers. The communication demands will vary considerably and the TRAN must be able to handle applications with different quality of service (QoS) requirements. The requirements include real-time transmission for voice and video, and non real-time transmission for large data files. In many presumed combat situations one can foresee the need for extremely reliable and fast communications. Therefore, the presumed TRAN should be equipped with a priority system, in order to guarantee extreme reliability and/or speed for especially urgent and vital communications.

To use many concurrent networks, each designed with a special application in mind, is one solution. Communication between networks could then be handled by special gateways. However, the TRAN we have in mind, can be considered as one network, at least in terms of equipment. One multimedia radio terminal must be able to handle all types of traffic. A large number of platforms, with communication needs, will be present on the future battlefield. Examples of platforms we have in mind are: vehicles of different types, soldiers, sensors, and helicopters. These platforms, or communication nodes, all have interfaces to user applications, and are at least capable of storing and forwarding information packets. Satellite communication is possible, but then we assume by using civilian systems and equipment.
In conclusion, without modifications, a commercial cellular system, such as GSM, cannot be used. A

TRAN must be UMTS/UTRAN compatible on an Iu interface level. For more information see [1].

4 TRAN concepts

To design a radio network fulfilling the above requirements is a true challenge, especially with the ambition that the network should integrate with civilian infrastructure, a structure that in many cases conflict with tactical requirements. Here, we describe two strategies, or concepts, of designing a TRAN. The idea is to have two competing concepts, one representing a true tactical radio, still compatible with the UMTS infrastructure, and one representing a cost efficient COTS solutions using as much of the UMTS technology as possible, see Fig. 2.

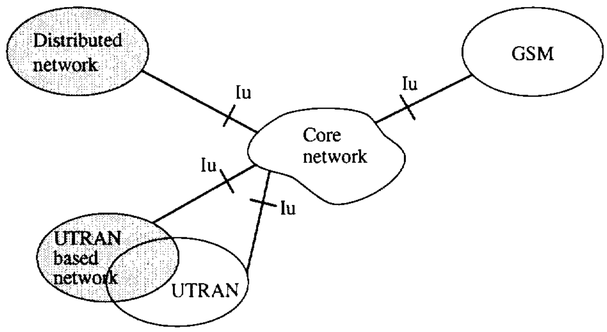


Figure 2: TRAN concepts (colored grey)

Concept 1: Distributed network

In the first concept, we primarily focus on fulfilling the tactical requirements described above to as large extent as possible. What we suggest here is a distributed multihop network architecture. Such a network can work without, or at least with a minimum of network planning. Network management functions can be distributed and are therefore robust against hostile attacks. Traffic can quickly be re-routed to handle topological changes. Urgent messages can be sent on several paths simultaneously to ensure high reliability.

We assume two types of radio units, standard nodes and Connection Control Nodes (CCN). The CCN are to maintain and distribute access schemes and routing tables, whereas the actual traffic can be relayed through all nodes with respect to some routing method. How to select a CCN based on, connectivity, computational capacity, and low mobility etc. is a still open question. The CCN may be a vehicle. A soldier with a small hand-held radio or a sensor may be a standard node. However, it is desirable that any node can act as an CCN. Furthermore, it is also desirable that any node can act as a GW, for connection to the CN.

The main reason to use two types of nodes, or hierarchical layers, is that the fully distributed management of network functions may be very difficult and cause a lot of overhead traffic in large networks. Whenever a flat network is large (say more than 100 nodes), the delay can be very long. Furthermore, a lot of processing is required in every node to handle routing, mobility management, and channel resource allocation. The amount of processing is reduced, in the standard nodes, by having a second layer, i.e., CCN:s. Also, different nodes, naturally have different communication and processing capabilities.

For channel resource allocation we investigate Spatial Time Division Multiple Access (STDMA). STDMA is a TDMA protocol, where nodes that are separated geographically can transmit simultaneously in the same time slots. An example of a STDMA protocol that also depends on the signal to interference ratio is given in [5]. Key problems are: to find efficient distributed algorithms for STDMA scheduling, slot synchronization and to handle mobility. STDMA is very efficient for stationary networks [5, 7], but a problem is that when the mobility is high an efficient STDMA protocol must be updated frequently [4].

Concept 2: UTRAN based network

The second concept aims to use as much of the UTRAN architecture as possible, introducing as few modifications as possible to fulfill the most essential tactical requirements.

The UTRAN consists of Radio Network Controllers (RNC's), base stations and user equipments. An RNC controls one, or several base stations. Also, it is the RNC which handles the connection to the CN by the Iu interface [2]. Examples of functionalities fully contained in the UTRAN are: radio channel ciphering/deciphering; radio bearer control; radio channel coding/decoding and control; initial access detection; power control/setting; data packeting; and radio channel reservation/allocation. Furthermore, all cell-level and cell structure mobility handling is contained within the UTRAN. The handover control functions are located in the UTRAN (RNC) and their execution and completion is located both in the UTRAN and the CN. The Macro-diversity function is typically located in the UTRAN (RNC).

The connection points to the CN are vulnerable parts and may be destroyed or lost, also in some cases, civilian infrastructure may be lacking altogether. Therefore autonomy with respect to the CN is one of the most essential tactical requirements in a TRAN. However, tactical requirements for robustness against hostile attacks and flexibility to different types of command and control structures lead naturally to a demand for autonomy also internally in the TRAN. To fulfill this requirement in the UTRAN based TRAN concept may be difficult. The concept is hierarchical, each level depends on higher levels. For autonomy each layer must be equipped with an ability to emulate higher levels if these are lacking. Full autonomy may have to be renounced.

Full mobility, and communication on the move, are also important tactical requirements. In UTRAN the mobility is handled by the base stations with handover routines. However, in some cases the base stations themselves might need to be mobile.

5 Concept feasibility study

Here we present a methodology to asses the feasibility of the concepts. The aim is to judge if a concept has the possibility to fulfill the requirements, and if so, if it is feasible in terms of practical realization. The main goal is to assess the following fundamental issues:

Issue 1: Can the service requirements be fulfilled?
Issue 2: Can the robustness requirements be fulfilled? This issue includes, jamming robustness, redundancy if vital units are destroyed, and good area coverage.

Another issue not treated further here is: how difficult and costly will it be to fulfill the compatibility requirements? Both concepts have to be tested towards these requirements.
A distributed multihop network may have problems providing sufficiently high user data rates and short transmission delays, due to poor links because of no network planning. On the other hand, the UTRAN based concept may be vulnerable, due to its hierarchical architecture. Therefore, we judge that Issue 1 is the most critical for the distributed multihop concept, and Issue 2 for the UTRAN based concept.

In the next part we develop Issue 1 and test it for the multihop network. Afterwards, aspects concerning Issue 2 is treated for the UTRAN based concept. Without sacrificing the relevancy, we try to simplify the situation as much as possible. Otherwise, the analyses will be very difficult, require extensive simulations, and be difficult to interpret. A main problem is that, routing, access, and traffic handling cannot be treated separately. These functions depend on each other and a joint methodology has to be adopted.

5.1 Distributed network

We assume ideal mobility management. That is, the network control protocols adapt them-selves ideally to the current situation, and the cost of sending information over the network for such an adaption is not considered. Thus, an optimistic estimation of the network performance at a given time (snap shot) is obtained. If not even such an optimistic estimate fulfills the requirements, another solution must be sought.

To assess the radio network, a tactical unit, and a communication (or traffic) model are defined. The unit is spread-out in the terrain and possible links and their capacities are estimated. Thereafter the communication traffic is routed, and the actual user data rates are estimated. In this process, an access protocol is also used. Finally, based on the obtained radio network it is possible to assess what communication that can be accommodated. We have the following goals.

1) Estimate the possible link data rates in the network.
2) Estimate the available bit rates for each connection, i.e., Origin-Destination (OD) pair.
3) Determine the services that can be provided. In particular, what is the time availability of the services?

Tactical unit

The tactical unit considered is a traditional mechanised battalion aimed for rural areas. This unit is simplified to consist of one communication platform only, a vehicle. Roughly, a battalion consists of 6 companies, e.g., four tank companies, one artillery company, and one pioneer (or support) company. If we assume each company contains 12 vehicles we get altogether 72 communication platforms in a battalion.

These 72 platforms are placed randomly within an area of either 20x20 km (large area network), or 3x3 km (small area network), according to an uniform distribution. Thus, no network planning is used. All 72 platforms are considered to be CCN nodes and no hierarchy is introduced. Smaller nodes, which could be soldiers or sensors, are not included.

Possible link data rates

To estimate the link data rates, assumptions about the communication platforms and the propagation environment are required. We consider the following two alternatives.

Carrier frequency $f_c = 450$ MHz: antenna height $h = 3$ m; transmit power $P_t = 10$ W; antenna gain $G = 3$ dB; receiver noise factor $F = 10$ dB; bandwidth $B = 45$ MHz.
Carrier frequency $f_c = 2$ GHz: antenna height 3m; transmit power $P_t = 10$ W; antenna gain $G = 10$ dB; receiver noise factor $F = 10$ dB; bandwidth $B = 200$ MHz.

Based on these assumptions possible links can be estimated. The radio network is modeled as a set of platforms, called nodes, V and the *link* (i, j) between any two nodes (v_i, v_j), for $i = 1, 2, ..., N$, and $j = 1, 2, ..., N$. To asses the network, firstly the possible data rates $\tilde{R}(i, j)$ between any two nodes are estimated.

We use an advanced electromagnetic propagation model, based on Vogler's knife edge diffraction model [3, 7], to estimate the path loss L_{ij} between nodes i and j. To model the terrain a digitized map is used. Thereafter, the data rate $\tilde{R}_{ij}$ of the link (i, j) is estimated as

$$\tilde{R}_{ij} = \frac{1}{D} B \log_2(1 + \frac{P_t G^2}{L_{ij}(BFKT_0 + I_{ij})}); \qquad (1)$$

$(1 \leq i, j \leq N)$, where K and T_0 are given constants. We also include a diversity factor D set to 10. Then at least some basic jamming protection is provided. By the diversity it is also practically feasible to handle the expected multipath propagation. In (1) an interference margin I_{ij} is also included. It is set to be equal to zero when possible link data rates are estimated. However, it is needed when the channel resource is assigned by the access protocol and the user data rate is estimated. When transmission on a link is considered, transmission simultaneously on an adjacent link cause interference and reduced data rate on the considered link.

In Fig.3 possible link data rates are illustrated for a rather difficult terrain in terms of radio communication. The terrain chosen, south of Linköping, Sweden, is hilly with mixed meadows and forest. A large area network (20x20 km) and f_c =2 GHz is tested. The poorest link in the connected network has a bit rate of 42 Kbit/s. Links where not even 42 Kbit/s is possible, are not shown.

Communication model and routing

We assume a uniform broadcast traffic model. That is, every node transmits situation awareness data, as frequent as possible, to all other nodes. The data can contain positions, movements and other mission data.

The uniform traffic model implies that all node-

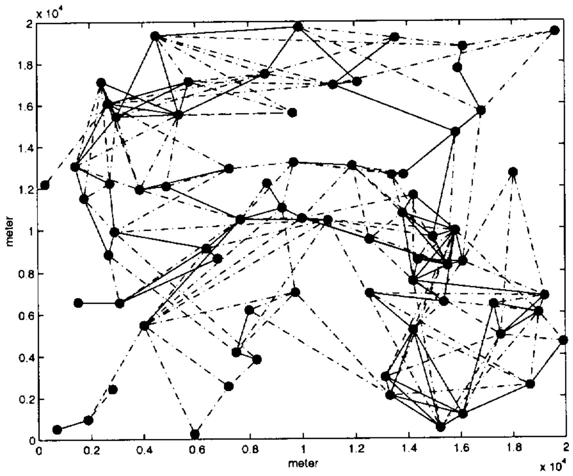


Figure 3: Example of possible link data rates; rates between 42 Kbit/s - 2Mbit/s: dash-dotted line; rates over 2Mbit/s: solid line.

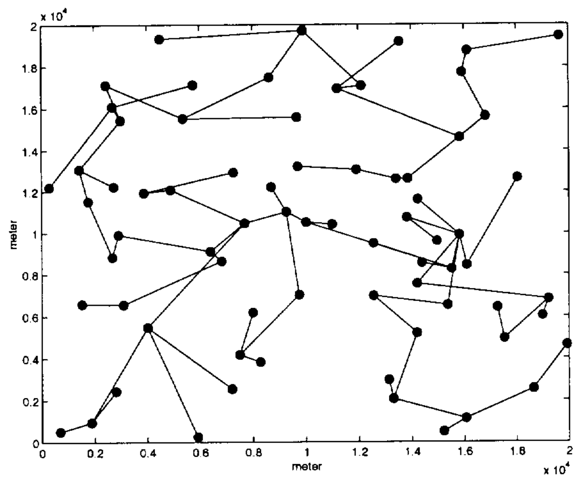


Figure 4: Example of spanning tree based routes for the communication model, compare with Fig.3

pairs (i, j) (OD-pairs) communicate with each other. However, multicasting can be utilized since every node sends the same data to all nodes [10]. Here, for simplicity, no delay constraints are included. Therefore, all routes can be obtained from one spanning tree only [9, p.2]. Fig.4 shows the spanning tree, maximizing the sum of data rates, over all links in the tree. The situation from Fig.3, is illustrated.

Data rates for user connections (OD-pairs)

Based on the possible link rates, and the traffic flow on each path obtained after the routing, the channel resource is assigned by the STDMA protocol devised by Grönkvist [5]. Notice that transmission on a link is only possible when it does not interfere to much with another link transmission. Thus, the actual data rate on a link, may be much lower than the possible link rate. The fraction of time, the link (i, j) is used for transmission, is denoted τ_{ij}. Thus, the actual used data rate is

$$R_{ij} = \tau_{ij} \tilde{R}_{ij}.$$

In this way we get an used radio network defined by a set nodes, V and data rates $R(i, j)$ on used links. Achievable data rates for every OD-pair can thereafter be obtained. Let us point out that an advanced traffic controlled STDMA algorithm is used, where τ_{ij} is assigned according to the relative traffic load. The communication in the network will be to a large extent equalized, by that STDMA algorithm. That is, all OD-pairs will get fairly similar data rates.

We have tested several different types of network cases, i.e., different carrier frequencies, different network size and different terrain. In Fig.5 to Fig.8 four of those cases are plotted in a histogram. Each of the four histograms is generated based on 100 different networks. For each network, it is the OD-connection with the minimum data rate that is given. The terrain chosen in all cases, south of Linköping, Sweden, is hilly with mixed meadows and forest. It was the most difficult terrain of those tested.

As we can see from the figures, spreading out the platforms over the larger area (20x20 km instead of 3x3 km) reduces the OD data rates about 20-50 times. At the high frequency, the used bandwidth is much higher. Therefore, for the small area network, using carrier frequency 2 GHz is clearly better than using carrier frequency 450 MHz. On the other hand, at 450 MHz the propagation properties over difficult terrain is much better. This fact, compensates for the smaller used bandwidth in the large area network case.

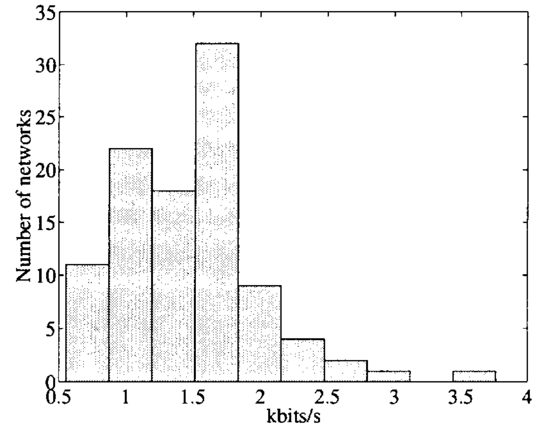


Figure 5: Minimum data rate for OD-connections; large area network (20x20 km) and 2 GHz carrier frequency.

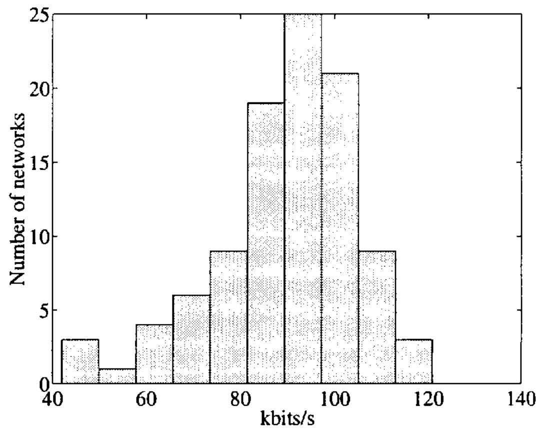


Figure 6: Minimum data rate for OD-connections; small area network (3x3 km) and 2 GHz carrier frequency.

Some remarks

In a high mobility combat situation, e.g., an attack, the unit will probably not be spread out over a 20x20 km area. It is plausible, that it is for the small area networks, the mobility can be the highest. On the other hand, for those networks substantial higher data rates can be provided than for the large area networks, and this will give us a mobility margin. That is, data rates will be reduced due to practically realizable (non-ideal) mobility management.

It is the vital links of poor quality that largely determines the network capacity. Therefore, some network planning would improve the situation considerably, in particular when the network is large. No delay constraint is included in our evaluation. For the large network the transmission delay between OD-pairs can be several seconds (many hops). This can be avoided by delay constraint routing. However, this in turn, results in a higher traffic load. Fortunately, using just a few more links will remove the real long delays (many hops) and this will cost very little in terms of additional traffic load, compared to the example in Fig.4.

5.2 Tactical UTRAN based network

The UTRAN is designed for multimedia applications, and the traffic flexibility requirement is acceptably fulfilled, at least as long as the development proceeds as expected. However, the problem with UTRAN

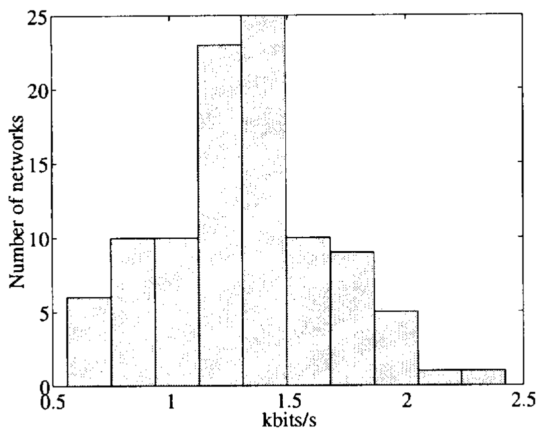


Figure 7: Minimum data rate for OD-connections; large area network (20x20 km) and 450 MHz carrier frequency.

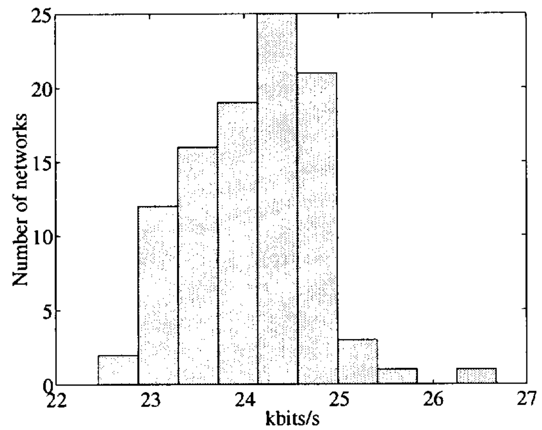


Figure 8: Minimum data rate for OD-connections; small area network (3x3 km) and 450 MHz carrier frequency.

is its lack of military robustness, and possibility to ensure good area coverage in remote areas (the second issue). It is vital to find good solutions to compensate for these deficiencies as far as possible.

The main robustness issues that the TRAN concept have identified are: autonomy, security and robustness against jamming. To fulfill the tactical requirements for autonomy, it is desirable to first of all extend the functionality of the UTRAN and include autonomous base stations. This topic will be treated below. Other desirable extensions are, relay capability between units within range of a base station and peer-to-peer communication between units without contact via a base station. Relay capability would considerably increase the coverage of the network, as it can be used to ensure radio coverage for users in a difficult environment. Peer-to-peer communication is judged to be more complicated than introducing a relay capability, as this needs the support of a base station or a central database.

Due to the centralized control and management of the civilian communication infrastructure, the UMTS architecture and consequently the UTRAN does not fulfill the autonomy requirement. The procedure for call set-up requires successful signaling through the Core Network, that is, without the CN there will be no traffic. Jamming, meaning signaling with the explicit purpose of disrupting the communication is not a threat in civilian systems. Accordingly, the air-interface is not designed to be robust against that threat. However, in a single-user case, the used 4 MHz wideband DS-CDMA waveform in UTRAN, is at least in itself fairly robust against jamming. Further investigations are necessary, to be able to judge whether the air-interface has to be modified.

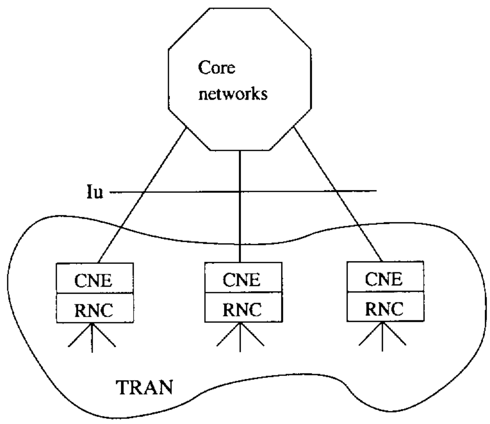


Figure 9: Additional functionality in the UTRAN based concept to allow for autonomy.

Autonomous base stations

To make the base stations autonomous one must identify key functions, normally facilitated by the CN, and emulate these in the base stations, by what we call a Core Network Emulator (CNE).

The CN has a hierarchic structure, and keeps all information about the users. All the user information that is sent from TRAN to the CN must be kept on a local level also (or perhaps not at all forwarded to a central node in the CN). Some of the information in the CN must also be kept in the CNE. If the connection to the CN is broken, the CNE is to have full knowledge about who the reachable users are, and be able to identify legitimate users in the TRAN.

How to make the TRAN network autonomous, is illustrated in Fig. 9. As long as the CNE/RNC is connected to the CN, the CNE is transparent, and only monitors the traffic in order to update its databases. When the connection is broken, the CNE has to substitute the CN. Note though, that the functionality of the CN can be simplified when implemented in the CNE. In the autonomous mode, the CNE must only be able to handle the units within range from the RNC, and incoming military units. These units are quite few compared to the number of terminals handled by the registers in the CN. The functionality included (in terms of the GSM standard) in the CNE, in order to make the TRAN autonomous is:

- EIR, Equipment Identity Register, provides identification so that lost, or stolen, equipment can be identified when they attempt access.
- AUC, Authentication Register, authenticates the users. Usually integrated with the HLR.
- HLR and VLR, Home Location Register and Visitor Location Register. HLR is a database that manages the records of the users. It also control services associated with incoming traffic. The VLR has a temporary record of users that are not included in the HLR.
- MSC, Mobile Switching Centre. A switching node, responsible for the coordination of mobility and the initiator of handover between location areas.

5.3 Conclusions

The feasibility study aims to test whether our concepts can fulfill the requirements. If not even by an optimistic evaluation, a concept works, another solution has to be sought.

Whether a distributed network can fulfill the service requirements is a central issue. We have tested a number of different situations, and for all those situations a basic service can be provided. That is, situation awareness data can constantly be transmitted between OD-pairs. About 120 bit/s is enough (a 240 bits packet per two second) according to an estimation made in [7, 8]. For the cases we have tested, at least 0.4Kbit/s can be provided for all OD-pairs. However, the possible data rates will be much higher in most situations. Notice that this means that all nodes can communicate with all other nodes simultaneously at such rates. Therefore, using a distributed multihop network solution, even without network planning, is possible. At least based on our investigation, we cannot discard such a solution due to bad connectivity and low user data rates.

For the UTRAN based network the robustness issue is vital. In particular, autonomy and good area coverage have to be provided. We believe that improving the robustness as described earlier is possible to realize. An issue is, however, if the concept still after all the desired modifications, constitute a cost efficient alternative as compared to the distributed network concept.

6 Further work

At this stage, our concept study focus primarily on network control functions, and the ability to provide good, robust and secure bearers for the communication. Then, services and applications can be adapted after the future demands. However, in a later stage all network levels need to be addressed. Compatibility issues with other networks, and future CN technologies, are important.

The aim is to continue to study these two concepts in more detail during the next year, and thereafter to be able to decide on a main TRAN architecture to develop further. However, taking that decision involves a lot of issues, first of all, technical ones, but the cost of realizing the concept is also a very important issue.

References

[1] E. Englund, et.al., *Tactical radio Access Networks - A Concept Study; Report 1,* FOA-report, FOA-R-98-00940-504-SE, Sweden, Dec. 1998.

[2] ETSI SMG2, *The ETSI UMTS Terrestrial Radio Access (UTRA),* ITU-R RTT Candidate Submission, http://www.etsi.fr, Jan. 1998.

[3] B. Asp and G. Eriksson and P. Holm, *Detvag-90 — Final Report,* Defence Research Est., Div. of Command and Control Warfare Tech., FOA-R-97-00566-504-SE, Linköping, Sweden, Sept. 1997,

[4] U. Sterner, *Mobility models for mobile radio networks,* Master's thesis nr: LiTH-ISY-EX-1975, Linköping University, S-581 83 Linköping (In Swedish), Sweden, Dec. 1998.

[5] J. Grönkvist, *Traffic controlled spatial reuse TDMA for multi-hop radio networks,* Proc. IEEE PIMRC'98, vol.3, pp.1203-1207, Boston, U.S.A, Sept. 1998.

[6] C. Jönsson, et.al., *Modeling terrestrial mobile networks in real terrain,* Proc: IST Symposium on Tactical Mobile Communications, Lillehammer, Norway, June 1999.

[7] C. Jönsson, *The cost of transmitting position data,* personal communication, chj@lin.foa.se.

[8] J. Kjellgren, et.al., *Teknisk-strategisk studie av digital slagfältet, Delrapport 4 - Igenkänning och positionsrapportering,* FOA-report (In Swedish), FOA-R-98-00903-504-SE, Sweden, Nov. 1998.

[9] B. Bolloba's, *Graph-Theory, An Introductory Course,* Springer-Verlag, New York 1979.

[10] K. Makki, N. Pissinou, and O. Frieder, *Efficient solutions to multicast routing in communication networks,* In: Mobile Networks and Applications 1, pp.221-231, J.C.Baltzer AG, Science Publishers, 1996.

Energy-Efficient Wireless Networking Techniques:
Fundamental Issues and an Application to Multicasting in Ad Hoc Networks

(June 1999)

Jeffrey E. Wieselthier
Information Technology Division
Code 5521
Naval Research Laboratory
Washington, DC 20375
wieselthier @ itd.nrl.navy.mil

Gam D. Nguyen
Information Technology Division
Code 5521
Naval Research Laboratory
Washington, DC 20375
nguyen @ itd.nrl.navy.mil

Anthony Ephremides
Electrical Engineering Dept. and
Institute of Systems Research
University of Maryland
College Park, MD 20742
tony @ eng.umd.edu

Abstract

Energy is a scarce resource in military ad hoc networks. In this paper we address the fundamental research issues associated with energy-efficient networking, and we apply energy-efficient networking techniques to the problem of multicasting. We discuss the impact of the wireless medium on the multicasting problem and the fundamental trade-offs that arise. We propose and evaluate several algorithms for defining multicast trees when transceiver resources are limited. The algorithms select the relay nodes and the corresponding transmission power levels, and achieve different degrees of scalability and performance. Our performance results demonstrate that the incorporation of energy considerations into multicast algorithms can, indeed, result in improved energy efficiency.

1 INTRODUCTION

In this basic research study, we consider the problem of energy-efficient networking in ad hoc (i.e., non-cellular, or infrastructureless) wireless networks. This problem arises in applications such as the Digital Battlefield and Littoral Warfare, as well as in other peer-to-peer wireless architectures. The intent is to maximize communication performance subject to a limit on the available energy or, alternatively, to minimize energy use subject to meeting given Quality-of-Service (QoS) performance requirements. After a discussion of the fundamental research issues that arise in this communication environment, we apply energy-efficient networking techniques to the problem of multicasting. This study is perhaps the first attempt to address the energy-related issues that arise in wireless multicasting.

Although energy conservation is clearly important in mobile communications of any kind, it is critical in military networks that consist (solely or partially) of "light" mobile users (such as individual warriors). In such networks, operation is ultimately limited by the constraint of finite battery life at the individual users because it may be impossible to recharge batteries during the course of a mission. Thus, there is a need to develop networking techniques that make efficient use of the limited energy that is available. A novel feature of this study is that instead of viewing energy efficiency from the perspective of low-power equipment or highly efficient batteries, we address it as a network design problem. Issues under study include choice of power level, choice of channel-access protocol, the design of asymmetrical protocols that minimize the energy expenditure of light users, the impact of relaying (i.e., the consumption of energy to support communication needs of others), and trade-offs between signal processing and communications.

In view of these issues, and the interrelationships among them, we argue that it may be necessary to abandon the traditional layered network architecture in favor of new approaches that permit the vertical coupling of protocol layer functionality, thereby permitting improved energy efficiency; e.g., the routing algorithm should be coordinated with the choice of transmitter power levels because the latter determine the connectivities that are available to the former.

We have chosen the problem of multicasting (one-to-many communication) as the focus of our energy-efficient networking studies. Multicasting in wireless networks is fundamentally different from multicasting in "wired" or "tethered" networks. In addition to node mobility (and, hence, variable connectivity in the network), there are additional trade-offs between the "reach" of wireless transmission (namely the simultaneous reception by many nodes of a transmitted message) and the resulting interference by that transmission. We assume that the power level of a transmission can be chosen within a given range of values. Therefore, there is a trade-off between reaching more nodes in a single hop by using higher power (but at a higher interference cost) versus reaching fewer nodes in that single hop by using lower power (but at a lower interference cost). By contrast, the unicast (one-to-one) communication problem (although challenging in its own right) is characterized by a minimum-energy solution in which multihop relaying at low power is generally favored over higher-power transmissions because of the nonlinear attenuation properties of radio signals. Such generalizations cannot be made for the multicasting problem, however. In some of our examples it is better to transmit at low power, whereas in other cases high power is better.

This work was supported by the Office of Naval Research.

Few studies have addressed the crucial problem of multicasting in wireless networks. For example, the problem of multicast scheduling in cellular mobile networks was studied in [1], and a forwarding multicast protocol for noncellular networks was studied in [2]. Virtually all multicasting studies have been limited to the case of stationary networks that are not wireless (e.g., [3], [4], [5]). Since ad hoc networks lack a fixed cellular infrastructure, they cannot effectively use multicast algorithms that are based on fixed topologies.

In this paper, we develop several algorithms for multicasting and compare their performance. We evaluate the trade-offs between algorithm complexity (and hence scalability) and performance. Our focus is on the source-initiated multicasting of "session" or connection-oriented traffic. To assess the complex trade-offs one at a time, we assume in this paper that there is no mobility.[1] Nevertheless, the impact of mobility can be incorporated into our models because transmitter power can be adjusted to accommodate the new locations of the nodes, as necessary. After a discussion of the basic issues of multicasting in wireless networks, we consider the problem of broadcasting (i.e., transmission to all nodes in the network), in which the goal is the determination of the minimum-energy broadcast tree. We then return to the multicasting problem, in which we model the network's resources by means of "node capacity" (namely by assuming finite numbers of transceivers at each node), while assuming a large number of available bandwidth resources (i.e., unlimited number of frequencies or time slots or orthogonal CDMA codes, so that contention for the channel is not an issue). Our performance results demonstrate that the incorporation of energy considerations into the multicast algorithms can, indeed, result in energy saving.

Future studies will incorporate the impact of finite bandwidth. We do not address the protocol issues associated with determining connectivity and reserving resources, but instead focus on the fundamental issues associated with the determination of energy-efficient broadcast and multicast trees.

2 ENERGY-EFFICIENT COMMUNICATIONS

The problem of energy-efficient communications is a many-faceted one. Whereas traditional approaches to this problem have emphasized the development of improved batteries, low-power electronics, efficient coding and modulation, signal processing techniques, and antenna design, it has recently been recognized that networking techniques can also have a strong impact on the energy efficiency of such systems. For example, energy efficiency was identified as one of several research priorities at a recent National Science Foundation Workshop on Wireless and Mobile Communications [7]. The specific energy-related research topics recommended were:

- Development of protocols and architectures that minimize power usage in mobile, energy-constrained nodes;
- Exploration of techniques and trade-offs for integrated, system-wide, architectural optimization of energy consumption;
- Development of monitoring and management tools that enable energy-efficient designs.

Also of interest is a study on "Energy-Efficient Technologies for the Dismounted Soldier," by the Committee on Electric Power for the Dismounted Soldier, Board on Army Science and Technology, Commission on Engineering and Technical Systems, of the National Research Council [8]. In addition to chapters on the traditional approaches to energy efficiency, a chapter on networks, protocols, and operations is included.

A variety of networking-based approaches to energy-efficiency are possible. For example, protocols can be designed that minimize the occurrence of destructive collisions or the transmission of unnecessary (e.g., redundant) information. It is interesting to consider the case of the slotted ALOHA random-access protocol. Performance of such protocols is normally defined in terms of throughput (successful packets per time slot), and it is well known that the maximum throughput that can be achieved by this protocol (under standard modeling assumptions) is $1/e = 0.368$. Operation at maximum throughput is not very energy-efficient, however, since the fraction of packets that suffer collisions (at the traffic level that provides maximum throughput) is $1 - 1/e = 0.632$. Thus, a large fraction of the transmitted energy is wasted in the sense that it does not produce successful packet reception. In situations where energy is a precious commodity, it may be appropriate to redefine the performance metric to be throughput per energy unit, rather than simply throughput. Channel-access protocols should then be designed to minimize such energy-aware metrics. In Section 6.2 we discuss performance metrics that are appropriate for energy-efficient multicasting

Another aspect of energy-efficient networking is the design of asymmetrical protocols that place the energy burden on "large" users (such as vehicles and ships) rather than "small" users (individual warfighters). In addition to the energy used in transmission, a complete design would address receiver energy as well, including the energy expenditure to remain in the "on" state, even when not processing received signals. Ultimately, trade-offs between signal processing and communications should be addressed. Although data compression can result in the more-efficient use of transmitter power, signal processing costs at the receivers are also a major concern in evaluating the overall energy expenditure.

Additionally, energy-efficient routing schemes [9] can be developed, in some cases in conjunction with adaptive coding/modulation schemes that incorporate knowledge of link characteristics into network-level decisions [10].

[1] Mobility can be addressed later through soft-failure and hand-off mechanisms such as those in [6].

Our approach to energy-efficient communication departs from the traditional layered structure in that we jointly address the issues of transmitted power levels (and hence network connectivity, a Physical layer function) and multicast tree formation (a routing function, associated with the Network layer). We argue that such joint decisions on connectivity and routing can result in significant improvement in energy efficiency, as compared to a rigid layered structure that makes these decisions independently. Here we consider only the energy used for transmission, neglecting for the present the energy associated with reception and signal processing; the joint study of all forms of energy expenditure and the associated trade-offs are not considered here.

It is clear that turning "on" and "off" the transmitter and/or the receiver and choosing the transmission power prescribes a schedule for the "amortization" of the stored battery energy.[2] Another level of complexity is added to the control and scheduling of the network by the fact that, while in the "off" state, a node's transceiver cannot participate in a distributed control algorithm. In fact, this complication has potentially serious consequences on the delay performance and quality of service requirements of the network.

The generic problem of energy-efficient networking can be loosely defined as an optimization problem as follows:

- Maximize communication
  - for a given quantity of energy

or

- Minimize energy
  - for a given communication requirement.

In multihop applications, such as those considered in this paper, a major complication arises from the fact that a relay node's precious resources (its transceivers and its energy) are used by other network nodes. It is a major research challenge to develop methods to fairly use a network's resources in this manner. We speculate that an analogy to pricing in commercial networks may provide useful approaches in this area.

3 ARCHITECTURAL ISSUES IN ALL-WIRELESS NETWORKS

The ad-hoc wireless networks studied here are quite different from the cellular systems that have been developed in the commercial domain. Cellular systems have fixed base stations, which communicate among themselves using dedicated non-wireless lines; thus, the only multicast problems that are new in those systems involve tracking the mobile users. Otherwise, wireless communication is limited to that between mobile users and base stations. However, in ad hoc wireless networks it is possible to establish a link between any pair of nodes, provided that each has a transceiver available for this purpose and that the signal-to-noise ratio at the receiving node is sufficiently high. Thus, unlike the case of wired networks, the set of network links and their capacities are not determined a priori, but depend on factors such as distance between nodes, transmitted power, error-control schemes, other-user interference, and background noise. Thus, even when the physical locations of the nodes are fixed, many of the factors that affect network topology (and hence network control schemes) are (at least partially) influenced by the actions of the network nodes (either those directly participating in a link, or those contributing to the interference that affects a link). Furthermore, in ad hoc networks no distinction can be made between uplink and downlink traffic, thus greatly complicating the interference environment. Therefore, the wireless networking environment poses many new challenges not encountered in non-wireless or cellular networks, even when mobility is not addressed.

In this paper, we focus on wireless networks with fixed topology (i.e., the node locations are fixed, and the channel conditions unchanging). The wireless channel is distinguished by its *broadcast* nature; when omnidirectional antennas are used, every transmission by a node can be received by all nodes that lie within its communication range. Consequently, if the multicast group membership includes multiple nodes in the immediate communication vicinity of the transmitting node, a single transmission suffices for reaching all these receivers. Hence, there is an incentive to perform a multicast by using maximum power (and thus maximum communication range). Of course, doing so results in interference with more nodes than if reduced power were used. Thus, there is a trade-off between the long "reach" of a single transmission and the interference (and/or delay) effects it creates in its communication neighborhood.

Another undesirable impact of the use of high transmitter power is that it results in increased energy usage. Since the propagation loss varies nonlinearly with distance (at somewhere between the second and fourth power), in unicast applications it is best (from the perspective of transmission energy consumption) to transmit at the lowest possible power level, even though doing so requires multiple hops to reach the destination. However, in multicast applications it is not prudent to draw such conclusions because the use of higher power may permit simultaneous connectivity to a sufficiently large number of nodes, so that the total energy required to reach all members of the multicast group may be actually reduced. Furthermore, even for unicast applications, the use of lower power (and, hence, multiple hops) necessitates the complex coordination of more signals and therefore may actually result in higher total energy expenditure.

Thus, the choice of transmitted power levels depends ultimately on complex trade-offs between energy limitations and the demands of protocol operation. In view of the complex interdependencies

[2] Recent studies have shown that the total energy capacity of a battery is not fixed, but rather depends on the way in which the battery energy is used. For example, more energy can be obtained from a battery by means of pulsed, rather than continuous, operation [11], [12].

among many aspects of network design (e.g., transmitted power levels, signal processing considerations, spectral efficiency, mobility effects, etc.), the traditional layered architectures proposed for protocol design may not provide satisfactory performance. Therefore, it may be beneficial to design protocols that span several of the traditional layers to address appropriately the unique characteristics of the all-wireless environment [7], [13]. Our studies do, in fact, support this conjecture.

4 MULTICASTING IN WIRELESS NETWORKS

To date, virtually all of the research and development work on multicasting has centered on tethered, point-to-point (typically high speed) networks and on methods of bandwidth-efficient maintenance of multicast group addresses and routing trees. There are two basic approaches to multicast tree construction. The first is the use of source-based trees (SBT), which are rooted at the sender and which are designed to minimize the number of transmissions needed to reach all of the members of the multicast group. The second is the use of Center-Based (also known as Core-Based) Trees (CBT) [7], under which the same tree is used for all communication within the multicast group. The Sparse Mode of the Protocol Independent Multicasting (PIM) protocol [8] can be used with either SBTs or CBTs, whereas the PIM Dense Mode is based on the use of SBTs.

A crucial feature of the wireless medium, which distinguishes it from wired networks, is the noisy nature of the wireless channel, which can result in transmission errors. Background noise, other-user interference, and jamming contribute to channel impairment. Furthermore, there may be contention for receiver resources if a node is expected to listen to too many of its neighbors simultaneously. Therefore, a multicast message may be correctly received by one multicast group member but not by another. Thus, provisions are needed both for wireless multicast error control and for coordinating the multicast transmissions in an efficient manner; the goal is to mitigate the disadvantage of increased interference by exploiting the inherent broadcast advantage. Many of these research issues are addressed in [7].

Critical multicasting requirements include reliability and scalability. Packet-based Internet Protocol (IP) multicast schemes are typically based on "best effort" service, which does not guarantee reliable delivery. Although "absolute" reliability can be supported by TCP at the transport layer for unicast sessions, no such capability is currently available for multicast applications.

In this paper we focus on a single aspect of the multicasting problem, namely the incorporation of energy considerations into the construction of multicast trees and the choice of transmission power levels.

4.1 A Model for Wireless Multicast

We consider source-initiated, circuit-switched, multicast sessions. The network consists of N nodes, which are randomly distributed over a specified region. Any node is permitted to initiate multicast sessions. Multicast requests and session durations are generated randomly at the network nodes. Each multicast group consists of the source node plus at least one destination node. Additional nodes may be needed as relays to provide connectivity to all members of the multicast group.[3] The set of nodes that support a multicast session (the source node, all destination nodes, and all relay nodes) is referred to as a *multicast tree*.

The connectivity of the network depends on the transmission power. We assume that each node can choose its power level, not to exceed some maximum value p_{max}. The nodes in any particular multicast tree do not necessarily have to use the same power levels; moreover, a node may use different power levels for the various multicast trees in which it participates.

A constant bit rate (CBR) traffic model is assumed; thus, one transceiver is required to support each active multicast session at every node participating in the multicast tree throughout the duration of the session. Each node has K transceivers, and can therefore participate in up to K multicast sessions simultaneously. Since, as noted earlier, we assume in this paper that ample bandwidth is available, the only hard constraints we consider are the number of transceivers and the maximum permitted transmitter power p_{max}.

We assume that the received signal power varies as $r^{-\alpha}$, where r is the range and α is a parameter that typically takes on a value between 2 and 4, depending on the characteristics of the communication medium. Based on this model the transmitted power required to support a link between two nodes separated by range r is proportional to r^{α}. Without loss of generality, we set the normalizing constant equal to 1, resulting in:

$$p_{ij} = \text{power needed to support link between nodes } i \ \& \ j$$
$$= r^{\alpha},$$

where r is the distance between nodes i and j. If the maximum permitted transmitter power p_{max} is sufficiently large, the nodes will be able to transmit at sufficiently high power so that the network is fully connected.[4]

We assume the use of omnidirectional antennas; thus all nodes within communication range of a transmitting node can receive its transmission. It is important to note how the broadcast property of wireless communication can be exploited in multicast applications. Consider the example shown in Fig. 1, in which a subset of the

[3] Also, the use of relays (even when not absolutely necessary to provide connectivity) may result in lower overall energy consumption.

[4] We implicitly assume that the communication medium is uniform, i.e., α is constant throughout the region of interest, there are no obstacles (such as buildings or mountains), and that the region is totally flat (hence no line-of-sight limitations resulting from the earth's curvature).

multicast tree involves node i, which is transmitting to its neighbors, node j and node k. The power required to reach node j is P_{ij} and the power required to reach node k is P_{ik}. A single transmission at power $P_{i,(j,k)} = \max\{P_{ij}, P_{ik}\}$ is sufficient to reach both node j and node k, based on our assumption of omnidirectional antennas. This situation is fundamentally different from wired applications, in which the cost of node i's transmission to nodes j and k would be the sum of the costs to the individual nodes.[5] The ability to exploit this property of wireless communication, which we refer to as the "wireless multicast advantage," makes multicasting an excellent setting in which to study the potential benefits of energy-efficient protocols.

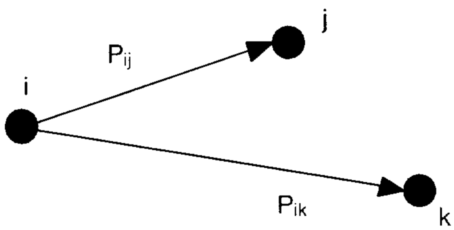


Fig. 1 — The "wireless multicast advantage:" $P_{i,(j,k)} = \max\{P_{ij}, P_{ik}\}$.

5 CONSTRUCTION OF MINIMUM-ENERGY BROADCAST TREES

Before addressing our specific multicasting problem, we first address the problem of constructing the minimum-energy broadcast tree for each newly arriving service request. Doing so involves a choice of transmitter-power levels and relay nodes. In addition to our assumption throughout this paper that ample bandwidth is available, we assume in this section that each node has a sufficient number of transceivers to accommodate all service requests. An insufficient quantity of either of these resources can result in the construction of trees that do not reach all destinations, use more than the minimum energy (because only suboptimal trees can be constructed), or both. In the simulations discussed later in this paper, we incorporate the impact of a finite number of transceivers.

We start with simple examples with two, and then three, destinations, and discuss how our results can be extended to larger examples by means of a recursive technique. Our examples in this section are based on the broadcasting problem, in which all nodes in the network (other than, of course, the source) are destinations. In Section 6 we return to the problem of multicasting, in which only a subset of the network nodes must be reached, while non-destination nodes may be used as relays. Including such nodes may result in reduced overall power consumption, or perhaps in providing a connected network where one was not achievable without the use of such relays.

It is important to emphasize a crucial difference between wired and wireless networks. In wired networks, the broadcasting problem can be formulated as the well-known minimum spanning tree (MST) problem. This formulation is based on the existence of a cost associated with each link in the network; the total cost of the broadcast tree is the sum of the link costs. The situation in wireless networks is different, however, because of the "wireless multicast advantage" property, discussed in Section 4, which permits all nodes within communication range to receive a transmission without additional expenditure of transmitter power. Therefore, the standard MST problem, which reflects the link-based nature of wired networks, does not capture the node-based nature of wireless networks. We do not know of any scalable solutions to the node-based version of this problem; the generalization of the MST problem to wireless networks is a possible approach, although we do not pursue it further here. In this paper we introduce one heuristic for the formation of low-energy broadcast trees, which takes into account the wireless multicast advantage. We use low-energy broadcast trees (including versions based on both link-based and node-based versions) as the basis for some of our heuristics for the construction of suboptimal multicast trees in wireless networks.

5.1 Minimum-Energy Broadcasting: Two Destinations

We consider a source node S (located at the origin) and two destination nodes D_1 (located along the x-axis, without loss of generality) and D_2, as shown in Fig. 2. The topology is specified by the coordinates of D_1 and D_2, which determine the angle θ. The distance between S and D_1 is r_1, the distance between S and D_2 is r_2, and the distance between D_1 and D_2 is r_{12}. It is assumed, without loss of generality, that $r_2 > r_1$. We define:

$P_{S1} = r_1^{\alpha}$ = power needed to support link between S & D_1

$P_{S2} = r_2^{\alpha}$ = power needed to support link between S & D_2

$P_{12} = r_{12}^{\alpha}$ = power needed to support link between D_1 & D_2

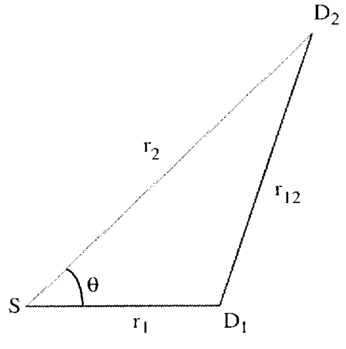


Fig. 2 — Broadcasting to two destinations.

[5] In wired networks, energy is not a concern; the cost of a link would typically be related to bandwidth and congestion (and hence delay) considerations. The case of wireless applications with highly directive antennas is similar to the case of wired networks in the sense that multiple beams may be needed to reach multiple destinations; thus the total cost of a node's transmissions to its neighbors would be equal to the sum of the cost of the individual beams needed to reach each individual destination.

In this simple example, there are two alternative strategies:

a) S transmits using P_{S2}: both D_1 and D_2 are reached

b) S transmits using P_{S1}: only D_1 is reached

D_1 then transmits to D_2 with power P_{12}, resulting in a total power of $P_{S1} + P_{12}$.

We would like to choose the alternative that results in the smaller value of total power consumption. For the case of propagation that follows a $1/r^2$ law, it is very simple to derive the following result from simple geometrical considerations:

- Use strategy (a) if $r_1 > r_2 \cos\theta$,
- Use strategy (b) otherwise.

For the general case of propagation behaving as $1/r^\alpha$, algebraic manipulation results in the following:

- Use strategy (a) if $x^\alpha - 1 < (1 + x^2 - 2x\cos\theta)^{\frac{\alpha}{2}}$,(1)

  where $x = r_2/r_1$;
- Use strategy (b) otherwise.

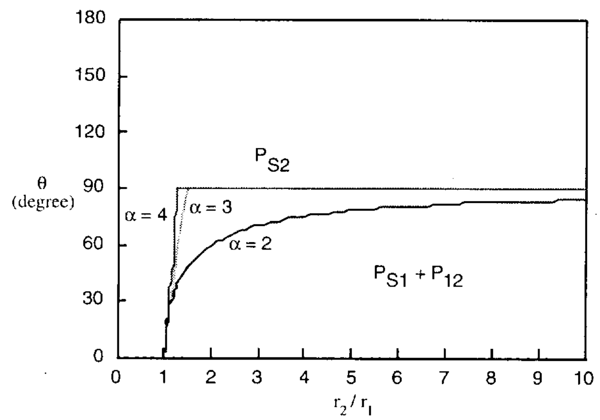


Fig. 3 — Transmission strategies for minimum-energy broadcasting to two destinations ($r_2 \geq r_1$).

This result is shown graphically in Fig. 3. For example, in the region above the curve (for each particular value of α) it is best to use strategy (a). It is of special interest to note that for $\alpha \geq 3$ (which is characteristic of many realistic environments) the boundary separating these regions is quite steep; therefore a simple heuristic that uses strategy (a) whenever $\theta \geq 90°$ and strategy (b) otherwise should be expected to provide nearly optimal performance. Thus, the incentive to use the shortest available links increases as α increases.

We acknowledge that, in practical applications, the locations of the nodes generally will not be known precisely. Also, the propagation characteristics are often difficult to characterize. Nevertheless, heuristics such as the one described here, which can depend on estimates of these quantities, are expected to provide insight into the development of good (although suboptimal) broadcast trees.

5.2 Minimum-Energy Broadcasting: Three Destinations

The minimum-energy broadcasting problem becomes more interesting as the number of destinations increases. In such cases, it is harder to make generalizations about the desirability of using the shortest available links because the use of a higher power transmission can often result in the ability to reach several nodes with a single transmission, thereby resulting in lower overall power in the complete tree. Figure 4 shows the case of broadcasting to three destinations. We enumerate the alternative strategies.

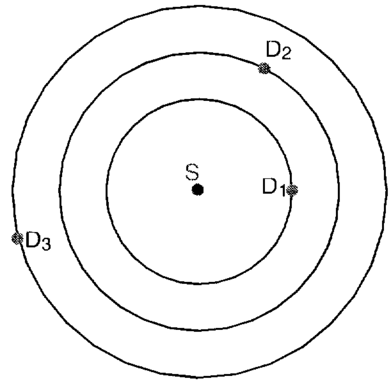


Fig. 4 — Broadcasting to three destinations.

a) S transmits using P_{S3}: All three destinations are reached.

b) S transmits using P_{S2}: Destinations D_1 and D_2 are reached by this transmission. One of these nodes must then transmit to D_3. The two alternatives are:

- D_1 transmits to D_3: total power = $P_{S2} + P_{13}$
- D_2 transmits to D_3: total power = $P_{S2} + P_{23}$

c) S transmits using P_{S1}: Only D_1 is reached by this transmission. D_1 must then form a tree to nodes D_2 and D_3. The three alternatives are:

- D_1 transmits with enough power to reach D_2 & D_3: total power = $P_{S1} + \max\{P_{12}, P_{13}\}$.
- D_1 transmits to D_2, which transmits to D_3: total power = $P_{S1} + P_{12} + P_{23}$
- D_1 transmits to D_3, which transmits to D_2: total power = $P_{S1} + P_{13} + P_{32}$.

As in the case of two destinations, the strategy that minimizes total power is chosen.

5.3 A Recursive Formulation

The number of alternative strategies increases rapidly as the number of destinations increases. However, the effects of complexity can be mitigated somewhat by means of a recursive formulation. For example, let us consider alternative (c) in Section 5.2. If the source transmits using power P_{S1}, it effectively delegates to D_1 the responsibility of reaching D_2 and D_3. This is simply the problem of broadcasting to two destinations, which is precisely the problem solved in Section 5.1. One can thus remap the origin to the location of D_1, and use the already obtained solution to the problem of broadcasting to two destinations in the evaluation of strategies for the three-destination example. In general, the solution for N_D destinations can be expressed in terms of the solutions for various subsets of the solutions for a smaller number of destinations. Unfortunately, the complexity of this formulation is high, making it impractical except for small networks. One way to roughly estimate complexity is to evaluate the number of times that the solution for the two-destination problem is called during the course of the algorithm. For the case of four destinations, it is called three times, which is certainly easy to handle. However, the number of calls to this subproblem increases rapidly as N_D increases; e.g., for $N_D = 10$, more than 51,000 calls are needed, and for $N_D = 13$ more than 14 million calls are needed. Nevertheless, this approach may serve as the basis for a suboptimal heuristic that provides less than an exhaustive search of all possible trees.

6 A MULTICASTING PROBLEM

We now address the problem of determining an appropriate multicast tree for each arriving multicast call request, so that a reward function (which incorporates both throughput and energy efficiency) is maximized. The establishment of a multicast tree requires the specification of the transmitted power levels and the commitment of the needed transceiver resources throughout the duration of the multicast session. If there is no tree that can reach any of the desired destinations (because the needed resources are blocked), the call is rejected. If there are trees that can reach only a portion of the destination set, they are considered. In some cases (depending on the admission-control policy), where one or more of the intended destinations is costly to reach, the "best" multicast tree may include only a subset of the reachable destinations.

In a wired network, the determination of the minimum-cost multicast tree is equivalent to the Steiner tree problem, which is NP-complete. By contrast, the MST problem (in wired networks) is polynomial in complexity. It would be of great interest to formulate and develop heuristics for the Steiner tree problem in a node-based context.[6]

6.1 Admission-Control Policies

Recall that the establishment of a multicast session requires the allocation of a transceiver at every node participating in the session (source, relays, and destinations) throughout the duration of the session. A destination can be *reached* if there exists a path from the source to it, and provided that a transceiver is available (i.e., not already supporting another session) at each node along the path. There are two basic aspects to the admission-control problem, i.e., whether or not to establish a multicast session for a particular multicast request, and (assuming a session is, in fact, established) which of the desired destinations to include in the multicast tree. Most of the results presented in this paper are based on the use of the "admit-all" admission-control policy, under which all multicast requests are accepted as long as one or more of the intended destinations can be reached; furthermore, under such schemes paths are established to all reachable destinations (i.e., the potential energy savings from dropping a subset of the destination nodes is not an option). In some cases, however, we do consider admission-control policies in which costly destinations are not included in the tree.

6.2 Performance Metrics

We define:

n_i = # of intended destinations by ith multicast arrival.

m_i = # of destinations reached by ith multicast session.

d_i = duration of ith multicast session (assumed to be exponentially distributed with mean = 1).

p_i = sum of the transmitter powers used by all nodes in ith multicast session.

E_i = total energy used by ith multicast session = $p_i\, d_i$.

v_i = *multicast value* of ith multicast session.

Since the quantity of information delivered is proportional to the duration of a session and to the number of destinations reached, we define the multicast value of session i to be:

$$v_i = m_i\, d_i. \tag{2}$$

A variety of performance measures can be defined for the multicasting problem, including the following.

Average (per call) multicast value per unit energy

The average (per call) multicast value per unit energy V_E, observed over an interval with X multicast requests, is[7]

$$V_E = \frac{1}{X}\sum_{i=1}^{X}\frac{v_i}{E_i} = \frac{1}{X}\sum_{i=1}^{X}\frac{m_i d_i}{p_i d_i} = \frac{1}{X}\sum_{i=1}^{X}\frac{m_i}{p_i}. \tag{3}$$

We observed in [15] that use of this metric alone tends to favor the hoarding of energy because this metric can often be maximized by transmitting to only those destinations that can be reached with very little energy consumption. Thus, only a small fraction of the desired destinations would typically be reached in multihop

[6] Although [14] addressed the multicasting problem with a goal toward reaching efficient and near-minimum-cost algorithms for wireless networks, their approach was link-based, rather than node-based, and hence does not take into consideration the wireless multicast advantage.

[7] Totally unsuccessful multicast arrivals, in which no destinations are reached, do not contribute to either throughput or energy expenditure.

networks when this metric is used as the basis for an admission-control policy.

Multicast efficiency

Also of interest is the *multicast efficiency* of the ith multicast session, which can be defined as the fraction of desired destinations of the ith multicast service request that are actually reached. Then, the overall multicast efficiency can be defined as:

$$e = \frac{1}{X}\sum_{i=1}^{X}\left(\frac{m_i}{n_i}\right). \tag{4}$$

This metric is maximized when all possible destinations are reached, without regard to the energy required to do so.

The "Yardstick" metric

To take into consideration both of the criteria discussed above, namely reaching many destinations per unit energy and reaching a large fraction of the number of desired destinations, we define a *local yardstick* measure of multicast performance to be:

$$y_i = \left(\frac{m_i}{p_i}\right)\left(\frac{m_i}{n_i}\right). \tag{5}$$

Our *global yardstick* Y is the average value of this quantity over the observation interval:

$$Y = \frac{1}{X}\sum_{i=1}^{X} y_i = \frac{1}{X}\sum_{i=1}^{X}\left(\frac{m_i}{p_i}\right)\left(\frac{m_i}{n_i}\right). \tag{6}$$

This metric is the primary performance metric we use to evaluate multicasting algorithms in this paper. In addition, we also consider the fraction of blocked calls as another metric.

Blocking Probability

We define k_X to be the number of multicast sessions that are completely blocked during an interval with X multicast requests, either because the resources are not available to reach any destinations or because the admission-control policy decides that it is not cost effective to form paths to any destinations. The *blocking probability* is defined as

$$P_B = \frac{k_X}{X}. \tag{7}$$

6.3 "Local" Cost Metrics

The problem of finding the multicast tree that maximizes the local yardstick for each new multicast request is highly complex, and not feasible, except for small examples. Therefore, we have found it necessary to take the approach of minimizing a cost function that is related to the ultimate objective, but only indirectly, and which is based on the use of local (i.e., per multicast request and/or link- or node-based) cost metrics.

6.3.1 Link-Based Costs

Consider link ij, which is established between nodes i and j. We define

D_{ij} = cost associated with link ij.

In this paper we define the cost of a link as the power level needed to support it, provided that at least one transceiver is available at both nodes.[8] If either node has no available transceiver (i.e., all are already committed to currently active sessions), the cost of the link is infinite. If the power required to support the link between nodes i and j is P_{ij},

$$D_{ij} = \begin{cases} P_{ij}, & \text{if there is at least one transceiver available at nodes } i \text{ and } j \\ \infty, & \text{otherwise} \end{cases} \tag{8}$$

The total cost of the tree can be defined (especially in wired networks) as the sum of the costs of all links in the tree.

The total cost to implement the multicast tree can be less than the sum of the link costs, however. It is sufficient for a node to transmit only once to reach all of its neighbors. The wireless multicast advantage applies here, since the total power required to reach several neighbors is the maximum power to reach any of them individually. However, the tree is selected without the ability to exploit the wireless multicast advantage in the choice of transmitting nodes.

6.3.2 Node-Based Costs

Since (under our assumptions of omnidirectional antennas and no interference) a node's transmission can be received by all of its neighbors, it would be best to design a tree that exploits the wireless multicast advantage. Tree formation would consist of a choice of transmitting nodes and their transmitting powers. The total cost of the tree is then the sum of the powers of all transmitting nodes. A minimum-cost tree is then one that reaches all reachable nodes with minimum total power. We know of no scalable algorithms for the minimum-cost broadcast tree problem, and certainly not for the presumably more difficult problem of minimum-cost multicasting.

7 ALTERNATIVE ALGORITHMS

In this section we discuss several of the multicasting algorithms we have studied; full descriptions are available in [17]. In this paper we define the notion of the cost associated with the support of a multicast tree to

[8] In [16] we discuss link cost functions that incorporate congestion. In this paper, however, we focus on the use of power (rather than congestion) as the cost metric to facilitate the comparison of several distinct algorithms without having to address subtle issues relating to the definition of alternative cost functions.

be the power required to reach all destination nodes; thus, it is the sum of the powers at all transmitting nodes. This is a metric that is used as the basis of some of our algorithms. However, performance is always judged by the "yardstick" metric, the multicast efficiency, and the blocking probability.

Each transmission by a node is characterized by its transmitter power level, as well as a designation of which (possibly several) of the nodes receiving this transmission are to forward it toward which of the ultimate destination nodes. In all cases, we use greedy algorithms, which attempt to optimize performance on a "local" (call-by-call) basis. Such greedy algorithms do not necessarily result in "global" (long-term average) optimal performance, even if they maximize a performance metric (e.g, the local yardstick y_i) or minimize the energy associated with each arriving session.

When the number of transceivers at each node (K) is finite, it may not be possible to establish minimum-energy trees (even on a local basis) because of the lack of resources (transceivers) at one or more nodes. In this case, the greedy algorithms discussed here are applied to the subset of nodes that have non-zero *residual capacity.*[9]

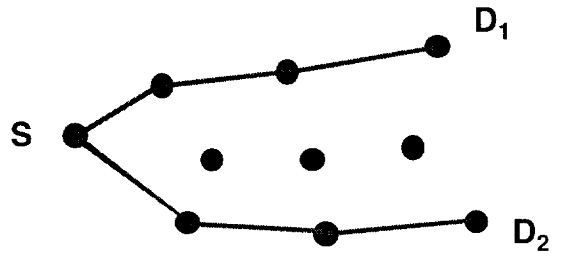
7.1 A Unicast-Based Multicast Algorithm

A straightforward approach is the use of multicast trees that consist of the superposition of the best unicast paths to each individual destination (see e.g., [18]). It is assumed that an underlying unicast algorithm (such as the Bellman-Ford or Dijkstra algorithm) provides "minimum-distance" paths from each node to every destination. However, the minimization of unicast costs does not necessarily minimize the cost of the multicast tree, as illustrated in Fig. 5, which shows a source and two destinations. Figure 5a shows the best unicast paths that reach the two destinations, and Fig. 5b shows the best multicast tree. The use of the best unicast paths fails to discover the path that reaches a neighbor of both destinations over a common path, thereby resulting in lower overall cost. Also, the use of the best unicast paths fails to incorporate the "multicast advantage," which was discussed in Section 2. Therefore, the trees obtained based on unicast information are not expected to provide optimal multicast performance. Nevertheless, they do perform reasonably well, and with considerably reduced complexity as compared to the calculation of truly optimal multicast trees.

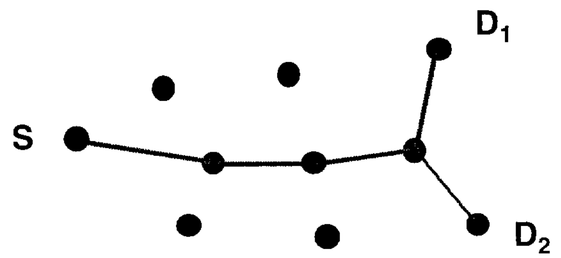
It is significant to note that, although algorithms based on minimum-distance paths are normally used for packet-switched applications, we are using this approach here for session-oriented traffic. We feel that it is appropriate to do so in wireless applications because a cost (involving power and possibly congestion) can be defined for each link in the network. By contrast, in circuit-switched wired applications it is difficult to define a link cost because energy is not of concern and because delay is not an appropriate metric (as it would be in packet-switched applications) since resources are reserved in circuit-switched applications. Summarizing the above, we have:

Algorithm 1) Least-Unicast-Cost Multicast

A minimum-cost path to each reachable destination is established. The multicast tree consists of the superposition of the unicast paths. Paths to all reachable destinations are established, regardless of the cost required to do so. This algorithm is scalable.



(a) best unicast paths



(b) best multicast tree

Fig. 5 — Unicast-based vs multicast-based trees.

7.2 Algorithms Based on Pruning MSTs

One approach we have taken in the development of heuristics for multicasting is the pruning of broadcast spanning trees. To obtain the multicast tree, the broadcast tree is pruned by eliminating all transmissions that are not needed to reach the members of the multicast group.

We noted earlier that, for the case of wired networks, the determination of minimum-cost broadcast (spanning) trees is considerably easier than the determination of minimum-cost multicast trees. Nevertheless, the determination of minimum-cost broadcast trees for wireless networks remains a difficult problem for which no scalable solutions appear to be available at this time. In small network examples we have determined minimum-energy spanning trees by using the recursive technique of Section 5.3; in moderate to large networks it is necessary to use heuristics. In this subsection we discuss the main features of three algorithms that are based on the technique of pruning. Further details are provided in [17].

[9] The residual capacity at node j is the number of transceivers at node j that are not currently supporting traffic, and hence are available to support new sessions.

Algorithm 2) Pruned Link-Based MST Multicast

This algorithm is based on the use of the standard MST formulation in which a link cost is associated with each pair of nodes (i.e., the power to sustain the link); thus the "wireless multicast advantage" is ignored in the construction of the MST. Since the MST problem is of polynomial complexity, it is scalable. To obtain the multicast tree, the MST is pruned by eliminating all transmissions that are not needed to reach the members of the multicast group. Once the MST is constructed in this manner, the evaluation of its cost (i.e., the total power needed to sustain the broadcast tree) does take into consideration the wireless multicast advantage.

Algorithm 3) Pruned Node-Based MST Multicast

This algorithm requires the determination of the minimum-energy spanning tree that is rooted at the Source node. Unlike Algorithm 2, the wireless multicast advantage is taken into consideration in the determination of the power needed to sustain the tree. An exhaustive search is needed to determine the minimum-spanning tree. Thus, this method is not scalable. Once the MST has been determined in this manner, it is pruned as in Algorithm 2.

Algorithm 4) Pruned Node-Based Spanning Tree Multicast

A heuristic is used to determine a suboptimal spanning tree, i.e., a spanning tree with low (but, in general, not minimum) power.[10] Once the spanning tree has been determined in this manner, it is pruned as in Algorithm 2.

Construction of the spanning tree begins at the Source node. Its transmission power is chosen to maximize the following metric:

$$\frac{n}{p} = \frac{\text{Number of "new" destinations reached}}{\text{Total power required to reach them}} \tag{9}$$

We refer to this as the n/p metric. In some cases, based on this metric the Source will transmit with only enough power to reach its nearest neighbor; in some cases, this metric will be maximized by transmitting at sufficient power to reach several destinations.

At the next stage, each of the nodes that has been "covered" (i.e., the Source node plus all nodes within its communication range based on the calculation in the first stage) evaluates the n/p metric for all possible sets of neighbors (however, in computing this metric, only "new" nodes, i.e., nodes not previously covered, are included in the number of destinations). Note that it is possible to increase the transmission power that was assigned to a node in an earlier stage. This procedure is repeated until all nodes are covered. Full details are provided in [17].

[10] In small examples, this heuristic typically provides a lower-energy broadcast tree than that produced by the link-based method of Algorithm 2, but does not provide the true minimum that can be obtained by means of the recursive scheme.

The motivation behind this algorithm is that it is beneficial to reach many nodes per unit of transmitted power. This is similar, in principle, to the Yardstick performance metric. Although it is easy to construct examples in which this algorithm does not provide optimal performance, it performs reasonably well and converges rapidly.

7.3 Additional Algorithms with High Complexity

The following algorithms require an exhaustive search, and are thus not scalable. Nevertheless, they provide a useful benchmark that permits us to evaluate the performance of the other algorithms.

Algorithm 5) Least-Multicast-Cost Multicast

As in Algorithm 1, paths to all reachable destinations are established, regardless of the cost required to do so. An exhaustive search of all multicast trees that reach all reachable destinations is performed. The tree with the lowest cost is chosen.

Algorithm 6) Local-Maximum-Yardstick Multicast

The yardstick function y_i is computed for each arriving multicast request i. Multicast trees are formed to all subsets of intended destinations. The tree that results in the maximum value of y_i is chosen. This tree does not necessarily include all reachable destinations.

8 PERFORMANCE RESULTS

We have simulated the performance of the six algorithms for the 8-node network shown in Fig. 6. The connectivities shown are based on a maximum permitted transmitter power value of $p_{max} = 10$ and $\alpha = 2$, which result in a maximum communication range of 3.16 (where the overall dimensions of the region are 5 by 5). The transmitter power actually used (r^2) depends on the distance (r) to the farthest neighbor to which a node is transmitting. The same connectivities apply to the case of $p_{max} = 100$ and $\alpha = 4$, in which case the power actually used is r^4.

In our simulations, multicast requests arrive randomly; interarrival times are exponentially distributed with rate λ. Service durations are exponentially distributed with mean 1 (i.e., $\mu = 1$). The multicast group, which must have at least two nodes (the source plus at least one destination) is chosen randomly at each arrival instant as follows. In a network with N nodes, one of the $2^N - N - 1$ subsets with at least two members is chosen with probability $(2^N - N - 1)^{-1}$. Then, a uniform distribution is used to choose one of the members of the multicast group to be the source. Each simulation run consists of $X = 1{,}000$ multicast session requests. We have run our simulations for numerous values of arrival rate λ.

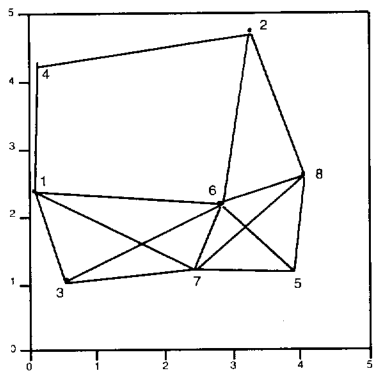


Fig. 6 — An example 8-node network ($p_{max} = 10$, $\alpha = 2$; or $p_{max} = 100$, $\alpha = 4$).

For all six algorithms, Fig. 7 shows the global yardstick metric Y as a function of arrival rate λ for the case of $T = 4$ transceivers at each node. The ordering of the algorithms in the legend of the figure is based on their relative performance (at low to moderate traffic loads); e.g., the best performance is provided by Algorithm 6 and the worst by Algorithm 2. It is not surprising that the best performance is obtained by Algorithm 6, which was designed specifically to maximize the local yardstick. The next best performance is obtained by Algorithm 5, which (like Algorithm 6) is based on an exhaustive search of all possible multicast trees. Thus, the two most highly complex algorithms provide the best performance. Although these two algorithms are too complex for practical applications, they are being studied because they can provide a benchmark of the performance that is achievable through appropriate choice of transmitter power levels and multicast trees.

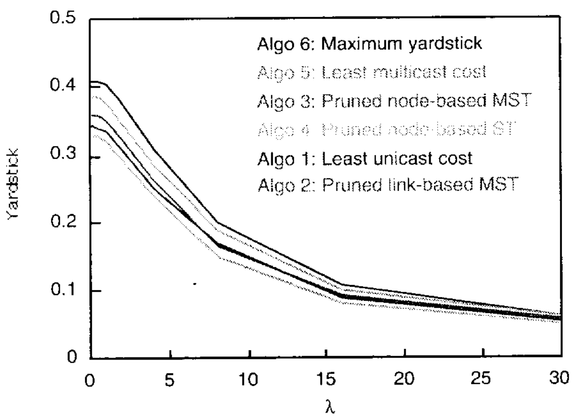


Fig. 7 — Global yardstick Y vs λ for $T = 4$, $\alpha = 2$.

Three of the four other algorithms are scalable. It is interesting to compare the performance of Algorithm 6 with that of Algorithm 1 (the first algorithm we studied and one of the simplest to implement). The fact that Algorithm 6 provides approximately 19% better yardstick performance than Algorithm 1 suggests that improvement can, in fact, be obtained through the exploitation of wireless networking properties, i.e., the choice of transmitter powers and relay nodes. On the other hand, the fact that simple algorithms can provide relatively good performance and the relatively small differences in performance among Algorithms 1 through 4 indicates a high degree of robustness in that a variety of well-motivated algorithms can provide similar, and possibly acceptable performance.

Figure 8 shows the multicast efficiency e as a function of λ. At low to moderate traffic levels, there is little difference in performance among Algorithms 1 – 5; however, Algorithm 6 provides considerably lower values of e. The low multicast efficiency provided by Algorithm 6 results from the fact that it, unlike the other algorithms, does not provide paths to costly destinations. Thus, as expected, there is a trade-off between the conflicting goals of providing high values of Y and high values of e.

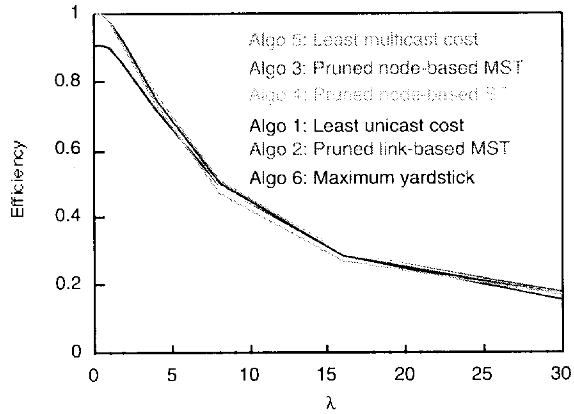


Fig. 8 — Efficiency e vs λ for $T = 4$, $\alpha = 2$.

Figure 9 shows the blocking probability P_B (the probability that none of the destinations are reached) as a function of λ. Algorithm 6 (uppermost in the legend, and at the bottom of the P_B curves) provides the best performance based on this metric. This behavior can be explained by the fact that since Algorithm 6 does not provide paths to costly destinations, fewer resources are committed to support ongoing calls than with the other algorithms (for a given traffic level λ). Thus, more resources tend to be available, and fewer calls are totally blocked.

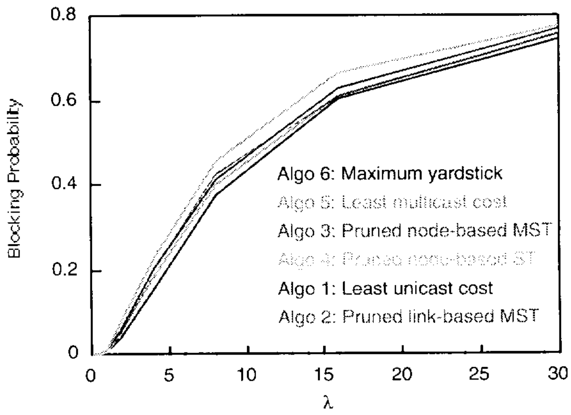


Fig. 9 — Blocking Probability P_B vs λ for $T = 4$, $\alpha = 2$.

Figures 10 – 12 show Y, e, and P_B for the same network, but with $\alpha = 4$ and $p_{max} = 100$. Qualitatively, the plots are similar to those for $\alpha = 2$, in the sense that Algorithm 6 provides the best performance on the basis of metrics Y and P_B, and the worst performance based on e. However, the difference in performance between Algorithm 6 and the others is much greater for $\alpha = 4$. As α increases, the incentive to use the shortest possible links increases. Also, as α increases, the cost of including distant destinations in the tree increases rapidly. Thus, there is an incentive to exclude such costly destinations from the multicast tree, which is an option only for Algorithm 6, but not for the other algorithms we have evaluated thus far. Based on these observations, our future studies will develop algorithms that do not necessarily use the admit-all admission-control policy.

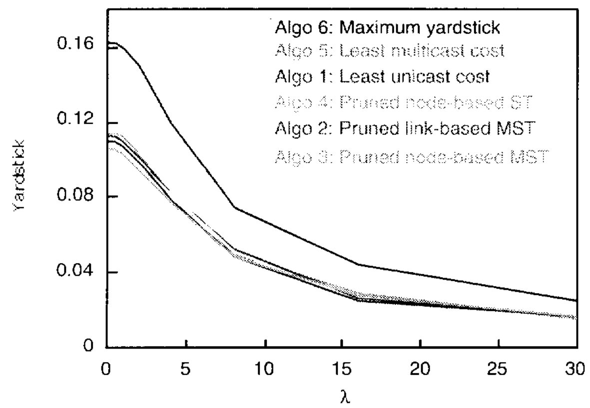


Fig. 10 — Global yardstick Y vs λ for $T = 4$, $\alpha = 4$.

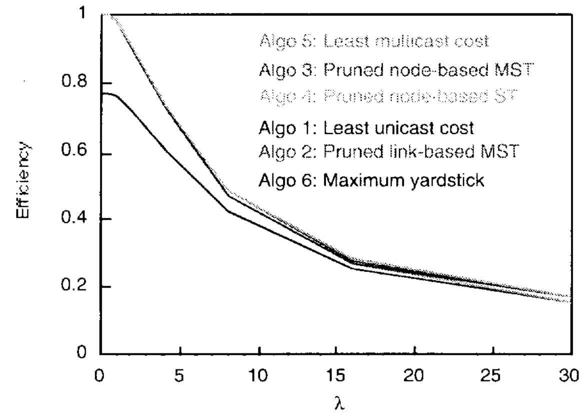


Fig. 11 — Efficiency e vs λ for $T = 4$, $\alpha = 4$.

8.1 A Larger Network Example: 100-node Network

We have also applied three of the scalable algorithms (1, 2, and 4) to a network consisting of 100 nodes. Like the eight-node network of Fig. 6, all nodes are located randomly in a square region of dimensions 5 by 5. Figures 13 and 14 show the global yardstick Y for the cases of $\alpha = 2$ and 4, respectively. Our first observation is that the yardstick values are considerably higher for the 100-node network than for the eight-node network. The higher yardstick values result from the generally much smaller communication ranges that can be implemented by means of extensive relaying in the considerably denser 100-node network (since 100 nodes are now located in the same region as eight nodes in our earlier examples). The impact of these smaller ranges is especially apparent for the case of $\alpha = 4$.

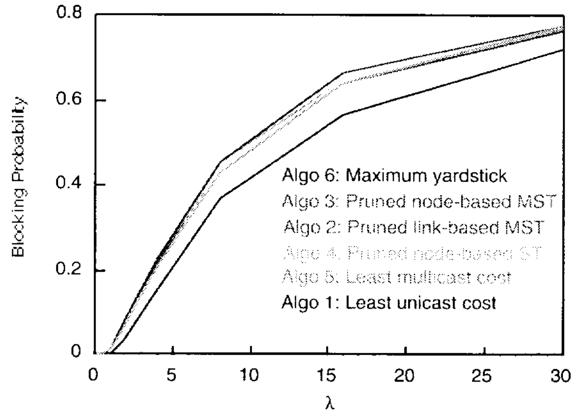


Fig. 12 — Blocking Probability P_B vs λ for $T = 4$, $\alpha = 4$.

For the case of $\alpha = 2$, the two algorithms that are based on the pruning of spanning trees (Algorithms 2 and 4) provide better performance than the unicast-based algorithm, especially at low levels of offered load. Algorithm 4 provides the best yardstick performance over the entire throughput range, while Algorithm 1 provides the worst.

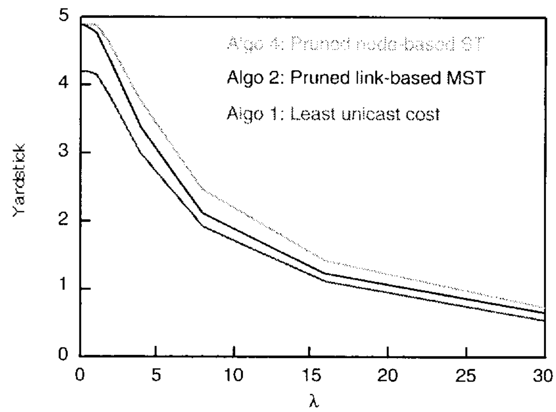


Fig. 13 — Global yardstick Y vs λ for $T = 4$, $\alpha = 2$, 100-node network.

For the case of $\alpha = 4$, the best performance is provided by Algorithm 2 (the pruned link-based spanning tree), while the performance of Algorithms 1 and 4 are virtually indistinguishable in the curves. We suspect that the relatively poor performance of Algorithm 4 may be a result of the implementation of Eq. (9), which can result in long communication ranges when their use results in reaching many destinations (many of which may not be helpful to the construction of the eventual multicast tree). The "penalty" of using long communication ranges is especially severe for large values of α. Further study is needed to provide a better understanding of the performance of the algorithms, and in particular the effects of pruning spanning trees.

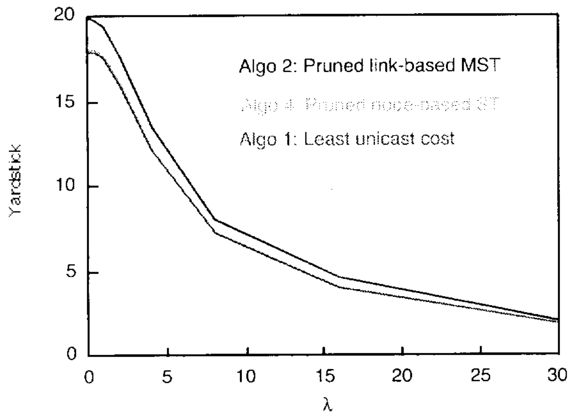


Fig. 14 — Global yardstick Y vs λ for $T = 4$, $\alpha = 4$, 100-node network.

9 CONCLUSIONS

In this paper, we have addressed some of the fundamental issues associated with energy-efficient wireless networking, and we have studied the specific application of multicasting in static ad hoc wireless networks. We have presented preliminary algorithms for the solution of this problem, and we have evaluated their performance via simulation. Our studies to date indicate that improved performance can be obtained when exploiting the properties of the wireless medium, e.g., the wireless multicast advantage. The fact that improved performance can be obtained by jointly considering physical layer issues and network layer issues suggests that novel approaches to wireless networking, which incorporate the vertical integration of protocol layer functions, may provide advantages over traditional network architectures.

However, further study of our algorithms is needed, including the study of additional (and larger) networks. Additionally, further research is needed to develop scalable algorithms that can achieve nearly optimal performance. On the other hand, we have demonstrated that reasonably good performance can be obtained by using simple, scalable heuristics. Ultimately, we plan to incorporate node mobility, as well as interference considerations that are associated with finite bandwidth. Additionally, we plan to address the impact of signal processing considerations on network operation, and to develop network control schemes that jointly address networking and signal processing issues.

References

[1] M. Nagy and S. Singh, "Multicast Scheduling Algorithms in Mobile Networks," *Cluster Computing*, **1**-2, pp. 177-185, December 1998.

[2] C.-C. Chiang, M. Gerla, and L. Zhang, "Forwarding Group Multicast Protocol (FGMP) for Multihop, Mobile Wireless Networks," *Cluster Computing*, **1**-2, pp. 187-196, December 1998.

[3] C. Diot, W. Dabbous, and J. Crowcroft, "Multipoint Communication: A Survey of Protocols, Functions, and Mechanisms," *IEEE Journal on Selected Areas in Communications*, **15**-3, pp. 277-290, April 1997.

[4] M. H. Ammar, G. C. Polyzos, and S. K. Tripathi, eds. "Special Issue on Network Support for Multipoint Communication," *IEEE Journal on Selected Areas in Communications*, **15**-3, April 1997.

[5] L. Zhang, S. Deering, D. Estrin, S. Shenker, and D. Zappala, "RSVP: A New Resource ReSerVation Protocol," *IEEE Network Magazine*, pp. 8-18, September 1993.

[6] A. Michail and A. Ephremides, "Routing Algorithms in All-Mobile Wireless Networks," *Proceedings First Annual ATIRP Conference*, pp. 119-123, January 1997.

[7] National Science Foundation, "Research Priorities in Wireless and Mobile Communications and Networking, Report of a Workshop," March 1997. (http://www.isi.edu/~workshop/wmcw97/)

[8] National Research Council, *Energy-Efficient Technologies for the Dismounted Soldier*, Washington, DC: National Academy Press, 1997.

[9] T. J. Shepard, "A Channel Access Scheme for Large Dense Packet Radio Networks," *SIGCOMM'96*, pp. 219-230, August 1996.

[10] M. B. Pursley, H. B. Russell, and J. S. Wysocarski, "Energy-Efficient Routing in Frequency-Hop Networks with Adaptive Transmission," *submitted to IEEE MILCOM'99*, Atlantic City, NJ, 1999.

[11] E. J. Podlaha and H. Y. Cheh, "Modeling of Cylindrical Alkaline Cells VI: Variable Discharge Conditions," *Journal of the Electrochemistry Society*, **141**, pp. 28-35, January 1994.

[12] M. Zorzi and R. R. Rao, "Error Control and Energy Consumption in Communications for Nomadic Computing," *IEEE Transactions on Computers*, **46**-3, pp. 279-289, March 1997.

[13] C. D. Rais and S. K. Tripathi, "Studying Vertical Dependence to Improve NFS Performance in Wireless Networks," *Cluster Computing*, **1**-2, pp. 225-235, December 1998.

[14] K. Makki, N. Pissinou, and O. Frieder, "Efficient Solutions to Multicast Routing in Communication Networks," *Mobile Networks and Applications*, **1**, pp. 221-232, 1996.

[15] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Multicasting in Energy-Limited Ad-Hoc Wireless Networks," *Proceedings of IEEE MILCOM'98*, Bedford, MA, pp. 723-729, October 1998.

[16] A. Ephremides, J. E. Wieselthier, G. D. Nguyen, and R. Bhattacharya, "Energy-Efficient Wireless Multicasting — a Beginning," *Proceedings Third Annual ATIRP Conference*, pp. 37-45, February 1999.

[17] J. E. Wieselthier, G. D. Nguyen, and A. Ephremides, "Energy-Efficient Networking — A First Look," to appear as an NRL Memorandum Report, Naval Research Laboratory, 1999.

[18] D. Bertsekas and R. Gallager, *Data Networks*, Englewood Cliffs: Prentice Hall, 1992.

Extending the Littoral Battlespace (ELB): Advanced Concept Technology Demonstration (ACTD)

(June 1999)

Edwin L. Althouse
Naval Research Laboratory
Code 5520
Washington, DC 20375-5337

Summary: This paper discusses the communications networking technology used to support a major U.S. Department of Defense Advanced Concept Technology Demonstration during April 1999 in support of modern concepts of littoral operations and operational maneuver from the sea. Commercial wireless communication products were modified to support extension of the Navy shipboard data networks to inland, non-mechanized troops and to mobile command posts. Demonstrated computer applications included IVOX "voice-over-data," Cu-See-Me (for multi-window laptop conferencing), video teleconferencing to large sites, fire-support software, command-control software, and government-developed, wearable-computer software called InCON™ to display the common tactical picture down to the lowest echelons

Background: The United States Navy and Marine Corps have embarked on an ambitious effort to integrate mature and emerging technologies to aid in the exploration of a new operational concept. This concept is contained in the well-publicized White Papers, "From the Sea" and "Forward From the Sea." Through these papers, the Navy and Marine Corps present a common vision for a future in which naval forces would enable the United States to project its influence wherever national policy requires. This vision places unprecedented emphasis on littoral regions of the world, requires more intimate cooperation between forces afloat and forces ashore, introduces the concept of the Naval Expeditionary Force, and provides the foundation for "Operational Maneuver from the Sea."

The objective of the ELB ACTD is to demonstrate the military utility of a revolutionary concept for joint expeditionary warfare enabled by advanced technology in the areas of remote sensing, communications, command center operations, and "fires and targeting." However, in accord with the Tactical Mobile Communications emphasis associated with this IST Symposium, this paper will discuss only the communication aspect of this ACTD. The ACTD program objectives will be achieved through a series of limited technical experiments leading up to two major operational demonstrations. Phase I of the program culminated in a major demonstration with operational forces in April 1999. Phase II will build upon the results and developments of Phase I to demonstrate an objective system with more complete capabilities and improved robustness in a major demonstration in mid-2001.

The ELB ACTD will attempt to demonstrate a seamless command structure between afloat and ashore units and shared situational awareness and understanding, thus allowing peer-to-peer communications and command by exception. The commander and staff will attempt to achieve total visibility of forces and their location and status through the command center. As forces engage targets, they will do so at great distances by targeting and calling for artillery support originating from weapon and loiter platforms at sea, on the ground, or in the air. Ground forces will operate in widely dispersed formations and will have the capability to mass fires rapidly and accurately from decentralized locations. With a communications infrastructure and fires-and-targeting capability designed to support the smallest combat unit deployed, a more flattened informational structure will result, allowing increased flow of information and greater optimization of resources. Sea-based Command-Control and Combat Operations Coordination precludes the establishment of large cantonments ashore and reduces the logistic footprint and vulnerabilities. Under this new operational concept, force versatility is increased under a variety of missions because their effectiveness is so much greater. Enhanced operations with coalition forces can be realized through equipping of these forces to allow them to inter-operate with our forces and shore situation understanding. Units will avoid direct firefights and rely upon remote weapons to engage the enemy. Survivability is increased by employment of numerous small and stealthy teams and because they present such a reduced target.

The objectives of the ELB project are synergistic with Vice Admiral Arthur Cebrowski's vision of Network-Centric Warfare (http://www.usni.org/Proceedings/Articles98/PROcebrowski.htm) in which operational advantage is achieved from the strong networking of a well-informed but geographically dispersed force.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

The communications infrastructure needed to support the ELB concept requires the ability to support reliable network connectivity at ranges that far exceed line-of sight. One cannot rely on the existence of commercial terrestrial communications infrastructure for these operations. It may not exist in the area of operations or, if it does, the adversary or adjacent country may be able to selectively disable service. Low Earth Orbiting (LEO) satellites can provide point-to-point voice, small paging messages, and low-rate data. However, they are not well suited to group voice sessions and cannot yet support high-rate data communications and multicast/broadcast service. Like terrestrial commercial service, large-scale usage of LEO commercial services by military forces could result in denial of service by an unsympathetic country who has control of the commercial services in that area. Military satellite service is already heavily congested among existing communication services and it does not easily provide medium data-rate service to Marines and dismounted soldiers in the context of the littoral operations defined above.

Consequently, the ELB ACTD has embraced the concept of employing a wireless network that reaches from the Command-Control and Combat Operations Center on Navy ships to forward-deployed Marines via one or more airborne relays. The relay platforms chosen for the demonstration were two Navy P-3s and two commercial Crownair aircraft; however, for an operational system other platforms including high-altitude UAVs will be considered. The objective was to be able to support packetized data and voice communications between the Combat Operations Center and Marines and dismounted soldiers over separations of about 300 miles without having to install massive supporting terrestrial infrastructure. Depending on the technology used to achieve the power-gain product required to support communications over the link, several airborne relays may be required to reach these distances. The use of an airborne relay that could be based on Navy ships (obviously not a P-3) could provide a quickly deployable system in the future that meets the operational needs in littoral engagements.

The desired qualities of the extended wireless network are as follows:

a. Ability to support point-to-point, multicast, and broadcast packet-switched communications among large- and small-capacity users over distances up to 300 miles.
b. Ability to support point-to-point and group voice service across the entire extended battle space.
c. Ability to include service to Marines and dismounted soldiers with battery-powered radio and computer at rates of at least 64 kbps.
d. Ability to include service to large users (ships, mobile combat centers) at rates up to 1.5 Mbps.
e. Ability to automatically configure and reconfigure the network topology and routing/switching databases as both users and airborne platforms move.

A competitive Broad Agency Announcement (BAA) was advertised by the Office of Naval Research in January 1998 to attract bidders and proposals for acquiring a prototype system. Because of the short schedule on this project and the high risk associated with the technology, a rapid down-select to four contractor teams was made from the initial proposals, and a six month competitive design process was used to pick the winning contractor. The winning proposal by General Dynamics was based on Lucent's WaveLAN products. During the ensuing period of system architecture definition and refinement, several other communications media were added. These consisted of the Marconi Aerospace Systems, Inc. VRC-99A network radio and SHF and Ku-band SATCOM. The VRC-99A was originally included as an alternative or backup for the WaveLAN devices, but later became the primary communication backbone. The SATCOM service was included to provide long-range network connectivity among ships and transportable combat operation centers during those periods when no communication relays were airborne.

Communication Objective: The communication objective for this ACTD was to provide the wireless communication network that is required to meet the operational goals defined above. It is synonymous with providing the high-performance, information grid described by Admiral Cebrowski's Network-Centric Warfare doctrine. Although communication networks to support this concept exist on land and large ships and between ships and land via satellite links, a suitable extension of this network into the tactical battlefield may not exist without the prior establishment of substantial terrestrial supporting infrastructure which is contrary to the notion of "Operational Maneuver from the Sea." Under this concept, troop insertion to great penetration is intended to be achieved with minimal pre-planted infrastructure and executed with speed and efficiency. The success of this type of operation depends on continuous and reliable information exchange in both directions at all times among all elements of command-control, logistics and re-supply, combat operations, sensing and targeting. Information will have to flow between Naval ships and dismounted soldiers and marines over distances of 300 miles or more. Satellite links can be used for transportable field-operation centers but they do not accommodate highly mobile operations, particularly at the lower echelons. Needed here is a highly dynamic, multiply-connected wireless network that supports mobility of the forces as well as the mobility of the relay platforms and the changes in network topology that this mobility introduces. The organization and reorganization of the wireless network topology should be automatic and without human intervention. It should support roaming in the sense that any node should be able to connect to the network at any available access point and then re-affiliate at any time to any other access point as required by the mobility of either the field unit or the elements of the network (platforms containing routers or switches).

The network must be capable of supporting numerous computer applications such as e-mail, file and image transfer, web browsing, interactive white-board, and collaborative voice. Video Teleconferencing is also desired among major platforms. The basic service is anticipated to be provided by connectionless datagrams. Where needed, reliable link- and end-to-end-services will be achieved by protocol overlays. Desired data rates were estimated to be 1.5 Mbps between major platforms and 64 kbps to the dismounted warrior. Between major platforms the network connectivity was expected to provide service similar to what would be achieved by common attachment to a LAN.

ELB Communication Systems: As discussed in the Background section, the winning industry proposal for providing the supporting technology for this ACTD was based on the use of Lucent's WaveLAN wireless LAN products. However, as the system engineering for this ACTD evolved, two other communication technologies were admitted to the overall architecture. These additional systems were the Marconi Aerospace VRC-99A radio and small aperture, Ku-band SATCOM. Each of these three systems will be discussed in detail below. SATCOM was added to the architecture to enable communications to large sites (trucks and ships) during those periods when there could be no airborne relay. The VRC-99A radio was added as a risk-mitigating alternative to WaveLAN; its major disadvantage is that its size and weight preclude use as a manpack radio.

WaveLan/IEEE. The latest WaveLAN product, known as WaveLAN/IEEE (http://www.wavelan.com/), provides a wireless network interface to any computer via an ISA card or a PCMCIA card, the latter being of primary interest for this ACTD. WaveLAN/IEEE cards have a pair of antennas embedded directly into the card and have a miniature connector to allow substitution of external main (transmit/receive) and auxiliary (receive only) antennas. For each incoming message, WaveLAN/IEEE selects the antenna on which to receive data based on how well the 802.11 physical layer preamble is received. WaveLAN also uses this preamble to adapt its channel-matched filters for multipath mitigation. WaveLAN/IEEE operates in the unlicensed Instrumentation, Scientific, and Medical (ISM) band in the vicinity of 2.4 GHz. The normal transmit power for a WaveLAN transmitter is about 32 mW and FCC regulations for the ISM band in the United States require that the power applied to an antenna (with up to 6-dB gain) not exceed 1 W. However, because of the extended ranges required for this ACTD, additional power was required. Lucent requested and obtained approval from the FCC to apply up to 30W into an antenna (with up to 20 dB gain) for operations associated with this ACTD. To adapt WaveLAN to higher power, a custom electronics component, as shown in figure 1, was developed. This adapter contained a very fast transmit/receive (T/R) switch, a 30W power amplifier, and a low-noise amplifier. The purpose of the low noise amplifier was to attempt to increase the sensitivity of the receiver to within a few dB of the thermal noise limit. To accommodate this adapter, the factory antenna was bypassed and the adapter was connected to the external connector on the WaveLAN card. The final switch on the adapter was added to support a single transmitting antenna and dual receiving antennas to accommodate WaveLAN's ability to support switched antenna diversity. RF filters and limiters were also incorporated to protect against overload from shipboard sources of co-site interference.

For the manpack terminal, the WaveLAN PCMCIA card was enhanced with a switched 6-W power amplifier and an integrated low-noise amplifier and RF filter. This device weighed about 1 kg and preserved the ability to switch between main and auxiliary antennas.

The channel access protocol for WaveLAN/IEEE is Carrier Sense, Multiple Access with Collision Avoidance (CSMA/CA) as specified by IEEE 802.11. Within a single channel, all WaveLAN participants share the channel by the CSMA time-sharing process. Only one transmitter is intentionally active at any time; otherwise, a collision will occur and require a retransmission. Multiple computers can exchange data in a wireless local area network by employing a WaveLAN interface card at each computer. Using the IEEE 802.11 CSMA/CA protocol, the exchange is

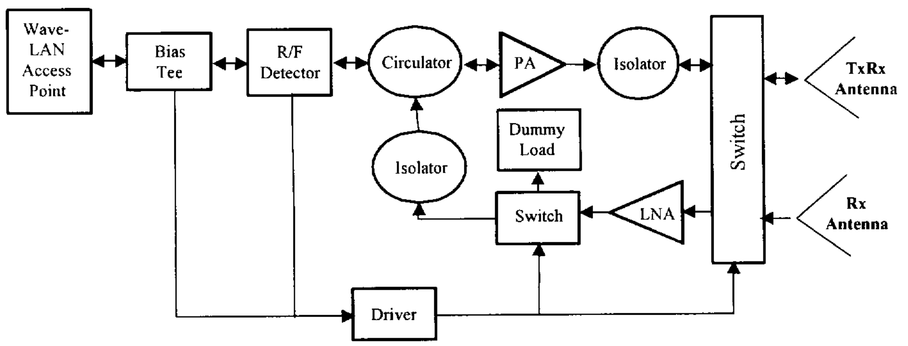


Figure 1. High-Power WavePOINT Assembly (HPWPA). Custom WavePOINT adapter to support higher output power and higher reception sensitivity.

very analogous to a wireless Ethernet. A transmission by any one unit is expected to be heard by all other units. This supports both collision avoidance and broadcast. However, a problem occurs when one node moves out of communication of one or more of the remaining nodes in the network. In this case both the broadcast and collision avoidance properties fail to be satisfied across the entire LAN. It is possible to configure a WaveLAN unit with special software that will allow it to provide an intranet relay function; however, this service has to be manually configured and it does not lend itself well to the mobility that will occur in military operations. A solution to this "hidden terminal" problem, when there is a common point of connectivity, is to use a WaveLAN access point operating under the Request-To-Send/Clear-to-Send (RTS/CTS) protocol. Here the CSMA/CA protocol is still invoked to control the RTS transmissions, and in the presence of hidden terminals, occasional RTS collisions are possible. The RTS/CTS protocol was used for the ELB ACTD.

The current WaveLAN card supports two data rates: 1 Mbps and 2 Mbps. Future upgrades are expected to support higher rates of 5.5 and 11 Mbps. Each WaveLAN link senses existing error statistics and adjusts its data rate to one of the available rates. All signaling is done at the 1 Mbps rate so that headers can be read without dependence on data rate. For communications from a manpack WaveLAN configuration to an airborne relay WaveLAN configuration, the minimum data rate is extremely high for supporting long range up-links. We will show later that power (battery, linear power amplifier quiescent power consumption) and gain limitations limit the length of the up-link to about 32 to 48 km (with very high demand on the battery pack). The sensitivity of the WaveLAN/IEEE receiver at 1 Mbps is -93 dBm for an average bit-error-rate of 10^{-5} or better.

WaveLAN provides a reliable link-level protocol when it senses that a reliable Transport protocol, such as TCP, is active. In this case each transmission is followed by an acknowledgement in the reverse direction. When an unreliable Transport protocol is used, as is common for multicast service, no acknowledgements are sent at the link level. End-to-end reliability can still be achieved by a higher layer protocol, independent of WaveLAN protocols. A problem was anticipated with the acknowledgement time out. The hardware counter used to track the time out normally expects a maximum delay comparable to a propagation range of about 6 to 7 km, which is far short of the propagation range required for this ACTD. Lucent was able to make a modification to the timer to support ranges up to 132.9 km.

Access Points. A WaveLAN access point is normally used to provide connectivity between a wired or fiber network and one or more WaveLAN wireless interface units. A standard WaveLAN access point, called a WavePOINT, provides two wireless WaveLAN interfaces and an Ethernet interface. The Ethernet interface can be connected to a wide-area network via a router. When a WavePOINT is used with two wireless interfaces, each interface is usually operated in a maximally separated RF band, and the antennas are separated to achieve as much isolation as possible. Ordinarily the Ethernet interface is used to interconnect multiple access points. However, one of the wireless interfaces can be used on each access point to provide wireless network connectivity among multiple access points. The ELB ACTD uses this mode, with a supporting beta software release from Lucent (WDS backbone beta-test software version 3.30), to support point-to-point WavePOINT interconnection on the airborne relay host platforms.

Although it was introduced too late for the ACTD, CellWave II, a third-party software product for Lucent's products, adds point-to-multipoint wireless interconnection of WavePOINTs through the use of a base station access point and multiple satellite access points. With the CellWave system, an access point may be a satellite to two base stations (one on each of its two radios) or may be a satellite to a base station on one radio and be a base station itself to its own group of satellites via its second radio.

A third wireless backbone system, not yet released, is the IEEE 802.11 Wireless Distribution system which will allow a WavePOINT with two radios to support simultaneously an Ethernet port, two cellular ports on the same or different sub-bands, and five WavePOINT links distributed arbitrarily between the two sub-bands.

End-to-end communications between computers interfaced to WaveLAN and WavePOINT devices are conducted by layer-two (link-layer) switching via the Spanning-Tree protocol. Hence, WaveLAN "networks" desire a flat network structure in which all attached computers have the same subnet address (but different host address). This has the advantage of simplicity and eliminates the need to run high-overhead router-to-router protocols like OSPF. However, it has the disadvantage that it does not scale well with network size. Each time a node attaches or detaches from the wireless LAN, the spanning tree has to be re-run. This process may become slow and unstable as the network size grows.

Roaming of a WaveLAN from one access point to another is also supported if the network structure stays flat (i.e., a single subnet). If a router is inserted anywhere in the network, spanning trees and roaming will not be supported by intrinsic WaveLAN functions beyond the first router.

VRC-99A. The VRC-99A is a militarized radio offered by Marconi Aerospace. It operates in the 1300 to 1500 MHz band (various options available), uses a spread-spectrum waveform to provide antijam and low-probability-of-intercept features, and provides U. S. Military Type-1 COMSEC. Error control is exercised via Cyclical Redundancy Check (CRC), and Viterbi/convolutional coding. Channel access can be controlled by a variety of TDMA, FDMA, and CDMA techniques. The present VRC-99A units occupy 16.4 cm^3 and weighs 11.4 kg. The receiver sensitivity at 1.25 Mbps is -94 dBm. The standard transmitter is rated at 10 W but an optional 50W amplifier is available. RAKE filtering is used to provide multipath mitigation up to 6 microseconds and Doppler correction exists up to 300 knots (570 km/hr). VRC-99A radios form a communications network capable of transferring data from host computers or groups of host computers connected in a LAN to remote hosts or remote LANs.

The VRC-99A has an embedded IP gateway that functions as a primitive router and provides interfaces to Ethernet, RS-422 serial bus, and the wireless medium. Normally, the Ethernet port is used to host one or more computers. Separate IP addresses are required for each of these ports and a unique subnet identifier is required for the wireless network and the wired LAN attached to each radio. Hence, there is a very fundamental difference between the networking scheme for this radio and WaveLAN. The VRC-99A expects unique subnets at each radio; whereas, WaveLAN expects a single subnet everywhere. To enable the two systems to be connected in a network via routers, the system engineers were forced to create WaveLAN subnets. Consequently, roaming of a WaveLAN radio from one subnet to another was not supported in the demonstration.

Because of a security concern on a previous large program that used the VRC-99A radios, the embedded VRC-99A router was configured so that it did not support the Address Resolution Protocol (ARP), which converts IP addresses into Medium ACcess (MAC) (hardware) addresses. Consequently, the ARP protocol could not be run over the ACTD network. Instead, this information was passed to the network nodes via IP tunneling.

SATCOM. SATCOM systems were employed in the demonstration to provide communications to large nodes when aircraft were not available. The technical objective for this aspect of the demonstration was to show that it was possible to provide high-rate data service to the forces with relatively small antenna apertures. This is quite feasible at Ku Band with antenna apertures in the range of 1 to 2.5 meters diameter. From a naval perspective, however, there is a complication with Ku-band in that there is no satellite coverage in the deep ocean areas; hence, ships do not normally carry Ku-band antennas. On the other hand, there is usually good Ku-band coverage in the littoral areas. The strategy employed for the ACTD was to use SHF SATCOM from the ships to a central hub, which provided a gateway to a Ku-band transponder on the TELSTAR 5 satellite. The entire capacity of a single 27 MHz transponder was purchased to support four simultaneous carriers at 2.048 Mbps. These carriers were used to support full duplex service from the hub to two HMMWVs equipped with 1.2-meter parabolic antennas. The 1.2-meter antenna had a gain of about 42 dB at 12 GHz. Less than 15 watts was required to maintain an average BER better than 10^{-11}.

Communication Platforms:

P-3. Two Navy P-3 aircraft were instrumented with WaveLAN and VRC-99A systems to provide a wireless network that extended from the Navy ships to the depths of the battlefield. Because of the small size, weight, and power consumption associated with WaveLAN, it was best suited to linking down to small terrestrial nodes such as a Marine or dismounted soldier with a manpack radio configuration. While WaveLAN was also capable of supporting long range links between aircraft and between aircraft and large terrestrial nodes such as mobile command posts, there was reason to believe that the VRC-99A, with its lower RF carrier and higher transmit power, would be better.

A major engineering concern for the aircraft was the antenna design. Link power budget calculations for WaveLAN indicated that a total link gain of 15 dB was required (assuming zero implementation loss) to support a link range of 160 km with a 9 dB system margin at a data rate of 1 Mbps. Increasing the data rate to 2 Mbps would either remove 3 dB from the margin or require an additional 3 dB of antenna gain to retain the same system margin. For the case where the communication link terminated at a dismounted soldier, the antenna gain at the terrestrial end of the link would be very small. Consequently, the design should attempt to place as much gain on the aircraft antenna as possible. An azimuthally-symmetric gain distribution is desired so that antenna steering of multiple beams is not required on the aircraft. Consequently, for long ranges a toroidal or donut pattern, similar to that produced by a stacked dipole, is desirable. However, because of the severe schedule constraints imposed on this demonstration and because of ACTD rules that forbid component development, we were not able to custom-develop a high-gain toroidal pattern antenna package for the P-3. The best available "stick" antenna, with advertised 6-dB gain, exhibited a 5 dB gain when mounted on the P-3. This type of antenna was mounted on the roof of the aircraft and also at two locations with 0.6 meter separation on the belly of the aircraft just behind the wings. The multiple antennas were installed to experimentally investigate the advantages of antenna diversity and alternate positions. Unfortunately, the lower two antennas were desensitized by electromagnetic interference believed to be caused by hydraulic pumps internal to the aircraft. Consequently, only the roof-mounted antenna was useful for receiving. Because of its low gain, a propagation range of 160 km could be achieved only by using a antenna gain of 9 dB at the terrestrial end of the link.

For the situation where the aircraft was nearly overhead of the participating terrestrial nodes, dual Cushcraft model SQ2303P "squint" patch antennas were used. One antenna was used to transmit while both antennas were used for spatial diversity during reception. These antennas were mounted on the belly of the P-3 approximately 6 meters behind the end of the wing. Because of its vertical polarization, this antenna provides no coverage directly below the aircraft but because of its "squint" characteristic, it provides lobes with 3.5 dB gain at elevation angles lower those covered by the stick antenna. The stick antenna and the patch antenna were connected to two different WaveLAN cards contained in two different WavePOINT access points as shown in figure 2.

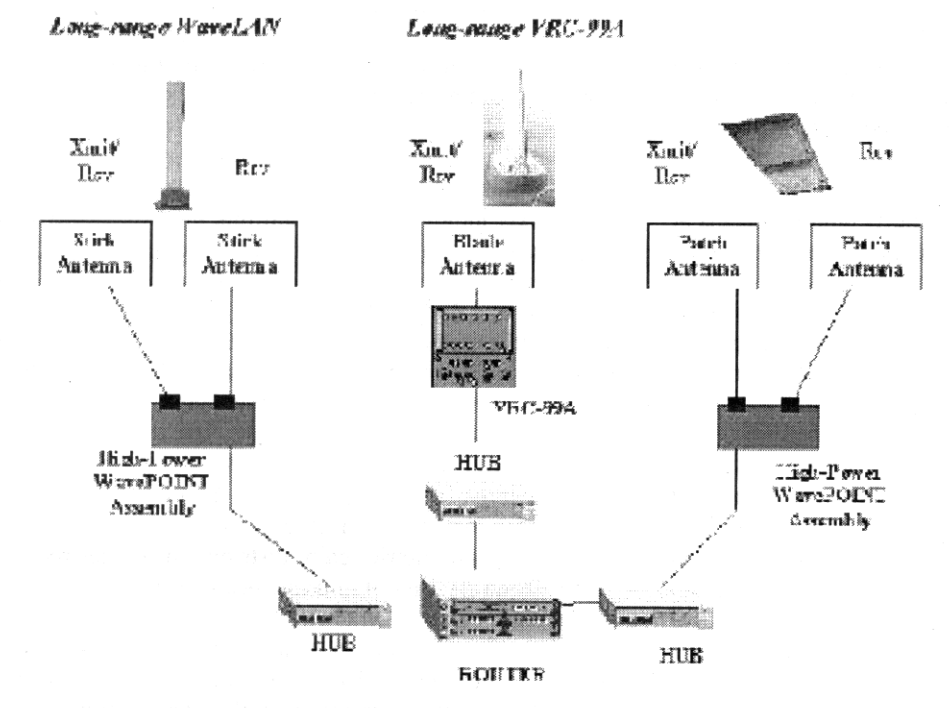


Figure 2. System Configuration for the P-3 and Crownair aircraft. S ant = stick antenna; P ant = patch antenna, B ant = blade antenna.

The VRC-99A antenna consisted of a Trivec Avant monopole "blade" antenna that provided approximately 3 dB of gain. This antenna was located on the belly of the P-3 approximately 10 meters behind the nose of the aircraft.

The interconnection of communication components on the P-3 is shown in figure 2. The box identified as HPWPA is the high-power WavePOINT assembly previously discussed in figure 1. One HPWPA was used with stick antennas for long-range communications and a second HPWPA was used with the patch antennas for shorter range communications. Ethernet hubs were used as a convenient way to interconnect components and to provide on-board access points for testing and to support voice communications to the airborne system engineer.

Separate subnet addresses were used for each router port; hence, it was not possible in this architecture for a WaveLAN to roam from the long-range to the short-range area of coverage without assigning a new IP address.

Crownair. Two commercial Crownair Partenavia P68C aircraft were also instrumented with WaveLAN and VRC-99A communications equipment. This style of aircraft had an overhead wing and dual wing-mounted propeller engines. There were fewer restrictions associated with antenna mounting on this aircraft. Consequently, a 9-dB stacked dipole stick antenna was mounted on the topside of the aircraft in the vicinity of the wing. A 6-dB stick antenna was mounted on the bottom side, approximately 2/3 the distance from the front of the aircraft. The blade antenna for the VRC-99 was mounted on the bottom side below the cockpit. Components were interconnected as in the same manner as for the P-3.

HMMWV. There were two types of HMMWV configurations. Two HMMWVs were equipped by NRL with 6-W WavePOINT and 1.2-meter parabolic antenna Ku-band SATCOM systems; these systems were interconnected by a router. The WavePOINT was used to provide a wireless connection to EUTs in the vicinity of the HMMWV. The remaining HMMWVs were instrumented with 6-W WaveLAN and 10-W VRC-99A. In this case one HPWPA was used with a 9 dB dipole antenna to support communications to the aircraft or to a surrogate-aircraft relay and a second HPWPA was used on a different WaveLAN RF channel to support either a wireless terrestrial LAN or longer range connection to another HMMWV. The VRC-99A could also be used to connect to aircraft or support a terrestrial connection to another HMMWV.

Ships. WaveLAN and VRC-99 antennas were installed on the ship mast yard-arms on the USS Coronado (AGF11), USS Bonhomme Richard (LHD6), and the USS John Paul Jones (DDG53). The WaveLAN stick antenna, low-noise-amplifier, and T/R switch were also located on the yard arm; the power amplifier was located below deck. All transmit/receiver electronics for the VRC-99A were located below deck.

End-User Terminal (EUT). The computing, display, and communication system used by the individual Marine was called the EUT. This equipment was built into a body harness containing a GPS receiver and a Toshiba model 100CT Libretto Laptop computer with a WaveLAN interface that was modified to support dual shoulder-mounted stub antennas with a 6-W external power amplifier. By positioning one antenna vertically and the other horizontally, reception was relatively insensitive to body position. The Libretto had a 166 MHz MMX processor, 64 MB of RAM, and supported two Type-II PCMCIA card slots. Non-rechargeable lithium and standard SINCGARS batteries were used for power.

Propagation Considerations: The calculated propagation range is shown in figure 3 for both the WaveLAN and VRC-99 systems as a function of power (delivered to the antenna terminals), antenna gain, and system margin. The calculation assumes zero implementation loss and neglects any performance improvement provided by the added low-noise amplifier. The antenna gain parameter is the product of the gains at both ends of the link. The system margin is a design safety factor to allow for additional loss mechanisms such as uncompensated fading. The link will be unreliable at zero margin and will become increasingly more reliable as more system margin is realized. If one considers a system margin of 3 dB to be an absolute minimum for consideration, then the maximum ranges are shown in Table 1.

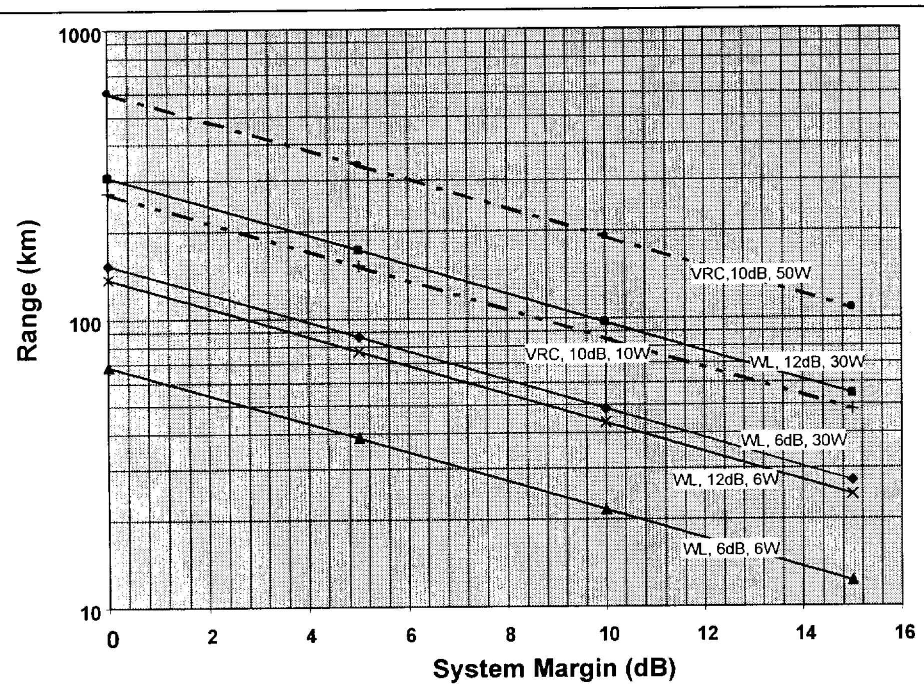


Figure 3. Calculated range for WaveLAN and VRC-99 systems as a function of power, antenna gain and system margin. The data rate was 1.0 Mbps for WaveLAN and 1.25 Mbps for the VRC-99

System	Power (Watts)	Gain Product (dB)	Range (km)
WaveLAN	6	6	48
WaveLAN	6	12	96
WaveLAN	30	6	107
WaveLAN	30	12	214
VRC-99A	10	10	188
VRC-99A	50	10	420

Table 1. Calculated range for WaveLAN and VRC-99A at 3 dB system margin.

User Applications.
Applications supported for this ACTD included networked voice via an application called IVOX; the Common Tactical Picture displayed via InCON™ software; CU-See-ME for tactical distributed collaboration; video teleconferencing and white-boarding; Advanced Combat System (ACS) for generating the Common Tactical Picture and supporting planning, analysis, and response execution; and Land-Attack Warfare System (LAWS) for coordinating fire support. CU-See-ME is a well-known commercial software package that allows video teleconferencing via desktop and laptop computers and a reflector. Information on CU-See-ME can be found at http://www.trey.com/cuseeme/. Information on InCON can be obtained at http://www.systech/sri.com. Interactive VOice eXchange (IVOX) is a tactical, networked, voice-over-IP application that was developed at the Naval Research Laboratory (NRL) (ftp://manimac.itd.nrl.navy.mil/Pub/ivox/). It was used in this project to provide a voice capability over WaveLAN networks that would otherwise be data only. IVOX provides selectable vocoding rates of 600, 800, 1200, 2400, 13,000 and 32,000 bps and provides a dynamic, self-optimizing buffering scheme that automatically adjusts for end-to-end network delays. IVOX supports multiple calls simultaneously, including point-to-point and group or conference calls. Group calls can be supported by multicast, multiple unicasts, or a combination of both. IVOX also contains an interface to the RSVP signaling protocol that can be used to request flow-based quality-of-service reservations for packet-switched networks with RSVP-enabled routers. Although IVOX was

initially selected for supporting WaveLAN-to-WaveLAN voice, a gateway is presently in development to support WaveLAN-to-telephone communications. ACS and LAWS are US military software used for command-control and combat direction and will not be discussed in this paper.

Security: The original intent for the ACTD was to include communication security in the demonstration. The system design included the use of Network Encryption System (NES) encrypters at command posts. Native WaveLAN encryption was intended to be used to cover sensitive-but-unclassified traffic to non-mechanized forces. At the time of initial planning, WaveLAN/IEEE was not available and the older WaveLAN units included Data Encryption Standard (DES) encryption, which is approved for sensitive-but-unclassified U.S. military use. When WaveLAN/IEEE products were released, no encryption was available for the initial products that were purchased for the ACTD. In addition, Lucent decided to support an IEEE security standard of RSA RC4 for WaveLAN/IEEE. (This is an example of a potential difficulty of using commercial products for military use: the commercial vendor determines the security architecture and can change it to something that is not backward compatible). The VRC-99A supports U.S. military Type-1 encryption. It became clear to the system engineers that an integrated security solution and implementation was too difficult under the existing conditions and schedule. Consequently, no encryption was included and all participating shipboard systems were hosted on separated, unclassified networks.

System Demonstration: The ELB demonstration was fit into a larger set of operational exercises identified as "Kernel Blitz," which spanned the period of March 22 to May 14, 1999. During the period of 22 - 30 March a set of demonstrations associated with the "Urban Warrior" advanced warfighting exercise were conducted. From 19 - 30 April a supporting arms coordination exercise and amphibious landing occured. During 5 - 16 April, the ELB demonstration was conducted. During 3 - 14 May a fleet excercise occured. By including the ELB demonstration with the Kernel Blitz exercise, a one-time opportunity was possible to involve a large base of Navy and Marine participation. On the negative side, this forced the project into a "non-slidable" window of opportunity. There was no possibility to repeat events in the occurance of an unanticipated technical problem, and there was very limited time in advance of the exercise for extensive testing and force training. Another potential impediment was that much of the ELB equipment was also used in the Urban Warrior exercise and there were only five days between the two exercises to reconfigure the systems. Since the system did not support ARP, all IP and MAC addresses had to be manually reconfigured.

The littoral region chosen for the demonstration was Southern California, including the specific areas of the U. S. Marine Corps. Camp Pendleton; Spawar Systems Center, San Diego California; the Naval Air Facility at El Centro near the southwest corner of the Salton Sea; the Anza Borrego desert area west of the Salton Sea; and

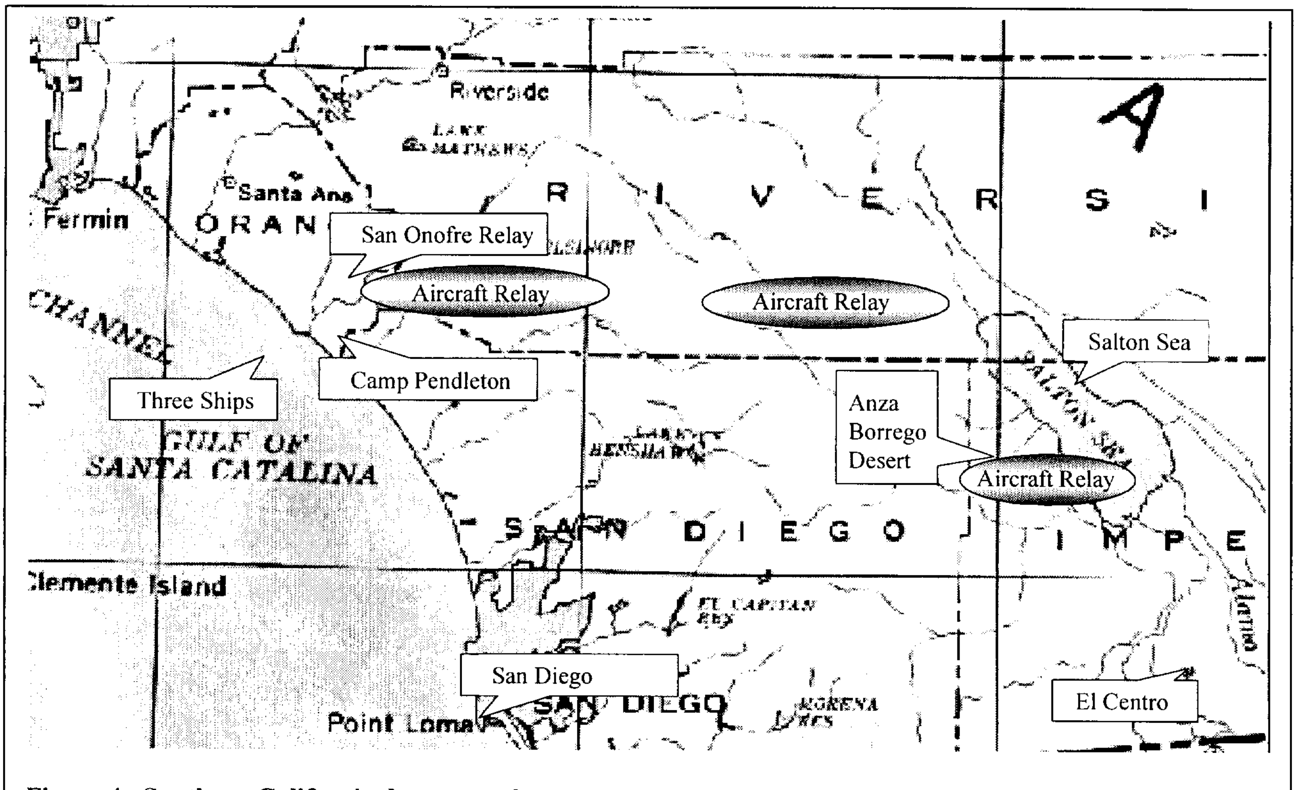


Figure 4. Southern California demonstration area.

Yuma, Arizona. Figure 4 shows this area with the exclusion of Yuma, which is east of the Salton Sea. Camp Pendleton is along the shoreline, midway between San Diego and Los Angeles, and consists of very mountainous, rugged terrain. Camp Pendleton was used to host the initial phase of reconnaissance and assault missions from support ships positioned off the coastline. These ships included the USS Coronado, the USS Bonhomme Richard, and the USS John Paul Jones. The USS Coronado was equipped with 30-W WaveLAN and 10-W VRC-99 communications equipment as well as SHF SATCOM. The sixth deck of the Coronado housed the ELB Enhanced Combat Operations Center (ECOC). The USS Bonhomme Richard carried similar communication equipment (but only 6-W WaveLAN) and hosted applications such as IVOX for digitized voice, ACS for providing the Common Tactical Picture, and a Collaboration workstation. The USS John Paul Jones (also 6-W WaveLAN and 10-W VRC-99A) was equipped with IVOX, ACS, a Collaboration Workstation, and LAWS. Only the USS Coronado had SATCOM service. Several High Mobility Multi-Wheel Vehicles (HMMWVs) were used as mobile command posts. Two of HMMWVs were specially configured by NRL with Ku-band SATCOM and 6-W WavePOINT systems. The other HMMWVs carried 6-W WaveLAN and 10-W VRC-99 communications equipment and hosted IVOX, ACS, and a Collaboration Workstation. The HMMWVs were repositioned at various locations throughout various "snapshots" of the demonstration scenario.

The Crownair and two P-3 aircraft were used to provide a wireless communication network that extends from the ships to shore sites at Camp Pendleton and to a "deep penetration" site in the Anza Borrego desert, northwest of the El Centro Air Station. Because of the long communication distances involved and shadowing effects in the mountainous terrain, more than one airborne relay was required to complete communications between the ships and various land sites. For those situations where an inadequate number of airborne relays could be put into flight, pre-placed "surrogate airborne relays" were activated at places such as San Onofre, in the Camp Pendleton area. These "surrogate airborne relays" were actually communication towers positioned on mountain tops and were configured to act as if they were an airborne relay. A disadvantage of these sites was the low elevation compared to the 4.6 km altitude achievable with the P-3, thereby experiencing more trouble with shadowing; an advantage was that the direction to the next site was usually well known so that directional antennas (9 to 14 dB) could be used to increase the link power-gain product for WaveLAN point-to-point links.

When airborne, the P-3 and Crownair platforms were used to complete the network between the ships at sea and various areas of operation at Camp Pendleton and in the Anza Borrego desert. Frequently, a racetrack pattern of 32 km by 16 km was flown. When aircraft were not available, the surrogate relay at San Onofre was used instead to extend the network to the Camp Pendleton area. In some cases the P-3 and Crownair relays were used in conjunction with the surrogate land-based relays to achieve a deep reach. Also, at times the WaveLAN and VRC-99 communications systems were used in a line-of-sight mode to support direct communications among ships.

A high-level description of the network configuration for the demonstration is shown in figure 5. The VRC-99A links were used as the primary long-range interconnection among ships and aircraft and from aircraft to HMMWVs. WaveLAN was used to connect from the aircraft to the EUTs and also to support terrestrial LANs centered around a HMMWV WavePOINT.

Test Results: The following range performance was observed during full-system tests prior to the final demonstration at aircraft speeds up to 740 km/hr.

WaveLAN (30Watts, 5 dB aircraft antenna, nominal 9 dB HMMWV antenna):
- 1 Mbps out to 185 km (becoming marginal)
- 2 Mbps out to 60 to 90 km (reliable link)

WaveLAN (EUT with 6 Watts and 3 dB antenna to 3.5 dB aircraft patch antenna,)
- 1 Mbps out to 37 km (becoming marginal)
- 2 Mbps out to 18 km (reliable link)

VRC-99A (10Watts, 3 dB antenna at each end of link)
- 625 kbps out to 204 km (reliable link)
- 1.25 Mbps out to 185 km (reliable link)
- 2.5 Mbps out to 148 km (reliable link)
- 5 Mbps out to 110 to 138 km (reliable link)
- 10 Mbps out to 55 km (reliable link)

Ku-band SATCOM (<15 Watts, 1.2 meter dish)
- 2.048 Mbps full duplex service between hub (2.4 m dish) and mobile command posts
- 99 % availability during demonstration

Summary and Conclusions: The ELB ACTD demonstrated that wireless networking technology can be used successfully to extend communications and data networking from ships in the littoral area to marines and dismounted soldiers at deep inland penetration. Individual wireless links were supported at ranges up to 185 km with a data rate up to 1.25 Mbps. By networking multiple links, the distance between support ships and terrestrial command posts and/or non-mechanized group leaders can be arbitrarily large. Most of the detailed link-range measurements were made prior to the final demonstration using a 30-W enhancement of WaveLAN and a 10-W version of the VRC-99A. For the final demonstration, a 50-W version of the VRC-99A was used for long-range connections between one P-3 aircraft and the San Onofre relay.

Because of the high power 50-W amplifier available with the VRC-99A (versus 30 W for WaveLAN) and because of the smaller free-space propagation loss at 1.3 GHz (versus 2.4 GHz for WaveLAN), the VRC-99A radio was favored in this demonstration for the longest range links. WaveLAN, with its superior size, weight, and power characteristics, was the only radio that could provide a direct network connection to a marine or dismounted soldier from the airborne relay. Hence, a heterogeneous network consisting of both WaveLAN and VRC-99 radios was used for the final demonstration. The WaveLAN uplink transmission from the foot soldier was limited to 6 Watts because of battery and human radiation-hazard considerations. Measured ranges corresponded closely with calculated ranges for a system margin in the range of 4 to 6 dB. Since the calculations assumed zero implementation loss, the 4 to 6 dB system margin may be considered to cover cable losses and antenna gain variations off the main beam direction. Doppler shift of the RF waveform caused no difficulty for either radio.

The stability of the network was marginal when extended across large numbers of nodes. Much improvement needs to be done in this area. Wide variations in success with establishing connections and time-slot synchronization were experienced with the VRC-99A radios. Our current conjecture is that the interaction between control and data packets produced some strange behavior. In particular, the network exhibited a tendency to break into geographically separated subnets with minimal data connectivity between them but with enough control packet exchange to cause problems with slot synchronization. Complete understanding of this problem requires further analysis.

Network services were also extended from the ships to terrestrial mobile command posts via satellite service. Ku-band satellite service over Telstar V was used to provide full-duplex service at 2.048 Mbps from a terrestrial hub-site to two HMMWV mobile command posts with 1.2-meter dishes. A 2.4 meter dish was used at the hub site. The use of Ku-band facilitated the use of tolerable-size field antennas. The terrestrial hub acted as a gateway to convert the SHF SATCOM transmissions from the ships to Ku-band for the field command posts.

Voice was supported over the network using a Naval-Research-Laboratory-developed application called IVOX. IVOX provided a "voice-over-wireless-network" capability that included the ability to support multiple, simultaneous point-to-point and group calls using any combination of unicast and multicast connections. An application called InCON™ was used to display common tactical picture information on a Toshiba Libretto laptop computer at the field units. CU-See-Me was used to provide multi-window laptop video teleconferencing to the field units. VTC was used for conferencing among major sites.

The "desired qualities (a - e)" listed in the background material, were met in this demonstration, with the following exceptions:

(1) Multicast was not supported because of the inability of the VRC-99 radios to support multicast addressing. This problem can be remedied by modifications to its embedded router.

(2) The "ability to automatically configure and

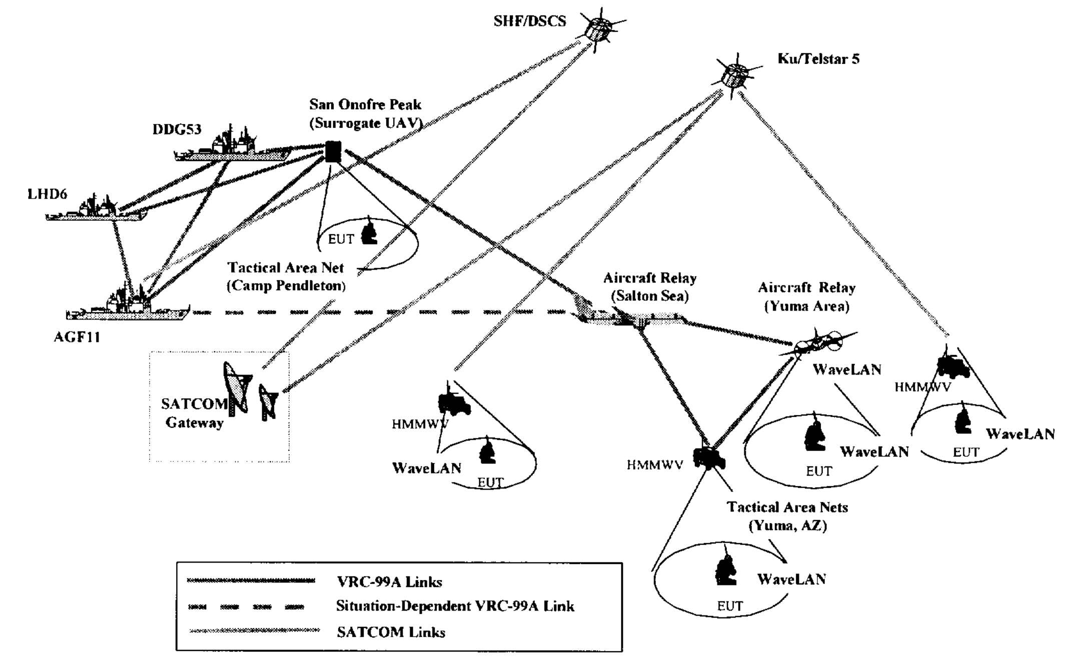


Figure 5. Network configuration for the ELB demonstration.

reconfigure the network topology and routing/switching databases as both users and airborne platforms move" was not achieved. In realty, roaming was not supported in the final architecture, and the necessity to manually reconfigure the association between IP and MAC addresses removed any possibilities of automatic reconfiguration. Nevertheless, the demonstration was a giant step forward toward the process of providing an over-the-horizon extension of the tactical network.

Many areas need improvement for the final demonstration in the year 2001.

* Roaming should be supported so that a user terminal can gain access to an alternate airborne relay without having to change its IP address.

* Address Resolution Protocol (ARP) and router-to-router protocols (such as MOSPF) should be supported to facilitate mobility and roaming.

* Quality of service provisions such as preferential queuing and/or flow reservations must be added to manage use of the limited, shared network capacity in a near optimal way. Otherwise, small but essential messages could get blocked by large file transfers.

* IP multicast was not supported. This difficulty must be eliminated because multicast service is essential.

* Encryption and suitable security architecture is needed for operational use.

* EUT connectivity to the aircraft must be achieved with less power drain on the manpack battery. This may require the use of communications equipment that supports data rates less than 1 Mbps. A 3-dB advantage could be gained by dropping to 500 kbps; another 6 dB could be gained by lowering the RF from 2.4 GHz to 1.2 GHz. This would also improve foliage penetration but would increase antenna length by a factor of two.

* Protocols need to be developed that allow the manpack (EUT) terminal to support access to both the airborne relay and terrestrial mobile, ad-hoc networks at the squad and platoon levels without having to reconfigure the terminal software.

The ACTD demonstration employed commercial wireless products such as WaveLAN and leading-edge military products such as the VRC-99, to allow the military operational exercise to experiment with the extension of data networks from Navy ships to the lower echelons of the battlefield. While this objective was well-satisfied in the demonstration, no one should assume that the commercial technology is ready for the battlefield. Commercial technology does not address the issues of denial of service (by jamming or any other means) or the unique needs during special operations for low-probability-of-detection waveforms. While commercial products, like WaveLAN, will soon include encryption for privacy, data integrity, and user authentication, it is still an issue of whether the commercial security will be accepted for military use. While employing commercial communications technology has the potential to bring an avalanche of advancement to military tactical communications, it also has the potential to provide an Achilles heel of vulnerability through the relatively inexpensive use of Information Warfare techniques. The commercial world controls the software and security upgrades for these products. What assurance does the military have that future upgrades will not introduce serious vulnerabilities? How long will the commercial vendor support old product versions? Will upgraded products always be backward compatible with older products that are in the military inventory? Will the military be willing to make frequent software upgrades for the complete inventory of equipment? Resolution of these and other similar concerns will have to be obtained before commercial technology is ready for use in life-threatening operations. Finally, there is the issue of spectrum ownership. Many of the systems being tested for tactical military use are in the ISM band. While the use of the ISM band is acceptable for experiments and demonstrations, it was not intended for routine military use. Therefore, unless the rules are changed, the commercial equipment will have to be modified to work in a different frequency band.

This demonstration was the first of the two major demonstrations under the ELB ACTD. A second demonstration is planned for the year 2001; however, there are no definitive plans in place at present.

Acknowledgements: The successful execution of the ELB demonstration is the result of the program management, planning, system engineering, training, and field implementation efforts of a very large group of people and organizations including the Office of Naval Research and their support staff, Spawar Systems Center, General Dynamics, Lucent Technologies, Naval Air Warfare Center, Marine Corps. Combat Development Command, USCINCPAC, Litton/PRC, Raytheon, SRI, and Marconi Aerospace to name a few. The Naval Research Laboratory played a roll in the initial planning and contractor selection, and served on the communications Integration Product Team with General Dynamics. We also provided IVOX, the Ku-band SATCOM services, and an engineer on a P-3 aircraft to configure and monitor the resident communications equipment. Much of the material reported in this document is a product of the team effort and should not be attributed solely or even largely to the Naval Research Laboratory. Much of the background material was drawn from the ELB Overview documentation that was made available to bidders on the Broad Agency Announcement.

PCS Study Plan

N. Young

NATO C3 Agency / Communications Systems Division / Techniques Branch
Communications System Division, PO Box 174
2501 CD Den Haag, The Netherlands
young@nc3a.nato.int

Dr. D. Yavuz

NATO C3 Agency / Communications Systems Division / Radio Branch
PO Box 174, 2501 CD Den Haag, The Netherlands
Davras.Yavus@nc3a.nato.int

Motivation

- **Traditional Military Comms Requirements**
  - Security, Mobility, Flexibility, Reliability, Alternatives
- **Recent Trends in Military Comms**
  - Blurring of Tactical & Strategic Comm Requirements
  - Integration of Voice & Data networks (The Internet)
  - Move away from Stove-Piped single purpose systems toward multi-purpose IP networks
  - Move toward COTS solutions
- **Attractive Features of PCS Technology**
  - Relatively Low Cost
  - Interoperability Potential with NATO and non-NATO organisations
  - Some based on open standards with Global Availability
  - Earth Coverage of satellite-based systems (ICO, GlobalStar, Iridium etc.)

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

NATO C3 AGENCY

Basis of Study

- **Scenario** developed for GSM study (update as required)
- **Identify** all technologies within 5 years of commercial service
- **Evaluate** those which will offer service by 2000
- **Focus** on 2-3 space-based and 2-3 terrestrial systems representing inherently different approaches (e.g. LEO vs. MEO vs. GEO satellite systems and CDMA vs. TETRA terrestrial systems)

NATO C3 AGENCY

Study Methodology

- **Derive Comms Requirements, Operational & Technical**
  - **Flow from identified scenario and security architecture**
- **Develop Evaluation Criteria**
  - **Mandatory & Optional/Desirable Capabilities & Features including security**
  - **Scheme for ranking candidate systems**
- **Identify candidate technologies which meet study criteria from**
  - **Literature**
  - **Web**
  - **National agencies & laboratories**

Methodology (cont.)

- Evaluate each candidate technology against requirements
- Produce Study Report including:
  - Overall Findings
  - Recommendations for potential adoption by NATO
  - Recommendations about acquisition strategy (i.e. lease vs. buy options)
  - Summary of emerging/promising technologies which merit future study

Task Breakdown

1. Overall Umbrella Task
2. Scenario, Requirements & Evaluation Plan
   - Update scenario developed for GSM investigation
   - Include all phases of scenario from CJTF/Peace Making to Multinational Force/Peace Keeping operations
   - Comm requirements (including security) to be developed based on scenario (i.e. J6 considerations)

Task Breakdown

3. Security Architecture
   - Link (RF) encryption
   - End-to-end security requirements for voice and data
   - Interoperability with other secure voice systems
   - User identification & authentication including system management requirements
   - Security requirements at interfaces & gateways to other NATO & non-NATO voice & data networks

Task Breakdown

4. Technology Investigation & Evaluation
   a) terrestrial systems,
   b) space-based systems

- Select candidate technologies
- Compile details sufficient for evaluation against defined comms requirements
- Conduct evaluation in accordance with Evaluation Plan
- Draft respective sections of Final Study Report

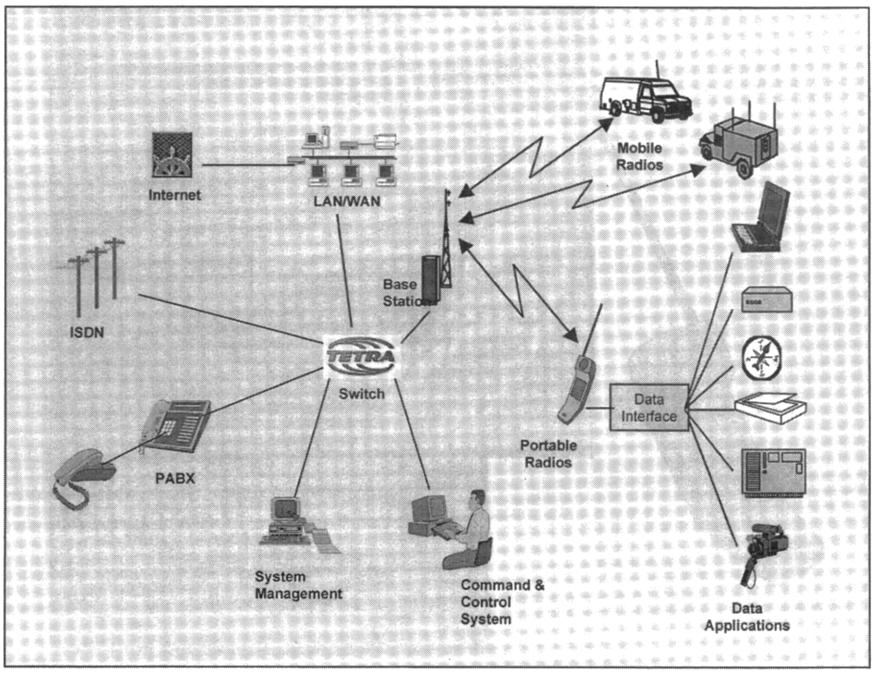
Hands-On Evaluations

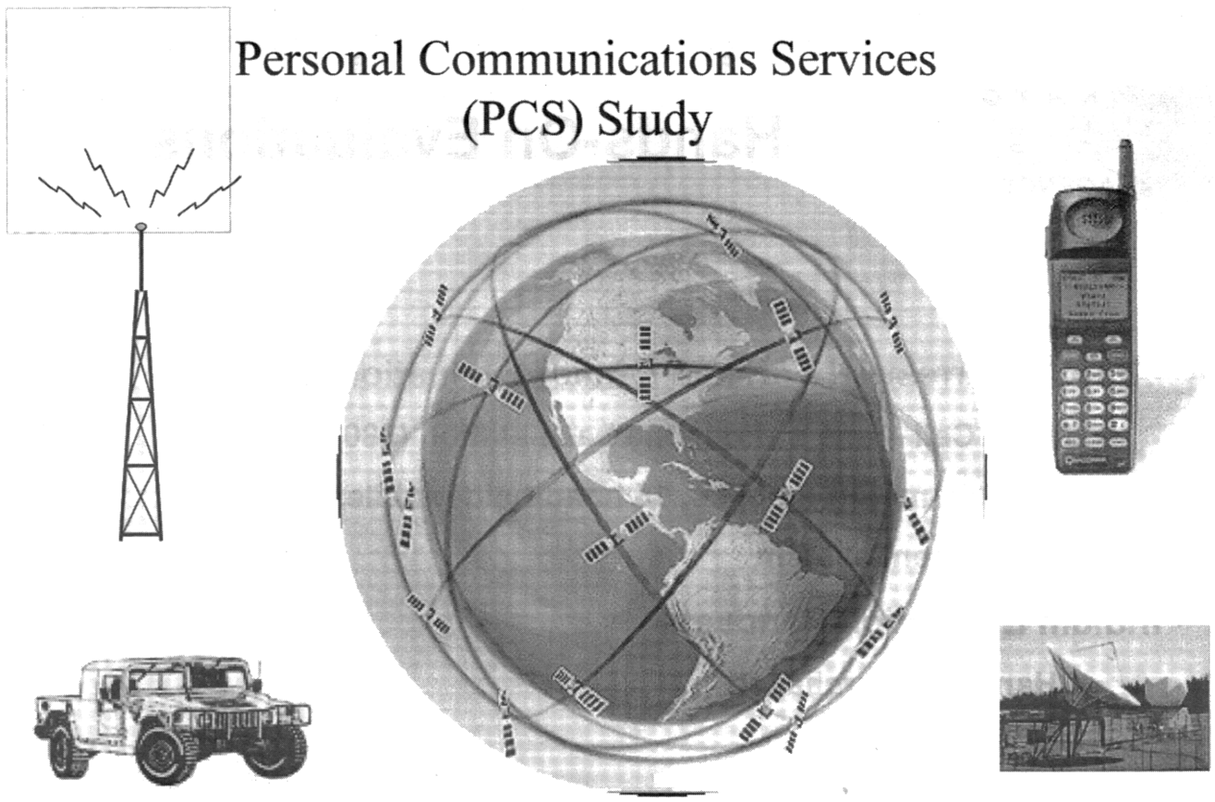
- TETRA (TErrestrial Trunked RAdio) Evaluation System
  - Single Cell – one switch, one base station (380-400 MHz)
  - 5 Vehicle-mount mobiles, 15 hand-held mobiles
  - Management system
- Iridium Evaluation Handsets
  - 4 Handsets purchased
  - Ongoing trials

TETRA Features

- Direct Mode Handset – Handset when out of range of base station
- Inherent relay capability in all mobiles
- Priority & Pre-emption of calls
- Half-duplex Group Calls (simulate combat net radio)
- Robust RF link encryption provision for end-end security features
- Data rates up to 28.8 kbps
- Simultaneous voice and data
- Operates in 380-400 MHz (NATO UHF) band
- International (ETSI) open standard

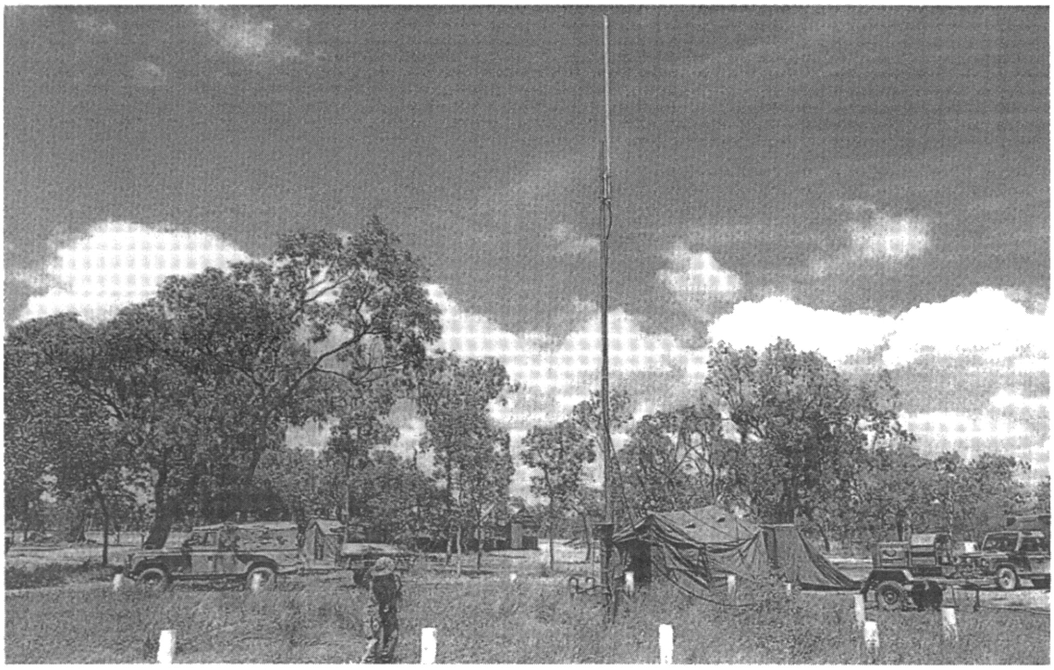
Terrestrial Trunked Radio (TETRA) Evaluation System




Personal Communications Services
(PCS) Study


NATO
C3
AGENCY
Iridium

Australian Army Mobile GSM



Tactical Mobile Communication using Civilian Standards

a preliminary study 2000-2003

Dipl.-Ing. Dirk Müller

Deputy Head of Section, Tactical Mobile Networks
BWB-FE II 2
Ferdinand-Sauerbruch-Str. 1
D-56073 Koblenz, Germany

Summary: This paper describes a preliminary study that starts in 2000 ending in 2003.

The task performed in this study is to evaluate the usability of civilian mobile communication networks for military purposes. Military add-ons needed to meet military requirements are to be identified. Technical prototype solutions will be implemented and evaluated in an experimental system.

Finally, the results of this study will give advice for the procurement of a tactical mobile communication system in 2005+.

History: The Netherlands and Germany developed successfully a military mobile tactical communication cellular system based on EUROCOM during 1986 - 1997. This SCRA-System (Single Channel Radio Access) was developed as a highly mobile extension of the tactical networks AUTOKO 90 (Germany) and ZODIAC (The Netherlands) and fulfilled all military requirements. Main characteristics were: ECM hardened by frequency hopping, fully encrypted connections (up to SECRET), vehicle mounted, typical communication range approx. 20 km from Mobile Subscriber Terminal (MST) to Radio Access Point (RAP), direct mode (MST-MST) within approx. 7 km.

Due to budgetary reasons the German Chief of Staff (Army) decided to delay the procurement until 2005. A situation evaluation has shown, that it seems appropriate to look for a new solution based on modern COTS (commercial of the shelf) systems instead to conserve technology from the 80's. For this reason, the German secretary of defense decided to terminate the SCRA program in 02/99.

A new approach with COTS: The use of COTS based systems has significant benefits but disadvantages also. Main benefits are: COTS systems are available at a reasonable price on the market at any time, no or low in-house logistics are necessary and it is quite easy to participate in the technical progress on the civilian market. Main disadvantages are: not all military requirements are supported necessarily. Civilian standards need not be implemented in full scale by a service provider or equipment manufacturer. There is nearly no chance for the military to influence migration to new systems. Instead, one gets fully dependent on the civilian market.

The disadvantages show, that in some areas military add-ons will be necessary to meet military requirements on tactical communication systems. Unfortunately any change in COTS systems leads to a special solution for the military. This may cause a totally loss of the COTS benefits, especially in cost, if done improperly.

Therefore it is essential to evaluate the military requirements for mobile communications very carefully in order to identify the areas where add-ons, if any, are needed. When implementing add-ons it is essential to do it with minimized effects on the civilian standard and technology to preserve the COTS benefits.

Mobile communication is a common need for all military services. Up to now, Germany's services have tried to procure independent systems that meet their special services' needs. Another difference has been made between "administrative communication" and "tactical communication" that have their own services' budged. This budged separation combined with the service-dependent requirement evaluation leaded into the development and procurement of different systems within the armed services. An example for this situation is the terminated SCRA program that was developed as a special military solution without COTS for the army only.

Administrative communication means communication between barracks and other facilities and e.g. officers on TDY or on leave, mainly under peacetime conditions. That means that there are no unfriendly or hostile activities that may damage or jam the system but the risk of passive information loss by signal intelligence (SIGINT).

Tactical communication means communication on the battlefield in an unfriendly or hostile environment with great risks of active measures by the enemy to destroy, interrupt and jam communication systems in addition to SIGINT.

The basic needs for mobile communications are common for all military services: The mobile communication system shall provide Voice/Fax, Data (slow; e.g. e-mail and information services), Data (fast; e.g. video, sensor data), secure communications in all areas, where people may be deployed. The main

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

differences lay in the area that needs to be covered and in the environment that affects the performance of the system. Because of the common aspects there is no need for different communication systems for the different services in principle. In fact there is a big chance, that a common technical solution that meets the requirements of all services based on COTS is possible.

First estimate for military scenarios: Service independent needs show the following aspects: Access to senior people "worldwide" in any region, interrupt free access to information and databases, access to military networks (e.g. EUROCOM systems), access to external and civilian networks, man portable, usable in a vehicle, encryption of sensitive information, based on COTS.

For administrative purposes in a friendly environment any of the civilian systems is applicable (e.g. GSM in populated areas, IRIDIUM/GLOBALSTAR worldwide). No enhancements or add-ons are necessary due to the basically friendly environment. Military owned infrastructure may be reduced to mobile terminals if civilian service providers are used.

For tactical communications there is a need to look at different scenarios: e.g. humanitarian task and peacekeeping, peace enforcement.

The humanitarian task and peacekeeping scenario is quite similar to the administrative scenario but civilian infrastructure may be damaged or simply missing. Therefore the whole infrastructure for a communication system needs to be imported and built up. The environment may be unfriendly but not hostile (SIGINT but no fighting action). The area is well defined and a cellular system, once operating, is basically stationary.

The peace enforcement scenario means: hostile environment with fighting action, no civilian infrastructure available and massive ECM, SIGINT and COMINT by the enemy. The whole communication system needs to be highly mobile due to the ongoing area changes and the physical threats. As civilian systems are not specified for situations like these mainly this scenario will show the needs for military enhancements of today's COTS systems.

Possible fields for military add-ons: No military add-ons are needed in the administrative scenario. For the humanitarian task and peacekeeping scenario information protection measures (Counter COMINT) may be necessary. These can be dealt with COTS products for end-to-end encryption devices without affecting a mobile COTS communications system, if necessary. Due to the fact, that a civilian infrastructure may not be available, all equipment to build up a mobile communication system need to be brought into the area. Modern COTS systems are very narrow and can be loaded on trucks and operated from mobile shelters. Therefore there is a good chance to use a COTS System in this scenario.

The peacekeeping and fighting action scenario will bring up the biggest challenge for military add-ons. Possible fields are: deployment (mobility) in a highly mobile battlefield, camouflage and concealment, withstanding sabotage (e.g. software intrusion) and destruction, reduction of weapons' effects (Anti Radiation Missiles, Laser and Radio Frequency weapons, NEMP), withstanding SIGINT, Counter C3 measures (e.g. information gathering) and ECM. A basic overview of military requirements is given in EUROCOM D/0 (94) and the upcoming results from the TACOMS POST 2000 program.

Interesting standards: Current COTS standards that are of interest for military use both on the battlefield and for administrative use are GSM + enhancements, TETRA-25, UMTS, IS-95. These may be combined with a low earth orbit (LEO) satellite system like IRIDIUM or GLOBALSTAR.

First results: A study by IABG (German industry) on behalf of the German Armed Forces Staff advises for administrative purposes a combined GSM/SATCOM system.

GSM needs enhancements in case of deployment in a mobile battlefield for damage resistance, ECM, SIGINT, infrastructure mobility (high mobility in a mobile battlefield) and interfaces to existing tactical networks (e.g. EUROCOM).

It seems reasonable that other standards like TETRA-25, UMTS and IS-95 need enhancements in the same areas. This will be evaluated within the study.

High data rates, that exceed 9.6 kbit/s (GSM) will be necessary in future. As this requirement is not a unique military task but also of great interest on the civilian market it may be assumed, that the COTS systems will be enhanced accordingly without the need for a military add-on except for robust error correction measures.

Another study concentrates on end-user communication equipment. The bilateral Multirole Multiband Radio Program is concerned with a communication device, that works for several purposes like CNR (combat net radio), VHF, HF and civilian cellular networks' waveforms like GSM or TETRA 25. If the armed forces are equipped with such a multipurpose radio, there is a good chance, that there is no need to have a military enhanced COTS cellular network. It may be possible, that in peacetime or in a humanitarian scenario a MMR equipped person uses the infrastructure available (civilian or military owned COTS system) and simply switches into a strictly military tactical mode if battlefield conditions arise.

Resumée: The work performed in the upcoming study will be concentrated on the evaluation of military requirements for both tactical and administrative purposes against the characteristics of COTS cellular networks.

Areas for possible enhancements will be identified with respect to the different scenarios as described before. Technical solutions to implement enhancements with minimized effects to COTS standards will be implemented by prototyping with respect to estimated lifecycle cost and tested in an experimental system.

A possible outcome may be a military enhanced COTS system that is applicable for all scenarios for all services. On the other hand a COTS system without or little enhancements (e.g. interfaces to EUROCOM systems) may be the key to the communication needs for most scenarios with a very specialized equipment for certain situations like MMR.

One way or the other: The time has come to leave special military solutions in favor of COTS products to preserve the taxpayers' money, to shorten development cycles and to participate on the technical progress in modern communications as part of standardized equipment within the armed services.

FE II 2

MANAGEMENT SYSTEM FOR MOBILE COMMUNICATION NETWORK USED IN THE POLISH ARMY

Krzysztof Mamcarz, Marek Małowidzki,
Robert Matyszkiel, Konrad Szczygielski
Network Management Department
Military Communication Institute
05-130 Zegrze, Poland

Abstract

This paper describes a general conception and implementation details of the Mobile Communication Network Management System (MCNMS) developed for Polish Army. The prototype of MCNMS has been finished this year and checked in the field conditions during exercises in November '98.

The following topics will be considered in the paper:

- standarizations and recommendations of international organizations and groups taken into account while building conception and prototype of MCNMS,
- general characteristics of MCNMS in terms of tactical aspects – hierarchy of management system, exchanging reports between signal officers,
- the physical structure of MCNMS including its hardware elements: management shelters, management workstations, mediation devices up to network elements,
- interfaces and communication protocols defined and used in MCNMS,
- the structure of the management application,
- interoperability issues
- realized and planned management functions,
- some tested functions from field exercises,
- conclusions.

1 Introduction

The Mobile Communication Network Management System (MCNSM) has been developed in Military Communication Institute in Zegrze, Poland. The MCNMS is designed for signal officers who are reaponsible for proper organizing and controlling operational aspects of the tactical network. The MCNMS supports signal officers in taking into account the tactital situation and the network resources to make efficient decisions. Basic supported areas are the following:

- network planning,
- contolling the deployment of the network,
- network resources monitoring,
- remote configuration,
- raport exchange.

The project took into account ISO and ITU-T standards, especially those related to Telecommunications Management Network (TMN, see [TMN]). However, full compliance with TMN architecture and protocols used within TMN was not possible because of hardware and network limitations. Originally, the MCNMS was developed for circuit-switched low rate (with basic channel of 16 kbps) network with many different network elements coming from different vendors and offering diverse and nonstandard interfaces for management. At present, IP packet nodes are being deployed; this enables the use of TCP/IP protocol stack for communication of management system components and, of course, causes noticeable performance boost.

2 Management System Physical Structure

The physical structure of the MCNMS is presented in Fig. 1.

Hardware components of the MCNMS are as follows:

- two **Management Shelters (MS)**, one of them placed in the main command post and the second placed on the rear command post as a reserve. The MS is designed to be the working place of three signal officers and includes positions for network planning and monitoring.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

- **Management Workstations (WS),** NT-based computers with management software.
- **Portable Workstations (PWS),** which are intended to be used by the signal officers for local management of the base network or access nodes. PWSs allow applying hierarchical management structure – the WSs don't need to monitor local directly the local node and its Mediation Devices (see below) but instead they receive higher-level management data from PWSs.
- **Mediation Devices (MD),** which are management system agents placed inside managed shelters. They control Network Elements (NE) such as switches, radio-relays, encryption bulk units, modems, multiplexers etc. The MDs realize following main functions:
  - providing means for remote management of the NEs,
  - providing some kind of unification of the management of different network elements,
  - event collecting and generation on the base of continuous polling of the managed devices,
  - low-level event filtering,
  - event notification for WSs and PWSs.

  The MD is implemented on the base of industrial computer with QNX real-time operating system.
- **Network elements (NE),** i.e. all telecommunication devices equipped with management interfaces – commutation devices, transmission devices etc. Since these interfaces are mostly local (typical RS-232C connections) and nonstandard, the MD is required to enable remote management and some value-added fuctions mentioned above.

The transport network for management information is the managed network itself. There's no separated network resources for the management system. The communication between management system elements (WSs, PWSs and MDs) is possible in both circuit and packet-switched modes. The latter uses IP network implemented on the top of circuit-switched connections.

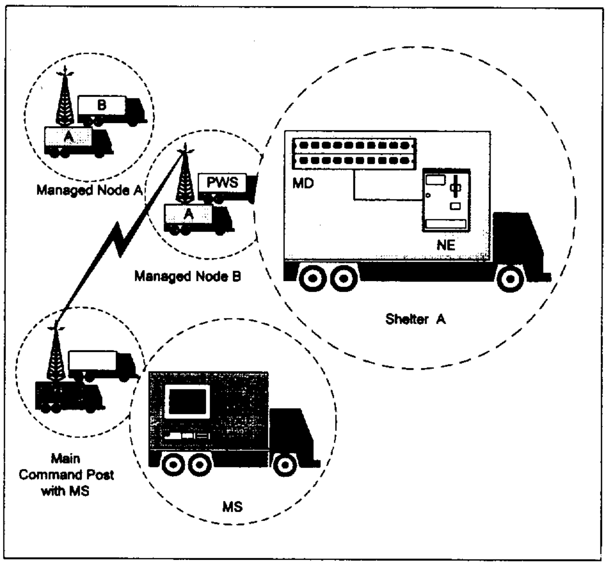


Fig. 1: The physical structure of the MCNMS

3 Management System Architecture and Interfaces

The MCNMS architecture follows the physical structure. Main components and interfaces used for communications between them are presented in Fig. 2.

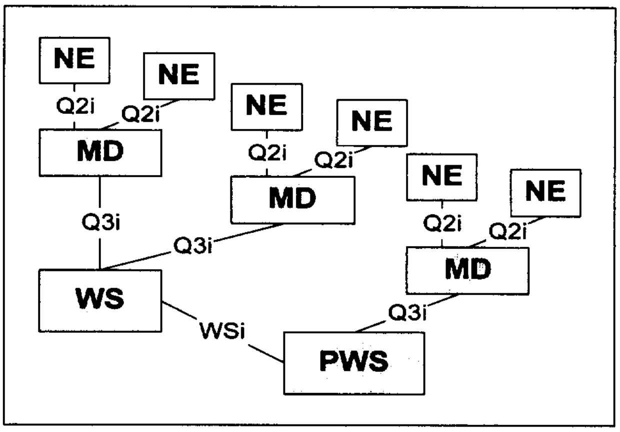


Fig. 2: Management System Architecture and Interfaces

The following interfaces were defined within the management network:

- Q2i – interfaces between MD and NE. They correspond to Q2 interfaces in old TMN's terminology or Q3 (see [Q3]) in the newer one. These interfaces are vendor-specific and most of them can be used locally only.
- Q3i – interface between (P)WS and MD. They serve the same goals as Q3 interface in TMN. However, as MCMMS was originally

developed for circuit-switched network (with connections established via low-speed modems), only simple, nonstandard protocols might be used. Even in packet-switched network, because of its 16-kbps basic channels, using of standard protocols (such as CMIP over IP) would be problematic and resource-consuming.

- WSi – interface between two (P)WSs (corresponds to Q3 interface in TMN, but is nonstandard and tailored to specific fuctionality) – conveys reports and databases synchronization data.

4 Manager Applications Architecture

Manager applications are designed for signal officers who are responsible for network command and control. They run on the workstations. Fig. 3 presents the structure of software modules.

The software structure may be divided into three layers:

- **Communication Layer**, which deals with protocols used by management applications
- **Core Services Layer**, which deals with event collecting and analyzing and provides database (Management Information Base, MIB) which is the central point for the whole software.
- **User Interfaces Layer**, which is designed to present management data for human operator (signal officer) as well as to enable management operations.

Two main components of Communication Layer are:

- **Switched Connections Server (SCS)**, which works as server for all management applications and provides simple protocol for transparent and reliable data exchange over modem connections. It also enables effective use of the connections.
- Operating System's TCP/IP protocol stack

Core Services Layer includes:

- **Management Information Base (MIB)**, which comprises relational database used for data storage and **Object Database Server (DBS)**, which maps rows of database tables to objects. Thus from applications' point of view the database is an object one. The DBS is the central application.
- Modules for event collecting and analyzing. **Event Collecting Module (ECM)** is responsible for collecting events from MDs. **Parameter Tracking Module (PTM)** periodically checks those values of NEs' parameters, which are not reported in events. Finally, **Event Analyzer Module (EAM)** analyzes information and updates database content in real time.

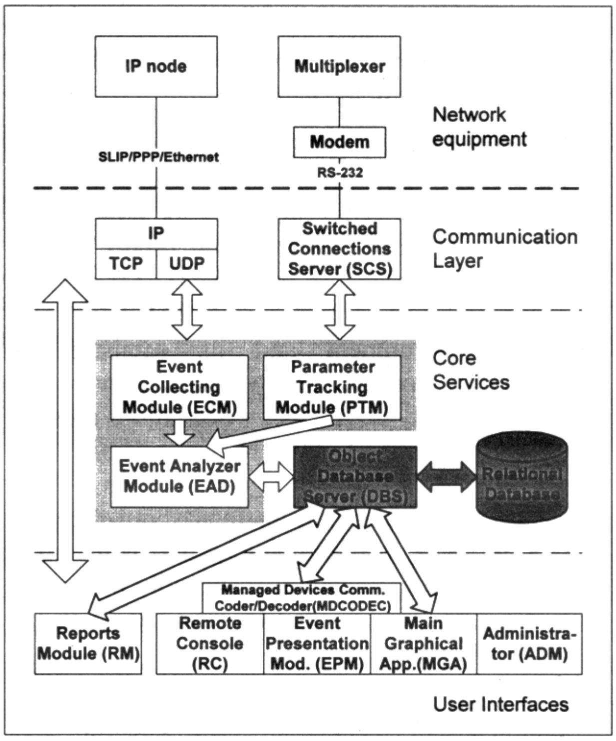


Fig. 3: Manager Applications Structure

User Interfaces Layer is composed of many different management applications. All the applications cooperate with DBS; some of them may use Communication Layer. The most important are:

- **Main Graphical Application (MGA)**, which presents managed objects as icons on the digital map. The colors of icons denote states and alarms in objects and are updated in real time. MGA provides powerful tools for network monitoring. It also supports basic network planning functions.
- **Event Presentation Module (EPM)**, which presents collected events in a table. The events are also updated in real time. EPM provides very strong filtering facilities so the type of events observed in a particular window may be chosen in a very flexible way. EPM enables also on-line examination of subscribers

connections with all detailed information about subscribers, connection times, priorities and additional services.

- **Remote Console (RC)**, which provides means for remote dialogue with MDs and NEs. RC may be used for network configuration.
- **Managed Devices Communication Coder/Decoder (MDCODEC)**, which bidirectionally translates protocols used by NEs to be understandable by human operator forms. The translation is made on the base of templates kept in the database.
- **Reports Module (RM)**, which is responsible for report generation and exchange. The possible report types are: network state report, network or node configuration report, alarms report, events report, subscriber connections report and free-content report. RM uses Lotus Notes environment and its database for document storage. Lotus Nodes servers are located mainly in WS. All workstations must be equipped with Lotus Notes clients. RM provides tools for report generation as well as ensures automatic notifications about received reports. RM is fully integrated with other management applications (especially MGA and EPM) thus effort needed for report exchange is minimal.
- **Administrator (ADM)**, which deals with administering – creating user accounts and assigning them privileges for performing management actions.

5 Management System Functions

The following Operation, Administration, Maintenance and Provisioning (OAM&P) functions are implemented in the MCNMS:

Planning. The MCNMS supports basic planning facilities. Signal officer responsible for network planning can design deployment of nodes, vehicles, devices as well as links between them. Object model and the relationships between objects support configuration checking and correcting (for example, it's not possible to connect two digital modems with an analog link). More advanced functions, comprising performace and fault tolerance assessment of the planned network project (for instance, with the use of network simulator), will be implemented in the future.

Configuration management. This function follows planning process during which all network parameters are established and are provided to support signal officers with the control of the whole network resources. The MCNMS enables on-line database content verification with real network configuration. It also provides means for network monitoring as well as remote configuration.

Fault management. Alarms generated by the NEs are collected and stored by the MDs. The ECM may be automatically notified about critical alarms by MDs or it may poll them periodically for both less important alarms and other events. Then alarms are sent to EAM, which analyzes alarms and updates the MIB. Operator is notified about new alarms by both MGA (real-time update of the icon colors) and EPM (alarms displayed in table rows). In both cases full information about alarm source, generation time and alarm reason explanation are provided. Alarms can be acknowledged by the operator to make them inactive.

Performance management. The MCNMS collects performance-related data such as bit error rates and quality assessments. They may be presented in a table and graphical presentation form. Automated functions for performance correcting are planned to be implemented.

Accounting management. The MCNMS collects information about connections within the network, but it's rather for control purposes. As tactical network is not a commercial one, this function is needless.

Security management. The security of communications between management system components is ensured by a secure network (all trunks are encrypted by bulk encryption units). The access control is supported by Administrator application. All security-related data such as incorrect login to system or events from bulk encryption units are collected and presented to an operator.

6 Interoperability with Other Management Systems

As MCNMS is designed for the tactical level (brigade or division), it will surely have to cooperate with some higher level management system and maybe some other management systems for cooperating tactical units. Internal protocols, as mentioned above, are nonstandard, thus they shouldn't be used outside the MCNMS.

The only solution is to use standard architecture and protocol. But TMN's OSI protocol stack based Q3 interface seems to be obsolete and too huge to use in low-rate IP-based tactical network. We propose solution based on CORBA (Common Object Request Broker Architecture, see [CORBA]) and its IIOP (Internet Inter-ORB Protocol) protocol. The most important reason from interoperability point of view is the presence of CORBA-based solutions in the market of network management and easy integration with TMN (see [JIDM]). It's most likely that cooperating management systems would implement TMN's Q3 or CORBA-based interface for interoperability purposes.

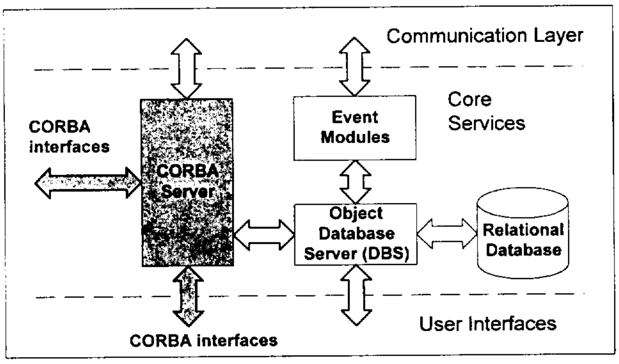


Fig. 4: Interoperability with CORBA

The CORBA Server, operating in Core Services Layer, would map objects and services specific to MCNMS into set of CORBA interfaces. Thus, all MCNMS-specific solutions will be hidden. Details are shown in Fig. 4.

7 Tests

The tests of the MCNMS were carried out during field exercises in November '98. The network was composed of 10 managed vehicles, grouped in 5 nodes connected with radioline links. The exercises included testing of the following basic system fuctions:

- creating project of managed network and its on-line verification with real network configuration,
- tactical network maintenance:
  - monitoring of states and parameters changes in both network elements and links,
  - alarm detection and reporting,
- remote configuration of network elements, including:
  - setting data rates for trunks,
  - frequency setting in radioline trunks,
- report exchange between workstations.

The MCNMS passed the examination and was qualified to be put into practice for Polish Armed Forces.

8 Future Plans

The most important project for the near future is the development of network planning application and integration it with the MCNMS. Capabilities offered actually by the MGA are not sufficient for network planning. First, new application must take into consideration technical characteristics of the managed devices (for example, maximum range of radio-relays), frequency compatibility etc. It should be also integrated with network simulator in order to assess performance and fault tolerance of the network project. Finally, verified network project should be load to the MCNMS's MIB and sent to vehicle crews.

The second project aims at implementation of CORBA-based interface for interoperability with other management systems, as mentioned above.

9 Conclussions

The MCNMS is intended to be an integrated environment for performing all the tasks related to the management of tactical telecommunication network. It's the aim of this project to provide similar functionality as is offered by commercial management systems. The project was based on TMN concepts, but full compliance was not possible because of nonstandard management interfaces implemented in network elements and network resources limitations. Nevertheless, the implementation of CORBA interface will ensure good interoperability capabilities and all the nonstandard solutions will be hidden.

It's also worth to mention that all protocols and interfaces used by the MCNMS are prepared for network development, for example, adding new telecommunication devices, with different management protocol and parameters. Such an extension would require only minimal software update at the MD level and adding new object definitions to the MIB.

10 References

[TMN] ITU-T Recommendations: M.3010 and related

[TMN-2] Salah Aidarous, Thomas Plevyak: *Telecommunications Network Management into the 21st century*, IEEE Press, 1994

[Q3] ITU-T Recommendations: Q.811, Q.812 and related

[CORBA] Object Management Group: *The Common Object Request Broker: Architecture and Specification*, http://www.omg.org/library/c2indx.html

[JIDM] Object Management Group: *JIDM Interaction Translation*, telecom/98-10-10

http://www.omg.org/techprocess/meetings/schedule/Technology_Adoptions.html#CORBA_TMN_Interworking_specification

Present and Near Future Tactical Air Defence Operations Requirements

K.N. Nesterov
Air Defence Department
"G.S. Rakovski" War College
"Evlogy Georgiev" str. 23
1504 Sofia, Bulgaria

N.L. Georgiev
Research and Technology Directorate-MoD
Totleben blvd. 34
1606 Sofia, Bulgaria

The Bulgarian Armed Forces are facing now a great challenge – urgent necessity of restructuring in order to gain higher efficiency and flexibility. This process must go on at the same time with the process of adopting the NATO standards. Well-timed execution of this process is a guarantee for the integration of the Bulgarian army's structures to the corresponding NATO structures and for performing new possibilities in accordance with the global political and military environment. In the current manuscript the focus will be on these problems and their solving in the Air Defence.

The existing Bulgarian Air Defence system is established and developed in full correspondence with the military science of the past. Characteristics of it are all elements of Air Defence system that is deep layered, is capable to cover main strategic objects, directions and zones, and definitely is assigned to operate in military conflict of high intensity or in levels of geopolitical uncertainty and tension before the end of the Cold War.

Main characteristics of the Air Defence are:
1. Strongly decentralized structure;
2. Many levels in the structure for gaining radar information and reporting the results of the surveillance;
3. In order to improve the survivability of the system there are large number of equipment and units.

It is important to notice that far from all of the levels in the chain of links, taking part in the process of gathering radar information, have direct, convenient and comprehensive access to the recognized air picture. Even more – it is impossible to realize better organization structure unless more sophisticated automation is applied.

Inevitably, in the future, should be worked on several directions. According to the priority, the modernization process should start with the reform of the communications and modernization of the radars.

For instance, it is necessary Air Defence units to be familiar with NATO short-range Air Defence (SHORAD) procedures and to be able for planning and executing defence according with Allied Tactical Publication – 42 (ATP-42). For fulfilling a similar task, there will be needed assets and manpower up to battalion with their weapon systems, which are no compatible with adequate NATO weapon systems. That's why it is interesting to find out the answer of the question: "How to gain operational interoperability without replacing all weapon systems (it is impossible to make it at once)?" It is obvious that we have to ensure the operational interoperability of the control system and unified and synonymous recognition of the delivered and received reports and orders. On the other hand it is obligatory to achieved a unified recognition of the air picture in our, commanding, neighboring and interactive command posts. It's clear that information exchange, i.e. the existence of the appropriate communications means, must be realized in accordance with the Allied Administrative Publication (AAP)-31.

Vital condition is the requirement for real-time working of the information system and thus providing the functionality of the command and control system.

That's why the principles, on which base the information system should be established and developed, should be formulated as follows:

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

1. Centralized control must be establish in order to assure normal decentralized execution of the tasks of information resource management;
2. For every one user of the information system should be provided integrated information console (work place) in order to ensure consistent, efficient and effective interaction with the information system and to maximize user productivity;
3. It must be implemented an open system to optimize information system integration and to support further development;
4. An unified data standards must be establish in order to maximize the shared use of the data and to ensure the quality and the integrity of the information flow;
5. Possibilities for further development (using the module principle) to provide timely and cost-effective satisfaction of the information needs and to eliminate the effect of the constant change of the information system components, and thus to gain stability and flexibility in budget funding.

The world changes set additional requirements to the weapon systems, using automated means in carrying out their missions. Naturally, from today's point of view, these weapons (or weapon systems) are at the same way users of one or another information systems.

It could be accepted that Air Defence structure, according the functional principle, consists from interconnected and overlapping surveillance and reporting subsystems, as well as surface-to-air (SAM) coverage, air defence fighters coverage, electronic countermeasures (ECM), logistic and command and control subsystems.

To the Air Defence structure is inherent two variety of links – inner and outer. The main part of the inner links integrates in one all of the system and in the same time surveillance and reporting subsystem and command and control subsystems are united for the entire system. Outer links provide the connection of the system with the surrounded environment, higher headquarter and interactive units.

Let's take a look to some of the Air Defence's system elements.

1. Command and control

Typical for the command and control system is strong and "multifloor" architecture. Such kind of organization chart gives the opportunity of more flexible using, command and control of the forces, i.e. it is related with the functional purpose. That's why it is obligatory the system structure, from organizational point of view, to correspond with the functional purpose. In these terms it is reasonable, when studying the Air Defence, to take in to account the speciality of the organizational chart of the structure, as well as the structure from functional point of view.

Every one of these structures has its own advantages and disadvantages and reflects mostly one of the features of the complicated Air Defence system. In practice these structures are mutual overlapping and supplementing each other.

The greater amount of levels in the Air Defence system is the feature that mostly explain the significant clumsiness when is needed to react in suddenly changed environment. The necessary time for receiving orders and dispatching information is relatively long and in complicated environment could be out of real one. This fact could be more obvious if we track the information flow in the Air Defence system.

The basis of the Air Defence information system is the radar surveillance system.

The existing radar surveillance system includes:

- Radar sensors with their coverage zones;
- First level that forms radar coverage and has the ability to perform independent combat missions for realizing radar surveillance and information distributing;
- Second integration level – Control and Reporting Center (CRC). In this place is realizing the centralized control over the radar surveillance and information distribution process;
- Third integration level - main command and control functions.

It's possible to resume that the system is relatively centralized, but the realized method of gathering, processing and distribution of the information leads to significant delay and causes reduced information quality.

The mentioned above four levels bring inertia and thus forming an environment which not support reasonable decision making process

The existing system allows decentralized control when the centralized one is lost, but the communication status will bring down the effectiveness.

Command and control system is using on an unreasonable way the existing radar resource.

The lack of automation means leads to insufficient flexibility and timeliness of the information flow.

2. Radar Resources

The existing radar resources in the Air Defence are mostly old-fashioned.

The **coverage zone** of the radar sensors characterizes with variety type of means working in different electromagnetic range. Thanks to that fact it is possible to form an optimal, according to the parameters, radar coverage over hilly terrain.

Information quality is significant factor when estimating the performance of the radioelectronic means in Air Defence, and thus estimating the information system at all. Main features for evaluation process are radar accuracy and resolution. The result of the estimating process shows essential need for modernization of the means.

Nowadays, the **information abilities** of the radar means are significantly important. The most of the radar sensors have analog output and low rate of information distribution. These radars, with few exceptions, are incapable for automation target tracking.

Jamming protection is vital for radar performance. Lessons learned from the recent conflicts convincingly show that the use of variety of types, intensity and methods for jamming is obligatory. The abilities of radar sensors from this point of view are far from desirable.

Reliability is another factor on which the radar performance depends. It defines, in greater degree, from the technological level of the radioelectronic equipment. Existing radar sensors are developed on an old technological basis, i.e. its impossible to have high level of reliability.

As for the **survivability,** radar means have relatively good parameters. There is full capability for remote radar control. Some components are dispersed. There are capabilities for cordless information distribution.

In current high dynamic Air Defence operations, the requirement of the **maneuverability** has great importance for the success. Analyzing the current status could be made a conclusion that more of the radar sensors do not have very good abilities, thus it is impossible to gain flexibility in combat conditions. Maneuver with Air Defence radar sensors is possible at the stage of prior combat preparation, with limitation on the stage of direct combat preparation and practically impossible in combat operation.

It must be notice that radar interrogator (identification friend or foe) is at a type incompatible with NATO standards.

The conclusions are as follows:

- Existing sensors could be used in case of applying digital output, distributing the information in common format (for instance Asterix), thus gaining interoperability with adequate NATO and Air Traffic Control systems;
- Modernization process besides acquiring new 3 dimensional (3-D) radars, must be supported by modernization of the existing. This modernization process have to consist of adding extractor units, thus improving the informational capabilities.

3. Communications

When analyzing information system inevitably must be stressed on the condition of the communication system where the radar sensors realize air surveillance.

Good quality and reliability of the communication systems is vital prerequisite for normal functioning of any Air Defence system.

The existing communication system is establish on the base of channels with low information transfer rate and is incapable of distribution the information at great distance. Doubling of the information channels does not lead to better condition. If we notice that communication system is rigid, it is obvious that Air Defence system is incapable of carrying out variety of missions except war missions. The necessity is for flexible, established on module base communication system, capable for optimal performance in rapid changing environment and thus ensures the carrying out of a large scope of missions.

The need is for establishing unified communicational-informational Air Defence infrastructure. This infrastructure must be capable of enlargement in accordance with the amount of modernized and acquired sensors and to remain adaptable to later stages of Air Defence system development. Communication infrastructure has to provide information and service data transfer to the lowest possible level in the structure. At that way will be ensure proper centralized control and decentralized execution.

The information flow must be doubled on the main directions while using different communication means for distribution the information to the users.

Establishing such a communication system will lead to communicational - informational *environment* in which every one user will have access, will be in it and work in real time (with adequate priority and limitations). In this environment will be carrying out all main activities, characteristic for any military system, as well as those activities, specific for Air Defence system.

4. Summary

Air Defence system modernization process can't be selfpurpose and can't be undertaken without serious analyzing study. This is the only way a new system to be able to meet the up-to-date requirements. It means that this system will achieve:

- Needful resource to plan and task the forces;
- Centralized direction of the forces;
- The capabilities to direct and monitor the execution of tasking;
- The ability to communicate with the assigned forces, higher headquarters and the service units.

All these thoughts were made with clear understanding that they serve for near Bulgarian Air Defence future and are connected with the NATO and adequate regional structures integration process.

Multi-Mode Radios – The Way Forward to Flexible Mobile Communications on the Battlefield

Gerhard Hering
DaimlerChrysler Aerospace
Landshuterstraße 26
85716 Unterschleißheim
GERMANY
Phone: +4989-31792742, FAX: +4989-31792568
e-mail: gerhard.hering@mchu.si-t.de

Abstract: This paper introduces an approach to a Multi-Mode-Radio (MMR) for the use in a military tactical environment. The approach has been investigated under a German R&D- programme. From the required functionality a suitable architecture is derived and reasons for the design are presented. Technology needs in the field of Analog Digital Conversion and Digital Signal Processing are considered. Advantages and limitations which appear when operating a MMR in the field will be discussed.

1 Introduction

In January 1995 NATO's Project Group 6 (PG/6) released its final report [5] on Phase II of their investigations of the Tactical Communication System for the Land Combat Zone Post-2000. It quickly became obvious that this report contained two really new revolutionary concepts. The one was the utilisation of the Asynchronous Transfer Mode (ATM), the other was the integration of all tactical radio communications into a Multi-Mode Radio (MMR). Whereas for ATM, experiences can be gained from many civil applications nothing comparable existed for MMRs. Therefore the German MoD launched a study [1] to investigate the MMR concept in some more detail. The study was contracted to a group which teamed up from four German companies SEL/Alcatel, DaimlerChrysler Aerospace, Rohde & Schwarz and IABG. Three major goals were associated with the contract:

1. Verify PG/6's requirements with respect to feasibility.
2. Develop a modular architecture which is suitable for both, migration of existing radios pre-2000 into a MMR and for a later full-scale MMR.
3. Identify critical components and technology needs.

2 Multi-Mode means ...

the combination of at least five different functions together with some expectations regarding the operational use (Figure 1). These functions and expectations are:

- ***Multi-Role Capability***

  Multi-role denotes the capability to support a variety of different teleservices to communicate into different nets. These services shall provide alternative communications media for the user and shall be available on the operator's selection without any HW change. TACOMS Post-2000 explicitly demands for a: Combat Net Radio Service (CNR), a Mobile Subscriber Access

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

Service (SCRA), a Packet Net Radio Service (PRN) and a Relay Service for range extension (REN). Each of the above services itself represents a group of different communications means. For instance, the Combat Net Radio Service may comprise the classical VHF push-to-talk communication, but also a long distance HF-communication and the ground-to-air communication in the UHF band.
Therefore multi-role implicitly also stands for the capability to spread communications over several bands. As a minimum the military HF-, the VHF- and the UHF-band shall be covered by a modern MMR – all together spanning over a frequency range from 1.5 to 600 MHz. This feature is often named multi-band. The multi-role capability allows the MMR to be deployed on different echelons of the military command hierarchy.

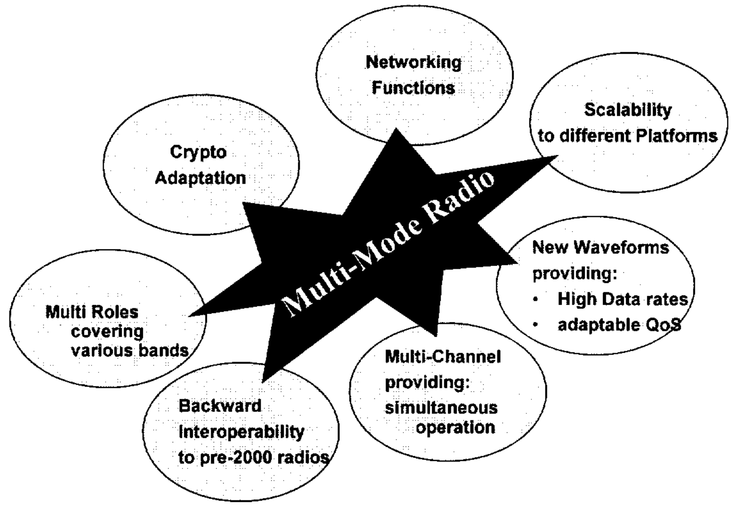


Figure 1: MMR Functionalities

- ***Multi-Channel Capability***

Multi-channel denotes the feature to facilitate several teleservices simultaneously. Highly attractive is the combination of multi-role and multi-channel capability as it opens the principal chance to replace several single-mode-radios by just one MMR. The advantages are obvious: It saves costs, eases logistics and frees space in the usually fully packed communications rack of an armoured command vehicle.
Technically, simultaneous operation is one of the most challenging features of the MMR. Therefore one must carefully trade-off the operational advantages versus the technical implications in the design.

- ***High Data Rates***

Today's tactical radios are really low speed communications devices. This has not been a major problem so far, as tactical communications have been dominated by voice. Advanced voice coding techniques have continuously lowered the source data rate so that even very low speed HF-radios could accommodate digitised encrypted voice.
However with the upcoming data applications, the requests for higher data throughput became demanding. In the future Digitised Battlefield, data applications will require data rates which are far beyond what is presently feasible in tactical radios (and maybe also for future MMRs).

Nevertheless an upgrade by a factor of 10 to 20 compared to legacy radios is anticipated and seem to be realistic.
The channel capacity must be flexibly allocable to different services, so that throughput on demand can be provided. This enables MMRs for new applications, like weapon control and remote sensor links. Parametric waveforms, as for instance the international FM3TR-waveform, will allow the MMR to adaptively adjust the quality-of-service parameters to the varying propagation and jamming conditions instead of falling back in a preventive, low performance mode as present tactical radios do.

- ***Network Functions***

A future MMR will not be considered anymore as just a transport device for information. Internet applications are at the dawn of the battlefield. With the before mentioned high data rates, the prerequisites are given. In addition, to really act as a network node, the MMR must be complemented with networking functions like IP-routers or packet switching. Once the MMR has become a real network node, then instantly the requirement for inter-working functions into other networks (SATCOM, public safety forces networks) will arise. As a consequence, the MMR design must provision for a computing power which is by numbers higher than that of present tactical radios with all the implications in power consumption, reliability and SW maintenance.

- ***Crypto Adaptation***

  Trusted communications will remain indispensable for future MMRs. Until now, all cipher algorithms have to be implemented in HW by authority regulations. Thousands of radios are fielded with such HW ciphers to which a future MMR must be interoperable to. Therefore, the requirement to adopt existing HW ciphers is mandatory for a MMR design. However, proprietary national HW-ciphers cannot be the answer to trusted communications in a future MMR. Civil applications already go for SW ciphers or smart cards. MMRs are believed to follow this approach. This means the national security agencies are anticipated to provide ciphers which can be implemented in SW and parameterised for dedicated missions or multi-national deployments. For the MMR design this means, the radio architecture must be flexible enough to accommodate both, existing HW ciphers for linking up with existing radios and future SW ciphers.

- ***Backward Interoperability***

  Experts estimate the migration period from present single-mode radios to future MMRs may last 10 to 15 years. Therefore interoperability with legacy radios is indispensable for a future MMR. Maybe a NATO-wide interoperability waveform will be defined sometime which allows military forces from different nations to communicate among each other. In a short-term view however, existing tactical waveforms need to be implemented in order to maintain backward interoperability with tactical radios of the generation pre-2000.

- ***Scalability***

  There will not only one type of a MMR. MMRs will need to be scaled for various platforms like a manpack radio or a vehicular radio. Therefore, scalability of HW building blocks and a generic design are essential. Moreover spontaneous configuration is required to set-up mission specific functionality. Here, SW-programmability is the key.

A radio which incorporates all the above mentioned capabilities is becoming the "service partner" for the multi-purpose terminals which, according to TACOMS Post-2000, shall act as a personal communicator to the user on various levels of the military command hierarchy. Such terminals will facilitate advanced multimedia services when connected to a high capacity LAN. Because of propagation constraints the user will need to accept some restrictions when communicating mobile, but services like facsimile, slow motion video and e-mail must be supported by a MMR. As a consequence from scalability and service adaptation, the architecture of a future MMR must be flexible with respect to HW, SW and functionality.

3 The Multi-Mode Architecture

What is a radio architecture? A commonly used definition [3] explains architecture as a mapping of functions to building blocks (antennas, DSPs, terminals) and resources (bandwidth, frequencies, power, etc.). Figure 2 illustrates this definition.

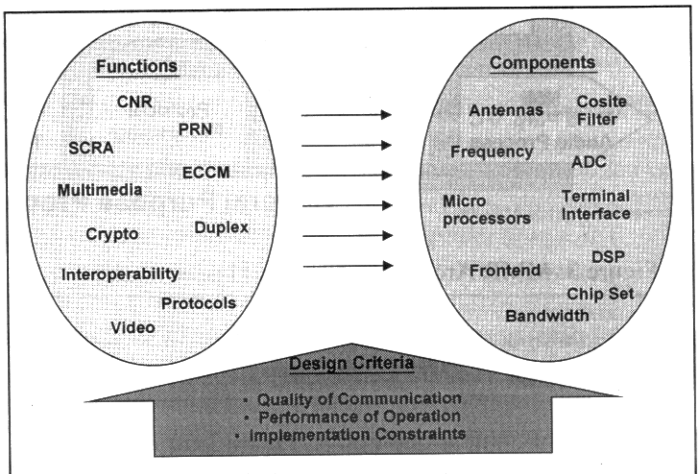


Figure 2: Definition of an Architecture

However, the mapping may not be arbitrary. It is restricted by design criteria which in the end define the characteristics of the device with respect to link quality and operational performance. The architecture we deduced from the above mentioned functionalities by applying this definition is shown in Figure 3. Configurability, multi-band capability and power consumption turned out to be the most dominant design criteria. In a rough view the architecture splits up into four major sections. These are the:

- Frontend Section
- General Purpose Processor Unit (GPU)
- Internal Bus System
- Antenna Interface Module

radio. Instead, we decided to foresee more dualmode frontends instead of an allband/allmode frontend for the following reasons:

The VHF- and the UHF-band is predestined to be combined in one dualmode frontend, but not so the HF-band. HF communication requires very special signal processing, filtering and channel adaptation methods which differ substantially from those in V/UHF. Even if it may technically be feasible, the combination of HF, VHF and UHF to one frontend would tremendously increase the complexity and will result in a suboptimal performance for all communication modes. Thus we restrained from allband/allmode frontends as a design option for a MMR.

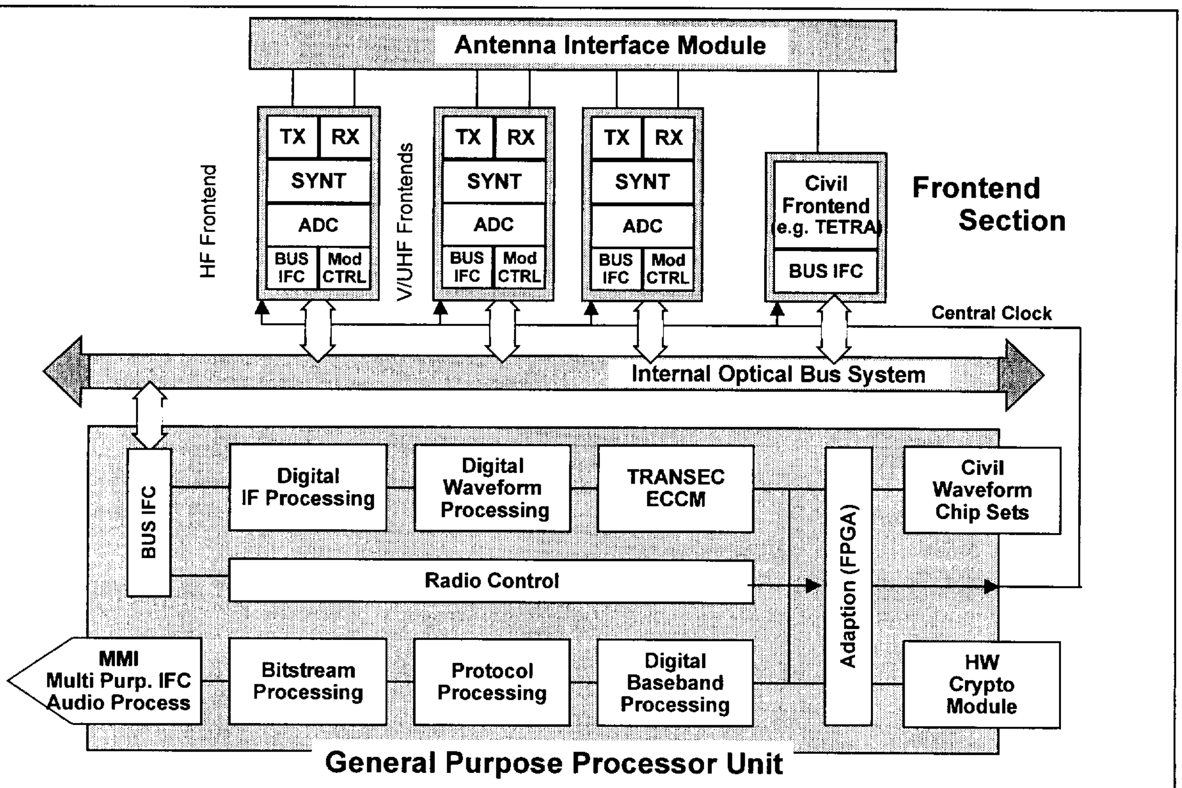


Figure 3: MMR Architecture acc. to [1]

3.1 Frontends

The frontend section as depicted in Figure 3 may look somewhat conservative. We did not follow the approach of a broadband frontend which seems obvious when targeting for a multiband

Contrary to this, VHF-CNR service and UHF-SCRA are rather similar with respect to bandwidth, signal structure and protocols. Therefore it makes sense to combine these two modes in a dualmode V/UHF-frontend. Dependent on the actual platform configuration, a number of these V/UHF-frontends can be incorporated in a task-dedicated MMR. Although from a logistic point of view it also seems quite attractive to have just one broadband frontend module, we believe that different platforms like a manpack radio or a helicopter-borne radio can more optimally equipped by using dual-band or single-band frontends rather than a broadband frontend.

Dual- or single-band frontends match the noise floor and the spurious free dynamic range much better to the AD-converter's input. Filtering, notching and interference cancellation techniques perform more effectively narrowband than broadband. Contrary to this, allmode frontends in general require high linearity amplifiers and low phase noise synthesisers. Often special supply voltages are necessary to achieve optimum performance. All these circumstances increase both the complexity of the power supply and the current consumption which can become a k.o.-criterion for a radio powered from battery.

Nevertheless, the frontends - even if just dualband- shall transmit and receive different waveforms. Therefore all functional blocks (synthesizer, transmitter, receiver etc.) must be prepared to be parameterised by the radio control SW. This turned out to be the key in frontend design. In that sense, the architecture developed in [1] rather aims for a wide range parametrisation than for broadband reception and transmission.

A typical frontend module in this architecture (Figure 3) includes the analogue parts of the transmitter, the receiver (pre-amp, RF-filter, anti-aliasing filter, gain control) and the synthesiser. The analogue signal feeds into an Analog/Digital Converter (ADC) which is the first (Tx), respectively the last element (Rx) in the signal processing chain of each frontend module. All frontends are controlled and monitored via a Module Control Unit which provides a standard interface to the Internal Bus System.

3.2 General Purpose Processor Unit

The General Purpose Processor Unit (GPU) forms the heart of the MMR architecture. It is the core element in which any multirole capability is actually realised. Roles materialise as threads of functions through the HW- and SW-resources of the GPU. Functions can be grouped into four categories:

- channel (IF) processing
- waveform processing
- bitstream processing
- radio control processing

Channel Processing deals with post-processing of the digitised IF-signals coming in from the frontends via the internal bus system. Typical tasks of channel processing are IF-filtering, notching, signal shaping and spectrum characterisation. Even if IF-signal processing is often straight forward and not very complex, it dominates the demand on DSP-power. Much attention must be paid in a proper allocation of IF-processing functions to DSP resources.

Waveform Processing realises the modem function. It typically includes modulation and demodulation, equalisation, carrier tracking and clock regeneration. These functions can be rather sophisticated but are usually less time critical than the channel processing. Real time requirements lie in the range of microseconds to fractions of milliseconds when applying appropriate buffering.

Bitstream Processing is applied to tasks like forward error control, signalling, protocol handling and source coding. These functions can become complex but tolerate delays of milliseconds to fractions of seconds. They fall into a range where microprocessors are superior to DSPs. Very often programmable dedicated HW-components are utilised from the market, which perform much more efficiently than any SW implementation on a processor will do. Field Programmable Gate Arrays (FPGA) can help to match the I/O-ports of these special components to the standard interfaces of the GPU.

Radio Control Processing is usually a low-speed task. In a radio that shall execute multi-role and multi-channel operations, radio control is expected to become a rather complicated issue. Mode agility and how to ensure a dead-lock-free interaction of all the different modes will probably drive the complexity of the Radio Control Software. Radio Control Processing also must keep the Man Machine Interface manageable for an unskilled radio operator despite of the complicated internal interactions.

There are already many components and COTS solutions on the market which can be utilised in MMR designs. For DSP a huge variety of SW for general and special signal operations is available from the market. A common object broker like CORBA will make best use of COTS DSP libraries and will ensure future upgrades. For radio control processing, Realtime Operating Systems are available which support open software architectures. They already provide runtime MMR functions, such as task switching, interrupt handling and access control to shared resources.

3.3 Internal Bus System

A modular architecture is nothing worth without a flexible inter-connecting link between the various building blocks. A communication bus was developed in [1] based on a TDMA oriented access scheme. The TDMA data frame is of 0,5 ms length and caters for 14 communication channels. An organisation channel heads each frame and a request channel closes the frame. Modules on the bus are administered by a Bus Operating System which automatically recognises when a module on the bus go online or offline. Because EMC problems were considered to become very critical in a MMR, the experiment investigated an optical bus system. Up to four PC's with multimedia features were inter-linked via an optical star coupler. The throughput and the flexibility could be successfully demonstrated up to a net bit rate of 112 Mbit/s. Further work needs to be done in order to optimise the block length to the sustainable transfer rate of the bus after the Bus Operating System has reallocated capacity.

In the field of internal bus systems there are a lot of activities in the civil segment. Most promising is the Fire Wire Bus which was developed by Texas Instruments and Apple Computers. It provides up to 400 Mbit/s throughput and has recently been standardised under IEEE 1394. Its future upgrade IEEE 1394.2 will even extend to 8 Gbit/s. Except from being optical the Fire Wire Bus meets almost all the needs of an MMR internal communications system.

3.4 Problem Areas

- ***Adaptation of Pre-2000 Designs for Backward Interoperability***

The present generation of tactical radios, to which a future MMR shall be interoperable, was designed 5 to 10 years ago. In this period ASICs and micro-controller were the preferred components for implementation. Their performance often depend on highly synchronous signal processing triggered by raising and falling edges of clocks. Sometimes even the transit time of logic gates was used to adjust timing. A re-implementation on the GPU for DSP execution will be rather inefficient and risky. We recommend to incorporate these components as they are. FPGAs are a suitable means to adapt the time critical, synchronous interfaces of these components to the more asynchronous design of the GPU.

- ***Red/Black Separation of Crypto***

Crypto devices for NATO classified messages impose the difficulty of separating encrypted data from non-encrypted data. This is often called the red/black separation problem. In a highly SW-driven MMR where the signal processing chain is not along a line of HW components but rather a thread through SW, the red/black separation may cause a severe problem.

- ***Chip Sets for Civil Waveforms***

For joint operations with public safety forces tactical MMRs are expected to support civil waveforms. Such a waveform could be TETRA 25 that may become the common European interoperability waveform for safety forces. For all civil communication devices, which are produced in high quantities, dedicated chip sets (Gold Chip Set GSM) exist on the market. These chip sets are extremely optimised and power-efficiently designed. It is does not make sense to re-engineer them for a SW-implementation on a General Purpose DSP. The MMR architecture must allow to adopt these chip sets and to interface them to the GPU and the frontends.

- ***Antennas***

A multiband radio consequently requires a multiband antenna. The attempt to cover the full frequency band from 1.5 MHz to 600 MHz with just one antenna touches physical limits and will not result in a practical antenna design. Much more promising are smart antenna arrays from which the *Antenna Module* of the MMR selects the appropriate antenna for a distinct service. Nevertheless this is just one option. At a manpack radio, where an antenna array cannot be used, the antenna problem remains unsolved [6] especially when HF is involved.

4 Technology Needs

The investigation we performed during the study made clear that there are two elements that really determine the feasibility of a software radio. These two are: the Analog/Digital Converters (ADC) and the Digital Signal Processors (DSP).

- ***ADC-Technology***

The product of sampling frequency (fs) and the full scale dynamic range is often used as a figure-of-merit which characterises the ADC performance. For contemporary technology, this figure-of-merit is about -135 to -146 dBc/Hz for monolithic ADCs and -150 to -160 dBc/Hz for hybrid ADCs.
Table 1 shows this value for some of the roles of a tactical MMR. One can also see from Table 1 that digital RF is still beyond the scope of present technology.

	fs (MHz)	full scale dyn. range (dB)	figure of merit (dBc/Hz)	Trend Index L
RF				
– Mobil - Mobil (CNR, VHF)	270	80 ... 90	165 ...175	43
– Mobil to RAP (SCRA, UHF)	1000	70 ... 80	160 ... 175	42
– Mobil to Mobil (CNR, HF)	75	100 ... 120	180 ... 200	46
IF				
– HF IF	0.2 ... 10	80 ... 100	130 ... 170	34 ... 36
– V/UHF IF	2 ... 40	60 ... 80	120 ... 156	34 ... 36

Table 1: Characteristic Demands on ADC

There is a technology trend which follows a logarithmic rule [4] of the form $L = n + \log_2 (fs)$. Herein (n) is the resolution in bits and (fs) the sampling frequency. The bit resolution (n) can be expressed as a function of the full scale dynamic range (DNR) by: $DNR(dB) = 6.02 * n$. The trend index L was $L = 28$ in 1978 and $L=37$ in 1993.

Following a linear tendency, this trend rule prognoses $L = 39$ for 1997. This value coincides quite well with available technology. Assuming the trend will continue as it has done over the last 15 years, digital RF may become feasible in 6 to 8 years. For the time being the dynamic range needs to be reduced to about 60 dB by an Automatic Gain Control (AGC) and the RF-frequency must be down-converted to an IF of some 10 MHz.

- ***DSP Capacity***

A precise estimate of the required DSP capacity would actually require a complete functional SW breakdown and an allocation of tasks to processors. Our study did not go so far. Instead we applied a simplified model of a SW radio [3], which allows a rough order-of-magnitude estimate of the necessary DSP power. From this model a formula can be derived with which one is able to estimate the minimum required processor power in millions of floating point operations pre second (MFLOPS)

$MFLOPs > [100 * (fs + Bc + Rb) * 4] * 10^{-6}$
whereas: fs ...sampling frequency,
Bc... channel bandwidth, Rb...bit rate

For a MMR executing a VHF CNR-role, a UHF SCRA-role and a HF-role in parallel, about 180 MFLOPs of processing power would be required. Following a good design practice which says, do not load the processor more than 60% in average, about 300 MFLOPs of processing power should be installed. This is not beyond today's capacity which a multiprocessor array can provide. The most critical problem however will become the power consumption. Even if modern 3.3V technology will relax the problem, a full scale implementation on General Purpose DSPs is not recommendable. Instead, a good mix of dedicated HW components, FPGAs, RISC Processors and General Purpose DSPs in combination with microprocessors seems more promising to cope with the power limitations of a battery supplied MMR. Such a mix can be easier realized if the architecture is dualband- and dualmode-oriented rather than allband/allmode.

5 Operational Aspects

SW-radios with multiband capability are undoubtedly the right approach for flexible mobile communications on the battlefield. However MMRs impose also new problem areas which have not yet been thoroughly investigated in [1]. Two of them are addressed below.

- *Radio Management*

In order to manage the different roles in an MMR, the radio needs to be pre-loaded with many operating data like frequency sets, keys, Time-of-Day (TOD) and identifiers. This can either be done before the radio goes operational or these data must be downloaded from a central station via the radio link. The Radio Access Point (RAP) of the mobile subsystem may serve as such a central station. In any case, configuring a MMR will turn out to be a rather complicated job. A soldier in a combat situation is certainly unable to do this job. This means an enhanced facility control as part of the network management becomes vital for a tactical communications network with MMRs. The problem is that such a sophisticated facility management is unlikely to be maintained in the chaotic circumstances of a real battle situation. Therefore emergency modes must be defined into which a MMR falls back in case it is disconnected from the network management facilities for some time.

- *Configuration to Various Platforms*

MMRs will service in various applications on different platforms. But as different the platform, as varying the radio configuration usually needs to be. For instance, a manpack-type MMR does not need the same multimode and multiband capability as a military commander's vehicular based MMR. Therefore MMRs must be HW-configurable. Ideally this could be realised by changing modules. However mechanical, environmental and platform specific aspects will not allow to follow this approach strictly. Standardisation on an architectural level can help to define interfaces which ease the incorporation of building blocks for different applications but it does not fully solve the problem. The Software-Design-Radio (SDR) Forum (former MMITS) is such a forum which aims for architectural standards. Contrary to this, standardisation at the level of functions and implementations may be even contra-productive because civil technology progresses much faster than common military standards can evolve.

6 Experimental Steps

6.1 The FM3TR Waveform Processing System

Two companies of the study team (SEL and Rohde & Schwarz) continued the theoretical part of the study [1] by an experiment. The aim of this experiment was to show how waveforms can be represented by a set of parameters suitable to pre-set an hardware and software accordingly. They started with a systematic registration of parameters which specify a waveform and found out that for complex waveforms up to 400 parameters could be necessary to comprehensively describe the waveform.
Based on these parameters a Waveform Processing System (WPS) was programmed to execute the SEM 93 VHF CNR and the SATURN UHF avionics waveform. The WPS consisted of a Digital Waveform Modulator, a Wideband Receiver, an Antenna Interface and a Waveform Pre-Processor. Digital bit processing was executed in a Data Processing Module realised with building blocks from existing radios. The experiment could successfully demonstrate interoperable communication to real SEM 93 and SATURN radios by setting the parameters in the WPS accordingly.

6.2 The Advanced MRR Demonstrator

The next step towards a MMR will be an R&D programme (MMR-ADM) in which a fully functional demonstrator shall be developed. This demonstrator (see Figure 4) will be capable of executing nine different waveforms both for interoperability to pre-2000 systems and future high data rate multimedia services. It will cover the frequency band from 2 to 600 MHz and will allow simultaneous transmission of two modes. The demonstrator is also to prove the adaptation of existing HW-ciphers into a general purpose processor platform and how to adopt COTS chip sets for civil waveforms. The implementation of the international parametric test waveform FM3TR is foreseen in both variants, the time division duplex (TTD) and the frequency division duplex access scheme.

The R&D programme is planned to run in three steps:

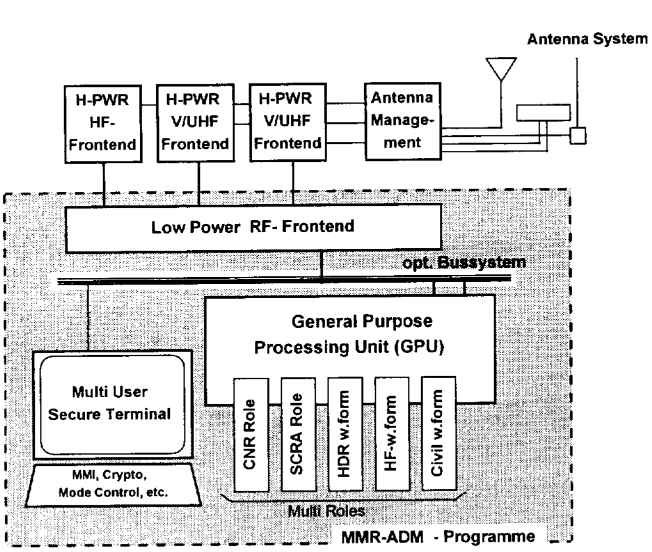


Figure 4 Advanced MMR Demonstrator

1. Demo system for interoperability to pre 2000 tactical radios
2. Demo system with HDR waveform
3. Demo system with multiband capability (HF, VHF, UHF + civil waveform)

The present time schedule presumes a start in 1999 and a duration of 4 years.

7 Achieving PG/6 Requirements

In the Mobile Subsystem of the Tactical Communication System in the Land Combat Zone Post-2000 (see Figure 5), the MRR plays a key role. Herein it shall operate as a CNR both in the VHF-band and the HF-band. It shall serve as Mobile Subscriber Terminal (MST) providing point-to-point dialled-up teleservices both net-internally and externally into the Wide Area Network. For the Packet Radio Network (PRN) the MRR must provide a connectionless data protocol and any MRR in the loop shall act as a potential repeater to forward messages to their final destination. Although PG/6 does not explicitly demand simultaneous operation of different roles, this feature becomes evident when combining PRN services with other roles. This is because a MMR must be able to receive and transmit data packages at any time, that means also when it is presently executing another role. PG/6 also requires a function called Range Extension Node (REN). The REN was primarily foreseen to link-up remote CNR subscribers which are out of range to their comrades in the radio loop group. With a MMR which is capable of processing several CNR waveforms in parallel, the REN function opens a further option. It will allow for inter-linking CNR nets even if they do not use a commonly agreed interoperable signal structure.

The architecture which has been developed in the study [1] is believed to be suitable for fulfilling the above mentioned TACOMS Post-2000 requirements. The architecture covers all the tactical frequency bands and can accommodate network functions. It allows for simultaneous transmission and the incorporation of existing crypto algorithms and civil chip sets. The design is open for both HW and

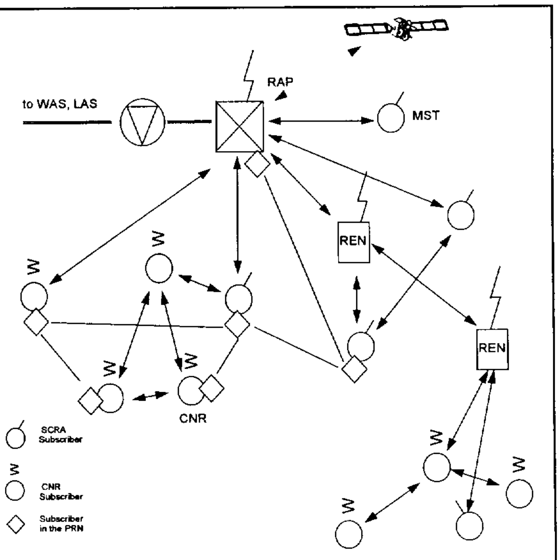


Figure 5: PG/6's Mobile Subsystem

SW-configurability. Therewith in principle all preconditions are given to build up a MMR. The ADM programme will prove the study results by implementing the architecture in HW and SW.

8 Conclusion

A radio which incorporates all the above mentioned capabilities is far more then just an enhanced combat net radio. It is rather becoming the "service partner" to the multi-purpose terminals which, according to TACOMS Post-2000, shall act as a personal communicator to the user on various levels of the military command hierarchy. It is believed that the overall life-cycle cost of the equipment may drop by 40% with MMRs compared to conventional single mode radios [2]. Interoperability, task specific configuration of communications means and seamless communications among all subsystems of the Tactical Net Post-2000 can become reality with MMRs.

In principle this is nothing knew. The Universal Mobile Telephone System (UMTS) which is planned to go in service by the year 2002 will already be a kind of MMR. Military tactical MMRs will require big efforts in defence industry, but technology-wise they are not a hazardous enterprise. The joint team which has worked on the study is convinced that MMRs are the way forward to flexible mobile communications on the digitised battlefield.

References

[1] Study „Integrierte Kommunikation auf dem Gefechtsfeld", German Bundesamt für Wehrtechnik und Beschaffung, Contract E/F21A/P0261/P5249, 1995

[2] Rohde&Schwarz, „Multi Role Radio (MRR). Ein Generationensprung in der Funkgerätetechnologie" Deutsche Heeresrüstung, Nov. 1997, page 42 - 43

[3] J. Mitola, III, „Software Radios" Symposium held in Stockholm May 14-16, 1997 by H. Silver & Associates (UK) Ltd.

[4] Study „Kritische Komponenten für künftige Truppenfunkgeräte", page 228 ff., German MoD, Contract T/R450/M0034/L2200, Feb. 1994

[5] PG/6 Phase II Report,

[6] Microwave & RF, „Future Radios Plans are Breaking up" , June 1997, page 23a

Modeling and simulation of the ZODIAC tactical communication system

Dr. Bert Boltjes and Ing. Roger J.G.M. Langeveld
TNO Physics and Electronics Laboratory
P.O. Box 96864
2509 JG The Hague
The Netherlands
Phone +31 70 3740746, fax +31 70 3280961
E-mail boltjes@fel.tno.nl
http://www.tno.nl

Abstract
This paper gives a description of the development of a simulation model of the ZOne DIgital Automatic Cryptographic (ZODIAC) tactical communication system of the Royal Netherlands Army (RNLA). Also, examples of simulation results and some lessons learned will be presented. The ZODIAC tactical communication system of the Dutch Army consists of access equipment and mobile switches. The switches are typically interconnected via line of sight radio links. A deployment of a ZODIAC network has, like all communication systems, a finite capacity to handle traffic. The ZODSIM simulation model was developed to study properties of deployments and the impact of new equipment and services on the capacity and to predict bottle necks. The introduction presents a brief description of ZODIAC, features implemented in ZODSIM, the OPNET simulation environment used to develop the model, and some modeling techniques. In the section thereafter, some examples of simulations of deployments of ZODIAC and obtained results will be presented, followed by a discussion and some lessons learned.

1. Introduction

1.1 Objective

The ZODSIM simulation model was developed at TNO to study properties of deployments and the impact of new equipment on their capacity and to predict bottle necks in deployments of the ZODIAC tactical communication system which might occur in future missions. The ZODIAC tactical communication system was designed to provide voice and data communication services during defense missions of the RNLA. Nowadays, novel equipment and data services which use ZODIAC are being tested and deployed. The RNLA asked TNO to develop the ZODIAC simulation model, called ZODSIM, in order to be able to study the behavior and capacity of ZODIAC deployments under various conditions and with new equipment. Beside testing novel ZODIAC deployments with standard voice and data traffic, ZODSIM can be used to study the impact of future applications and services. With the ZODSIM simulation model, the impact of changes can be studied and alterations modified before being used in the real world.

The commercial OPNET simulation environment was used to develop the model. This object oriented environment is especially designed to develop and perform simulations of telecommunication networks. This open environment has a Graphical User Interface (GUI) and many specialized simulation kernels which can be used to create models of communication systems. For instance, the kernels can be used to create models of systems that route or queue messages, model the behavior of users, or model communication protocols. Although models of a large number of COTS equipment is included in this environment, TNO had to develop models of the specialized ZODIAC equipment. The next section gives a brief description of the ZODIAC system, and the relevant features and functionalities which are implemented in ZODSIM simulation model. Also a brief description of the OPNET simulation environment, and modeling techniques

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

will be presented. In the subsequent sections, some examples of simulations of deployments of ZODIAC and obtained results will be presented and discussed, followed by conclusions and some lessons learned.

1.2 The ZODIAC system and features implemented in ZODSIM

The ZODIAC system consists of access equipment such as digital encrypted telephones, mobile switches and Line Of Sight (LOS) radio equipment. Typically, voice and data users of ZODIAC are connected to a Multiplexer Access Point (MAP). A MAP can handle up to 31 local users. These users use either an encrypting telephone for voice communication or a data terminal to exchange information. In the MAP, the user data is multiplexed and passed on to their local switch. A MAP is connected to a switch via an S4 cable. The switches, where information is passed onto and from local users or routed to other switches, are usually interconnected via LOS radio links (trunks). A ZODIAC switch performs tasks such as, routing of processornet messages and call information, and for instance updating of routing and subscriber location tables. The interconnecting trunks typically have 15 or 31 channels for switched connections, with 16 or 32 kb/s digital channel transmission speed respectively. When operating at 16 kb/s per channel, a trunk can have 64 channels of which 31 can be dynamically allocated for switched connections and 31 channels for dedicated sole user channels. One of the remaining channels is used for framing, the other one is a common channel to transport system messages called processornet messages.

As for all simulation models one has to know which properties of the real system, and to what level of detail, must be incorporated in the model to enable it to give useful answers. Only those properties of ZODIAC were incorporated in ZODSIM to obtain useful simulation results and to avoid extensive and unnecessary modeling effort and computations.
Basic properties of ZODIAC such as;

- Eurocom transmissionmechanisms (with a go-back-n error correction mechanism),
- switches and their relevant properties (routing, maintaining tables et cetera),
- processornetsignaling and the effect of a non-zero Bit Error Rate (BER) on it,
- channel occupation,
- properties of ZODIAC equipment (maximum capacities, data rates, et cetera),
- re-routing of connections after link failures,

and usergroup traffic modeling were included in the simulation model. The modeled network must try to set-up, maintain, and disconnect requested connections. Traffic statistics from a number of basic real ZODIAC deployments were used to validate the simulation model. Some of the implemented features shall now be described in more detail.

A first example of a relatively straightforward approximation is to use statistical models to generate traffic. It is not required for the traffic scenario to know the call behavior of individual users. Therefore, voice users connected to the same MAP are treated as a usergroup whose call behavior (frequency and duration) is simulated using a tailored statistical model so it can function as a representation of the behavior of the call behavior of the group in reality. For useful simulation results, each tailored traffic generation model of a usergroup has to know how many users it has, how often, for how long, and with which statistical distribution, they will call users in other usergroups. The model for the usergroups keeps track of statistics such as how many calls were initiated, ended successful, and how long they lasted. Traffic generation and MAP functionalities are in ZODSIM modeled in one type of node model called usergroup.

The models of switches only mimic the behavior that is needed to study those properties of real deployments that are of interest. For instance the switch model keeps track on how many channels are occupied in a trunk and it tries to route a call over an alternative trunk when the channels in the preferred one are all occupied. A switch model does not perform the routing of the actual frames of the calls. The model ignores details below the call level. However, switch models do respond to the processornet messages they receive. The MAP's and switches use processornetsignaling to set up, maintain, and terminate calls. For instance, when a user hangs up, the local switch is notified with an appropriate processornet message which causes the switch to undertake the necessary actions to release the network resources that where allocated for the call. The switch models keep track of the number of occupied channels as a function of time, how many calls were routed through them, the load of processornet channels of trunks, and other statistics of interest.

In the ideal case, only abstractions of processornet messages with the necessary information and length would suffice to obtain the required simulation results about the performance of the processornet. The abstractions of processornet messages are called packets in OPNET. In these packets there are data fields defined to specify the type of the message, how many bits there are in the message in reality, and other information such as the ID of the sender, time stamp, et cetera. When an OPNET model of a transmitter sends such a packet over a model of communication link with a specified data rate, it calculates from the data rate of the link, the distance between the sender and receiver, and the speed of light how long it takes for the last bit of the packet to arrive in the receiver. This way, packets don't have to be send bit by bit and simulations run more efficient. However, errors that occur during transmission over the radio links and the transmission protocol, call for somewhat more detailed models of message transmission. More detailed information than present on the level of processornet messages is required to answer questions regarding the properties and behavior of the processornet under imperfect conditions. Transmission protocols try to correct errors in received messages and cause an increase of the load of the processornet or even the breakdown of a link. A model of the transmission of processornet messages was created to mimic the effect of transmission errors.

In ZODIAC, processornet messages are transmitted on a block-by-block basis of 32 bits each, of which 16 are information bits. The rest is for numbering, an OK/RQ bit, FEC bits, and one parity bit. A non-zero BER causes error bits in transmitted blocks. The receiving side will request retransmission of a block when it can not correct for the errors in it. ZODIAC uses a go-back-n correction mechanism with a 4 block retransmission cycle to recover from faulty blocks. The processornet message, and ones that might be waiting behind it, thus get an extra delay before it arrives correctly. The processornet has to carry more data due to the retransmission of blocks. In the worst case the BER is too high to recover from corrupted blocks. The switches then start a re-synchronization sequence to recover the link. The effect due to a non-zero BER and the go-back-n mechanism was modeled in a separate queuing process to which a switch can pass processornet messages. Using the BER on the link and a random number generator, the process mimics the retransmission of blocks of messages and computes the additional delay messages get and the extra load this gives on the load of the processornet channel in the link. The queuing process keeps track of the processornet statistics, such as the number of bits waiting in the queue of the transmitter, the additional delay, load, and other statistics which are of interest.

A model consists of a number of levels. The highest level is the network level. In this level the topology and equipment is defined. An example of a ZODIAC deployment in ZODSIM at the network level is shown in Figure 1.

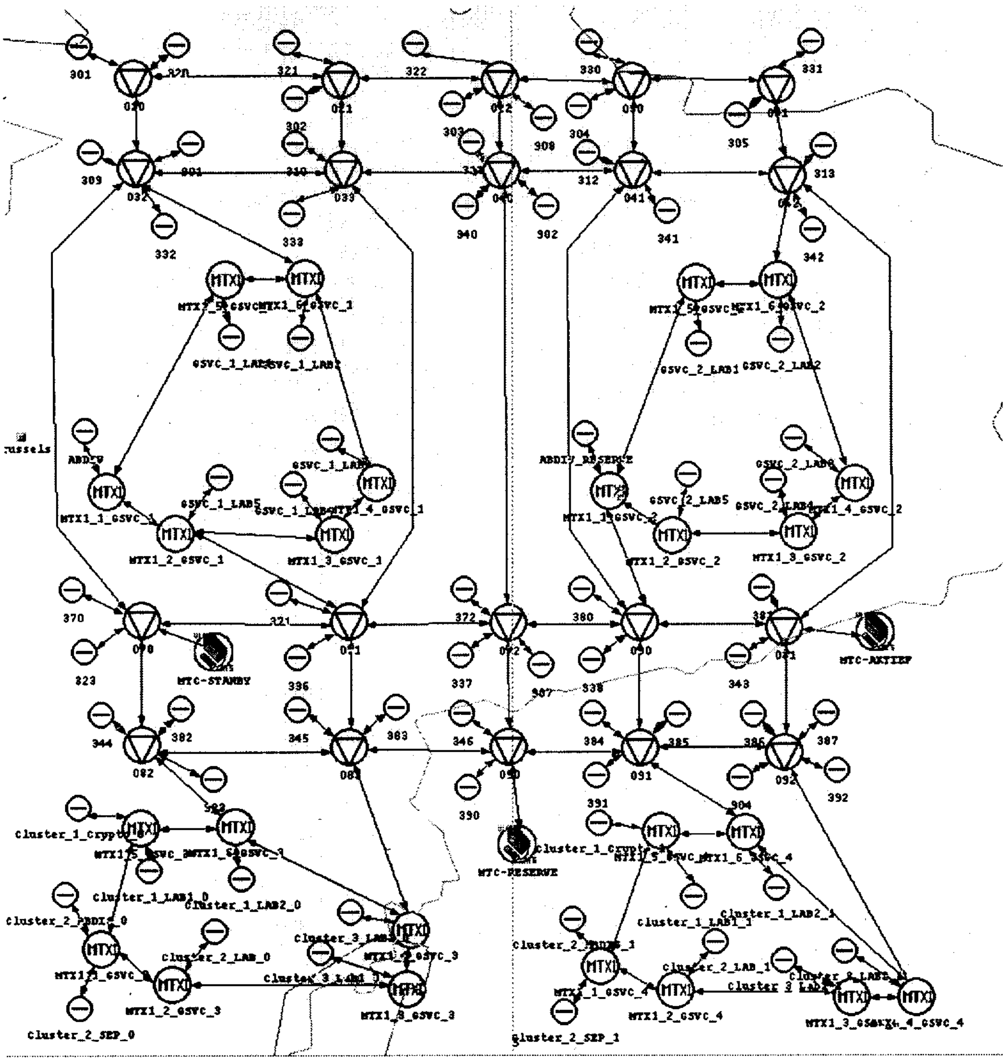


Figure 1 Example of a large ZODIAC deployment at the network level.

Figure 1 shows the various types of nodes that can be present in a ZODIAC deployment. The nodes with a circle-bar icon are usergroups from which calls are generated. The usergroups are connected to their local switch via a fixed duplex connection. The nodes with a circle-triangle icon are switches which can connect to up to 16 links. The nodes with a circle-MTX icon are novel small switches which can connect with only a small number of links. Typically, the switches are connected with a duplex LOS radio link. The radio links are represented by the same double arrow symbol as the fixed links. However, behind them are different models. For instance, for the fixed links it not possible to select a non-zero BER. In the radio link models allows any value for the BER. Note that the kinks in the LOS radio link were introduced in Figure 1 purely for graphical reasons. The nodes represented by a circle-terminal icon, are servers which can offer simple message handling services to data-users in the usergroups. As mentioned above, nowadays novel data services are being tested for use on ZODIAC deployments. Simulation results which offer insight in the requirements and impact of the novel services will be discussed in the next chapter.

Models of deployments such shown in Figure 1, are made quite easily in the OPNET GUI by selecting the required node type from a pallet, dragging it to the network area and connecting it with the required links. Names, data rates, BER on links, and other parameters are object attributes with a default value that can be adjusted when required. Usergroups in the deployment obtain their parameters for their statistical traffic profile from a ASCII scenario file. This file has to be edited to fit the deployment; the traffic matrix in it contains a set of parameters for each group that generates traffic.

Studies of properties of deployments of the ZODIAC network can give an insight in the characteristics of ZODIAC under conditions that are expensive, not yet possible, or difficult to achieve in reality. For instance, creating the deployment shown in Figure 1 in reality requires a large workforce. In the order of 20 to 50 men is needed per node to operate the LOS radios, switches, Map's, access and other equipment. If not impossible in peace time, testing such large deployments and various alternatives would be an expensive operation. Simulation in this case offers an elegant way to get an estimate of properties of the real system.

As stated above, traffic generation and MAP functionalities are in the usergroup node. Traffic is generated by voice- and data-users which pass their requests to an object inside the node called the request_administrator. This object performs the necessary MAP functionality's. The internals of the usergroup node are displayed in Figure 2.

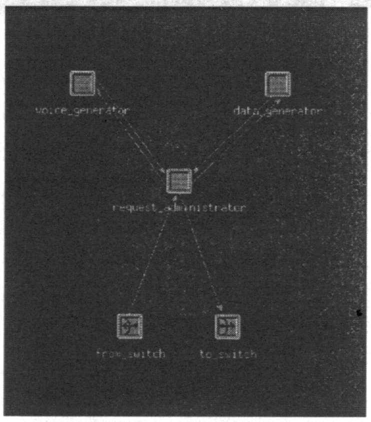


Figure 2 Graphical representation of the objects in a usergroup node. On top are the traffic generating objects in the node which performs the MAP functionality's. At the bottom-right is the object for transmitting requests to a switch, on the left is the receiver.

Figure 2 is a graphical representation of the functionalities of a usergroup node at the level below the network level in Figure 1. This lower level is easily accessed in the GUI by dubbel clicking on a usergroup node. In the usergroup there are two traffic generators, one modeled to represent voice users and another one to represent users which are transmitting datagrams. Frequency and duration of network resource allocation is significantly different for these two groups, so each generator has it own appropriate statistical traffic profile. The request_adminstrator object performs MAP functionalities and processes processornet messages. These messages can either be received and passed on by the receiver object, called from_switch in Figure 2, or send to the local switch via the transmitter, called to_switch.

The transmitter and receiver objects are connected to a model of a duplex link between the usergroup node and the local switch. The messages which arrive at the receiver objects in the switch are passed on to the dispatcher object which processes incoming messages and generates the required responses. The internals of the switch node are shown in Figure 3.

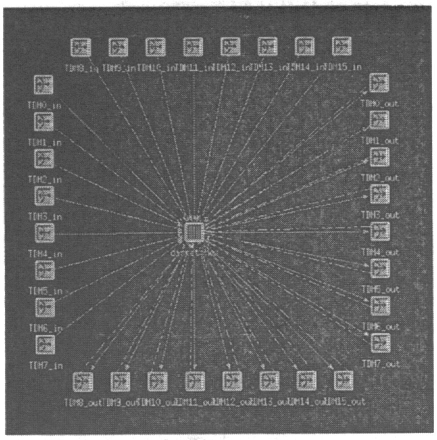


Figure 3 Graphical representation of the objects in the switch node. The object in the center is the dispatcher which performs the functionalities of the switch. The objects on the left and on top are receivers which can connect to 16 links. The objects on the right and at the bottom are the transmitter.

Figure 3 shows that a maximum of 16 links can connect to this switch model. However, models of switches can easily be modified by adding or removing transmitter and receiver objects and connecting them with an appropriate link. Models of the novel small MTX switches where thus simple obtained by removing receiver-transmitter object pairs. As Figure 3 shows, the transmitter objects are connected with double links to the dispatcher. One link represents a stream for outgoing processornet messages, the other one is a so called statistical wire through which the transmitter is able to inform the dispatcher that it has started or stopped transmitting. The main process residing under the dispatcher object can then send the next waiting message, perform updates of the statistics, et cetera. A graphical representation of the dispatcher process is shown in Figure 4.

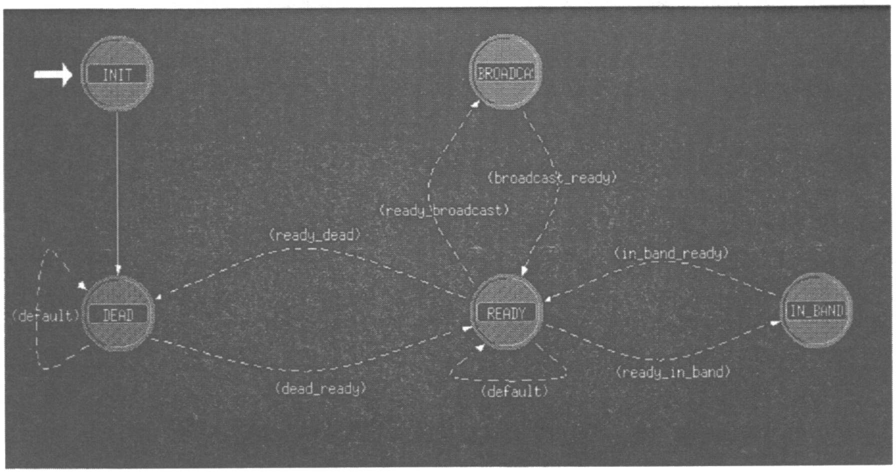


Figure 4 Graphical representation of the process residing in the dispatcher. The process is modeled as a finite state machine.

The process is modeled according to a finite state machine: the process can be in a finite number of states between which can switch over to in response to arriving interrupts or other events. The states are called INIT, DEAD, READY, et cetera. When the simulation starts the process starts in the INIT state where it performs initializations such as finding out which transmitter and receiver pair are connected, what is on the other side, who are its local subscribers et cetera. It then goes over to the DEAD state to wait until it receives an interrupt to become active. The process can then transit from state DEAD to READY when the Boolean condition dead_ready is true and wait there for arriving processornet messages and interrupts. This main dispatcher process can be the parent of many child processes. Instances of child processes are created each time a call has be set up. Each child keeps track of the state of a specific call and gets processornet messages passed on from the parent for the call that they are responsible for. Also, a child process monitors time out counters for the call. Kernels in OPNET can be utilized to create, communicate with, and destroy child processes.

The GUI facilitates the design of processes as finite state machines by guiding the programmer. State machines can easily be created by adding and connecting states and giving the transition the name of the Boolean which must be true to allow the transition to take place. OPNET collects all the C code in the states and creates a program with the functionality as laid out in the GUI. The programmer is responsible that the C code in the states performs as required (get or create packets, responses to arriving packets and interrupts, update statistics, broadcast messages to connected links et cetera) so the model mimics the functionalities and properties of the real system.

When the simulation is started, OPNET creates for each object in the deployment an instance of the appropriate main process, adjust their settings to the required values (names, data rates, BER's et cetera), connects all pieces of code according to the network topology and creates the simulation program. This way the programmer himself does not have to dig through piles of C code when program modifications are needed: he or she can navigate graphically through it with the aid of the GUI.

The next section of this paper presents and discusses results of simulations ZODIAC deployments.

2. Simulations and results

2.1 Large ZODIAC deployment with novel switches

ZODIAC equipment is mobile in the sense that LOS radio, switches, and other equipment are mounted on trucks. A deployment is not mobile. Once a communication center is created by connecting access equipment, MAP, radios, et cetera, moving it is not so easy. To be more mobile in the future, new and lighter equipment was developed. During the development of this equipment, questions raised on what their performance would be and whether they were properly dimensioned.

To study the traffic load on novel small more mobile switches in a large deployment, a large deployment was created and clusters of small MTX switches were inserted. The deployment is shown above in Figure 1. The small switches are symbolized by the circle-MTX icons. The 4 clusters with each 6 MTX-switches represent headquarter sites with each 3 pairs of stacked switches. An estimate was made for the traffic matrix for usergroups in the deployment. This matrix was based on the functionality of the usergroups and their relation to each other. For instance, the number of calls between a usergroup in the frontline, at the top of Figure 1, and a usergroup much further away from the front, can expected to be much smaller than the number of calls between the usergroups in the frontline. Estimates for the frequency and duration of calls under normal use was obtained from measurement in small real deployments. To study the traffic load on the MTX switch clusters due to outside traffic, the usergroups in the headquarters did not generate any traffic. This way an estimate for the capacity remaining for these usergroups was obtained.

To study the performance of the deployment, 3 simulations were performed with increasing network load. In the first simulation, the average time between call requests of users in a usergroup was 523 seconds. In the second simulation this was 105 seconds, and in the third one 19 seconds. These intervals correspond to light, medium and heavy traffic load. The average call duration was set to 25 seconds. The results are in Table 1.

Average time between two call requests [sec]	Calls ending successfully [%]	Calls for which called subscriber was occupied [%]
523	90	10
105	78	22
19	66	34

Table 1 Average time between calls and the number of successful and unsuccessful calls.

The network did not yet show signs of congestion during these simulations. There were enough channels available to handle the calls. When a call was not successful it was because the called subscriber was occupied.

Another result obtained from these simulations was that there was no clear correlation between the position of a LOS radio link (trunk) in the network and its average traffic load. For instance, the trunk between switch 041 and switch 080 had a smaller average load than the trunk between switch 033 and 071, although they both are parallel with the mirror axis of the deployment (the dotted line in Figure 1). The histogram in Figure 5 shows the average number of occupied channels in some of the trunks connecting switches in the deployment during the simulation with the highest traffic load. All the trunks had 31 channels available to connect the calls.

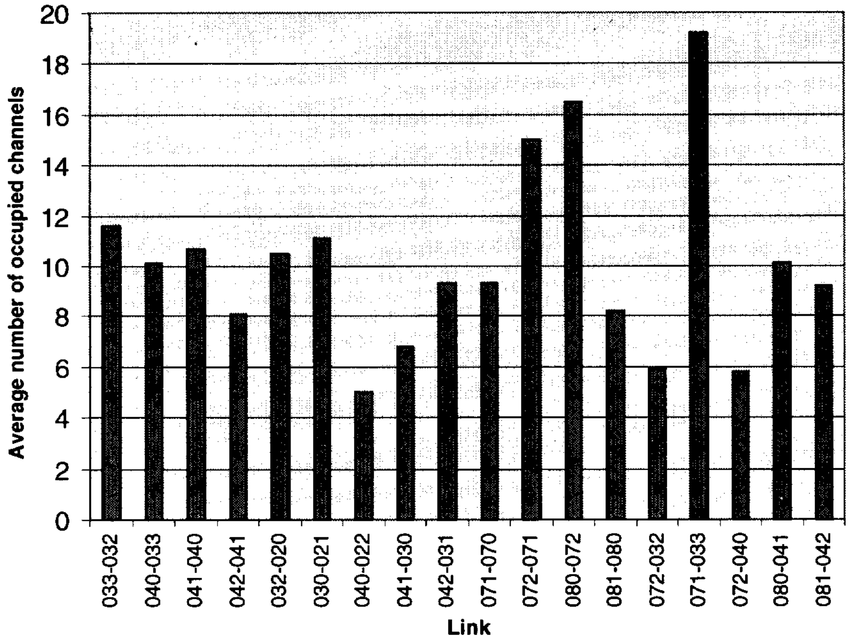


Figure 5 Average number of channels in the trunks occupied during the simulation with the highest offered traffic load. On the horizontal axis are the ID's of pairs of switches which the trunks connect.

These results indicate that there are no bottlenecks and that the deployment is capable of handling the highest offered traffic load and still has capacity left. In none of the 3 simulations there were any channels occupied in trunks connecting an MTX switch cluster with the rest of the network. Routing calls over these trunks was automatically avoided by the switches because there would have cost too many hops compared to other routes. This is a desirable feature because, as stated above, the MTX clusters represent switches of headquarter sites and routing calls over them should be avoided. Routing calls over the MTX clusters is in this deployment only expected when the network becomes congested.

Also, apart from the initialization at the start of the simulation, processornet statistics showed that the load of the processornet is not extremely high. At the start of the simulation all switches simultaneously generated a lot traffic to configure their routing tables, in reality this is not expected because they will not configure themselves exactly at the same instance. When the transient at the beginning of the simulation had disappeared, there was only one processornet message at the same time in the queues of the transmitters of the switches. These messages could be send immediately then. The processornet of the MTX cluster were not heavily loaded with external messages. Connecting switches send on average only 1 kbps to an MTX cluster, while 16 kbps is available.

2.2 Effect of a non-zero BER

In this section an example is given of the effect of a non-zero BER on the load of the processornet. As explained above, in ZODIAC the processornet messages are send in blocks of 32 bits. When bit errors in a block can not be correct using a FEC, a go-back-n mechanism is used to perform a retransmission of the previous n blocks. Subsequently, the load on the processornet channel in a trunk becomes higher when there is a non-zero BER. As an example, simulations when performed with a cluster of MTX switches. The deployment is shown in Figure 6.

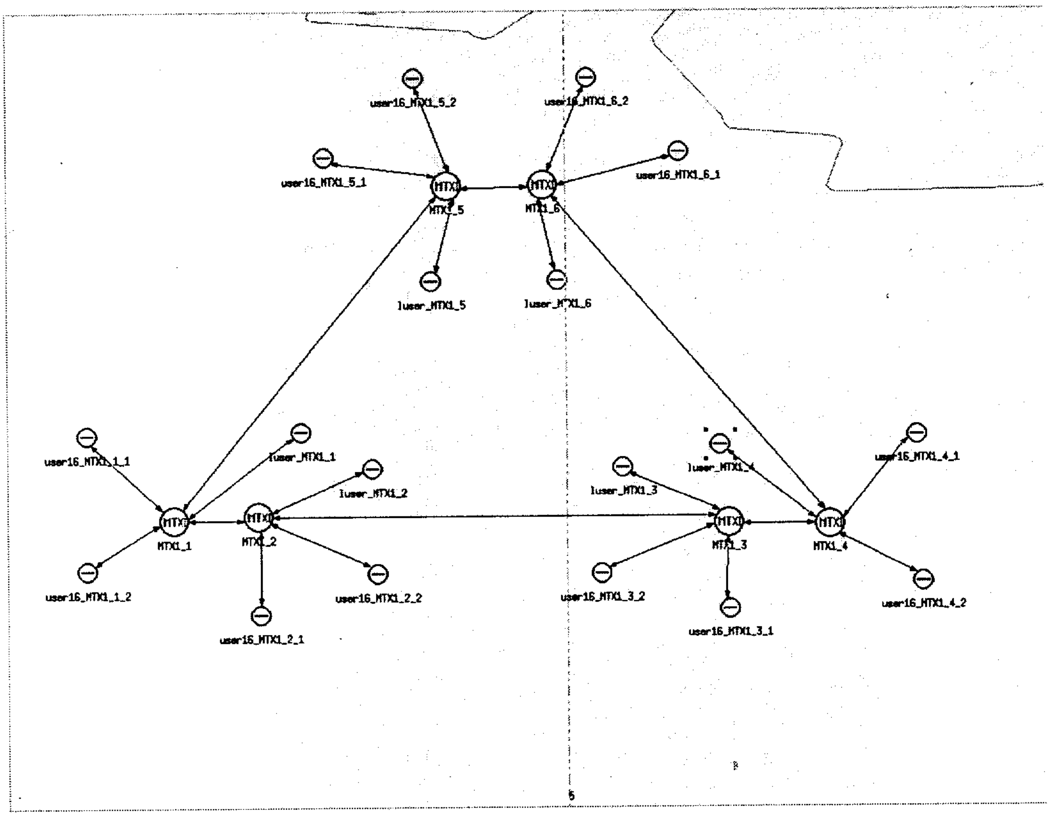


Figure 6 Topology of a cluster of MTX switches.

In the small cluster in this example there are 3 pairs of switches. Each stacked pair, for instance MTX1_1 and MTX1_2, are connected via a loopgroup. Also the usergroups are connected via a loopgroup to their local switch. The stacked pairs are interconnected via a with a model of trunk with a non-zero BER. In this simulation all 3 trunks had the same BER. Simulations where performed during which user in each group initiated calls with an average interval between calls of 125 seconds. The length of the calls was on average 25 seconds. Note that for a usergroup, with typically 30 members, the individual user on average initiates a call ones every hour. The total simulation time was 10 hours, during which 5200 call where initiated. On our Sun workstation with a 333Mhz Ultra SPARC processor, a simulation took about 200 seconds to perform. In total, about 23Mb of data was send over the processornet. The relative number of bits that had to be retransmitted is shown in Figure 7

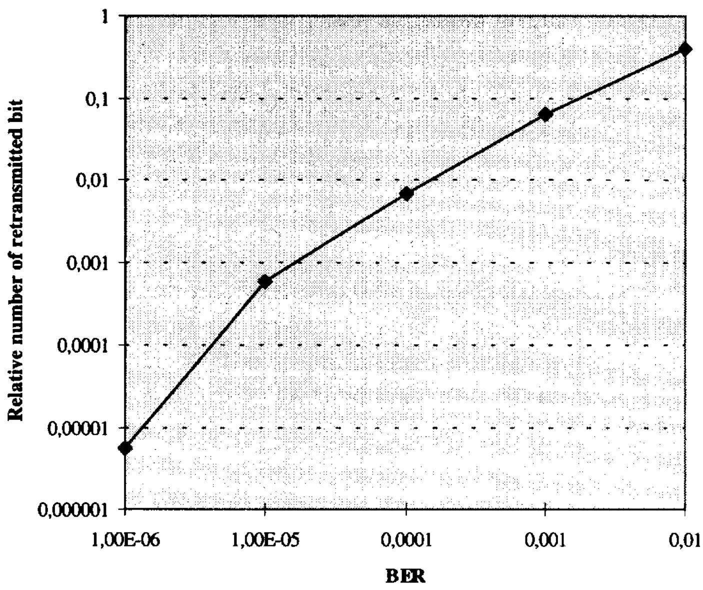


Figure 7 Total number of retransmitted bits relative to the total number of transmitted bits in the processornet versus the BER on the LOS trunks.

As is shown in Figure 7, at a BER of 0.01 about 40% of transmitted blocks of the processornet messages had to be retransmitted. No synchronization of the trunks were necessary.

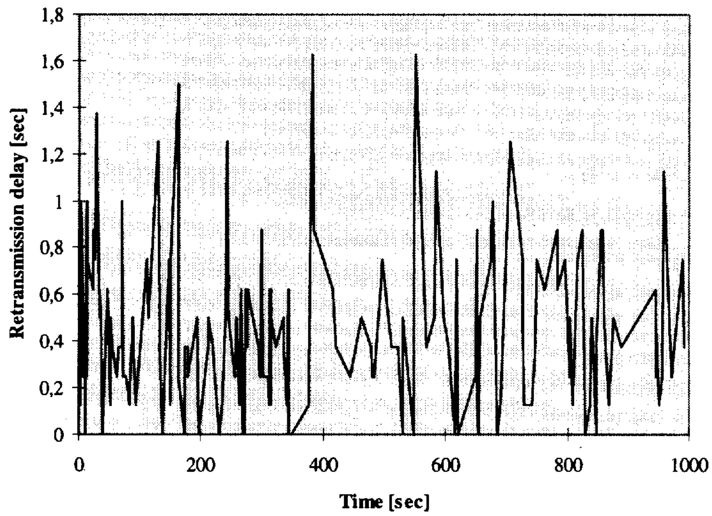


Figure 8 Delay in transmissions of processornet messages in a trunk with a BER of 0.01.

As an example, Figure 8 shows the delay which processornet messages can get when there is a BER of 0.01 on the trunk. Only part of the trace is shown here. Some of the messages could even get a delay of 2.375 seconds. Statistics of the queue of the transmitter showed that there where at maximum 512 bits waiting for transmission in the queue. For higher traffic loads the number of waiting bits can be higher.

The type of results discussed in this section can help the design or improvement of communication equipment and transmission protocols.

2.3 Novel data services on ZODIAC deployments

The RNLA asked TNO to perform a simulation study to gain insight in the requirements of a novel data service which is nowadays tested for ZODIAC. With this service, a user can get a LAN connection to a Central Router (CR) where data is stored and maintained. With there terminals, users can get information on the current status of numerous aspects of the mission. Points of concern were the impact of this new service would have on the workload of operators of ZODIAC switches and whether deployments would be able to make and maintain the required connections.

Typically 2 or 3 CR's, and at least one backup CR, are in the deployment. The CR's keep there data synchronized via backbone connections which are relatively high speed compared to the 16kbps LAN connections with the users. The data of this service can be routed through ZODIAC via various types of connections. One of them is to select the dedicated sole-user channels for the connections. These connections have the advantage that they are rarely used at the moment and that the ZODIAC system is designed to automatically reroute the connection when a trunk in their route fails. However, selecting the sole user connections implies an extra work load for the operators of the switches. These operators have to set up and monitor the sole-user connections and take action when the network can not automatically reroute the connections due to degradation or congestion. Simulations in the study were performed for a number of ZODIAC deployments of various sizes which might occur during missions at peace and war time. Methods used and some results obtained with this study will be discussed briefly in this section.

Discussions with officers of the RNLA made clear what might be some worst case deployment and scenarios under various circumstance. ZODIAC deployments were divided in deployments which can occur during peacekeeping operations, and deployments on a grant scale which might occur during war time. Another distinction was made in deployments which had 2 CR's and a backup CR and deployments with 3 CR's and a backup CR. Beside distinction between types of deployments, a number of traffic scenarios where decided upon. A deployment with 2 CR's and 2 satellitelinks which might occur during a peace keeping operation is shown in Figure 9.

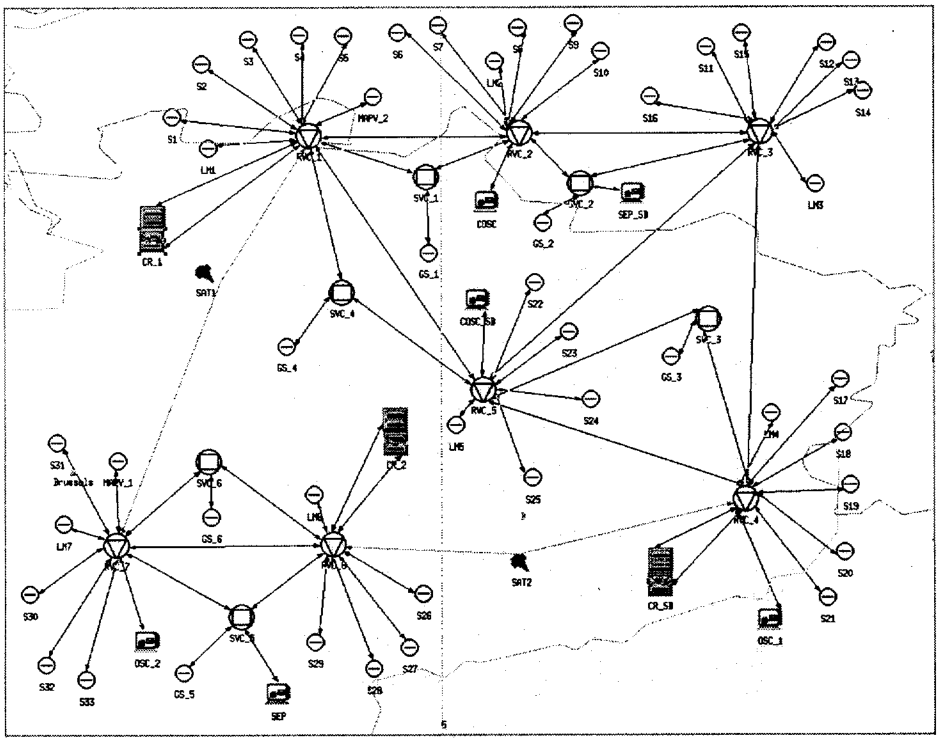


Figure 9 ZODIAC deployment with 2 satellitelinks which might occur during a peace keeping operation. Both islands in the deployment have one active CR's each. The backup CR is in the south-east of the largest island.

The 2 active CR's and the backup CR, called CR_1, CR_2 and CR_SB respectively, are represented in Figure 9 by the rectangular server icons. They are connected with 2 loopgroups to a switch so there are 2×31 channels available in the loopgroups for connections of which 2 channels are reserved for voice channels for the operators. An active CR's can thus handle up to 60 connections. In this deployment there are various types of users, the usergroups are again represented by the circle-bar icon and the circle-triangle icon switches. The circle-square icons represent switches for headquarter sites. The terminal icons represent SYSCOM usergroups which monitor the state of the ZODIAC deployment. Satellite icons, called SAT1 and SAT2, are placed next to the satellitelinks.

The performance, rerouting capabilities of the deployments where studied with the use of various traffic demands. The required connections between the CR's and between users and the CR's under standard operation was decided upon in discussions with officers of the RNLA. For instances, usergroups get 1 connection to a nearby CR. In the deployment shown in Figure 9 this means that usergroups must connect with the CR inside their island. Another connection rule during standard operation of the data service, is that headquarters sites get an extra connection to a different CR in the network without routing over satellitelinks otherwise it gets a connection to the same CR in the island. De active CR's themselves had to be connect 2×4 connections which then formed the backbone connection to keep the CR's synchronized, the backup router must be connected to a nearby active CR via 2 connections. When there 3 CR's in the deployment the backbone must have a ring topology. The SYSCOM groups are connected to 3 other SYSCOM groups via sole user connections to form a separate network and also get a connection to a CR.

All these rules are applied to the usergroups and CR's present in a deployment and put in a traffic matrix. The traffic matrix defines the begin and end points of the connections according to the connection rules. When the simulation starts, the models of the usergroup in the deployment read the traffic matrix and select the lines which define their connections and request their local switch to create these connections for them. The switches in the deployment will try and set up the connections according their routing tables. The route for the connections is thus decided by the switches in which the routing mechanism of ZODIAC is modeled. The relatively simple set of connection rules are thus applied to complex deployments where the (re)routing mechanism modeled in ZODSIM can be exploited to get an insight in potential problems that the novel data service might cause.

The following histogram shows the distribution of the number of occupied channels in the deployment after setting up all the, according to the traffic matrix, required connections in the LOS radio links (trunks).

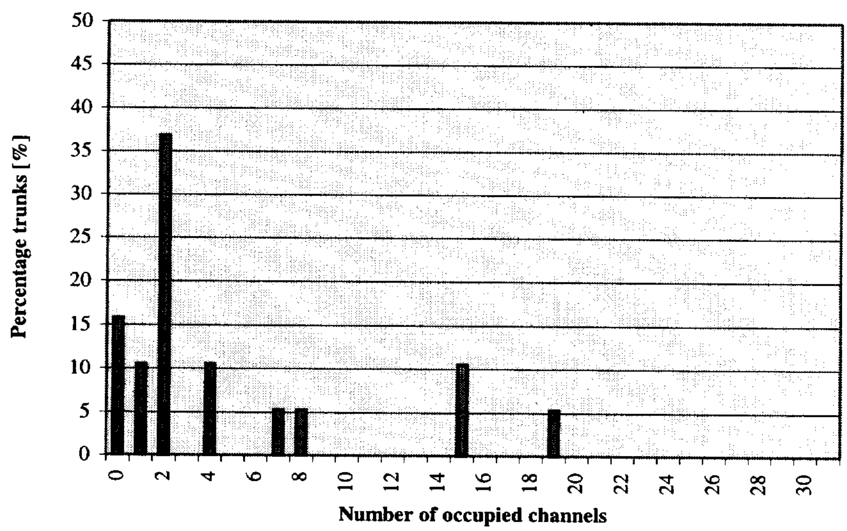


Figure 10 Distribution of the number of occupied channels in the peace time deployment with 2 CR's. The deployment is divided in 2 islands which are connection via 2 satellitelinks.

Figure 10 shows that in none of the trunk all available channels are fully occupied. Simulations in which trunk failure occurred after all connections where setup, showed that the deployment had enough capacity left to reroute connections as long as there where 2 trunks left to connect the local switch of the active CR's to the network. This indicates that the workload for operators is not increased by individual link failures in this deployment.

In the satellitelinks via SAT1 and SAT2, there where 11 and 2 channels occupied respectively by processornet channels, connections for the SYSCOM usergroups and the backbone connections. Simulations showed that when one of the satellitelinks failed, the network was able to reroute the required connections over the remaining one. After rerouting, 12 channels were occupied in the remaining satellitelink. Note that the processornet channel of the failed link is not rerouted. These results show that the data service requires satellitelink with a data rate in the order of $12 \times 16 \approx 2$Mbps in order to cope with link failures.

When all the standard required connections where setup, the loopgroups to the active CR_1 and CR_2 had 7 and 35 channels free for additional connections. This is not enough to provide with the headquarters with the additional connections which was proposed in a variation on the set of connection rules. This alternative set of connection rules prescribed that the headquarters where connected to the CR's via double 64 kbps links, or two set of 4 sole user links. In the large island this would require 4×6 channels in addition to the 2 each headquarter site already had. Note that the requirement in this variation could be fulfilled if additional connections were allowed to make use of the satellitelinks.

Other insights where also gained when the proposed simple set of connection rules where applied to deployments which might occur during missions. Examples of other important ones where bottlenecks which occur when a CR has be moved to a switch elsewhere in the deployment without the users loosing their data service and estimates of the workload that imposes on the operators.

3. Discussion and lessons learned

The examples discussed in this paper, show that the ZODSIM simulation model can be used to estimate various properties of ZODIAC deployments, predict their capacity and potential bottlenecks. In the simulation model it is easy to create deployments and run simulations. The OPNET simulation environment was used to create the simulation program and to analyze results.

Our experience shows that designing and programming of a simulation model from scratch takes a lot of the effort and time spend in a project. It is also very important to determine at the beginning of the development of a model which behavior, parameters, and use of the network elements are essential for the model and which can be ignored initially but have to be implemented in future refinements of the model. When a model which can be adapted to suit another project is present, the development time of a suitable model can be greatly reduced. Provided that, as in all computer programs, names of parameters, subroutines et cetera, were chosen logically and not abbreviated to a few letters of the alphabet. If programmers try to save time by using short abbreviations for names and typing sparse comments when they are introduced in the code, their future colleague who has to make modifications will annihilate the gain.

The GUI interface of OPNET and provided kernels shortens the development time of models and creating deployments. Navigating through the simulation code is make easy when the finite state machines behind the model is designed logically. We noticed that excessive use of states in a model should be avoided. When a model becomes to complex, it might be advisable to separate the functionality of the object into separate nodes and put these in a additional subnet layer. Using the OPNET simulation environment gives the advantage that models of COTS equipment can be used to model the use of COTS equipment in ZODIAC in the future. Simulating ZODIAC deployments with old an new equipment thus requires then less effort. However, a potential user should be aware that before he or she can develop models independently, quite some time is needed to learn the OPNET environment, the philosophy behind the way models are designed, and which tools and tricks are available to create them. Knowing the existence and purpose of the various kernels and tricks of the trade which are available to create models can reduce the complexity of the programs behind them.

During the development of the network elements for a model such as ZODSIM, it is important to keep contact with the customer and discuss the implemented features and show intermediate results. The developer of models of a network such as ZODIAC can in these discussions find out whether he or she understands the features and has implemented them properly. Systems such as ZODIAC can have features which are not well known to the present day operator. Also, information about features of switches which are usually inaccessible to the everyday operator but are needed to tune the model, such as transmitter queue sizes, can be hard to come by. Getting the details right from the start is very important for the model to be a model of the real system. Not only because it can save development time and because modifying an existing model easily induces errors, but also because it can very difficult to get data with which the model can be verified.

With the tools provided in the OPNET environment it is relatively easy for the programmer to create a large amount of statistics about a simulation and create impressive graphs and animations. However, he or she should always be aware the his customer is not accustomed to find is way is a sea of data with the wink of an eye. The programmer and simulator should see it as a challenge to find out what is most important to the customer and then present it in a simple way. The programmer and the persons who perform the simulations and digest the results should always be aware that they have sometimes been working for months with details of the problem and are used to focus in on important features and ignore the rest for the time being. They should also act as the filter for their customers and not only for themselves. We hope we have fulfilled that goal in this paper.

Modeling terrestrial mobile networks in real terrain

Corresponding author: Christian Jönsson
Co-authors: Jimmi Grönkvist, Ulf Sterner,
Eva Englund, Anders Hansson, Ingrid Söderquist

Swedish Defence Research Establishment
Div. of Command & Control Warfare Tech.
SE-581 11 LINKÖPING, Sweden

Abstract

The presentation given here, reflects the current state of an ongoing research project, with focus on tactical multi-hop radio near ground networks. For terrestrial mobile networks the terrain has a large impact, not only on radio link level, but also on network design and protocols. Our aim is to estimate and simulate network performance, and to evaluate network designs and protocols in a real terrain environment. Here, we describe our method and models for simulating network performance in digitized terrain. The method is based on estimating the basic-path-loss. It utilizes a digitized terrain and advanced electromagnetic propagation models to model the impact of terrain. We present the proposed methods as well as an overview of current work, including: spatial reuse access schemes, mobility model and analyses of the duration of an access protocol.

1 Introduction

In future military operations, the need to quickly acquire and assimilate information at all levels of the commands hierarchy is foreseen. In many information systems projected, low level units are expected to make well informed decisions. Combat information and sensor data must be available even at the level of the individual soldiers.

In crucial situations, for example communications on the battlefield itself, there will be a need for high performance wireless communication without the support of a preinstalled infrastructure. In these cases, a radio network should be able to be deployed in unknown terrain and with a minimal need of network preplanning. The radio units may be spread out in the terrain in an ad hoc manner and line-of-sight communications cannot be guaranteed. The terrain, in which the network is deployed, may have a large impact on the electro-magnetic propagation environment. Our work aims at estimating and simulate the effects of real terrain on network performance and to evaluate near ground radio network designs and protocols.

One method to obtain area coverage in these type of networks is to relay messages through one or several intermediate nodes. Networks with this property are called *multi-hop* networks. Multi-hop properties offer an important advantage in robustness in tactical radio networks. Most work on multi-hop networks described in literature have in common that they use very simple propagation models that do not take the terrain into account.

In our presented network model, we use a digitized terrain representation and Vogler's multiple knife-edge diffraction model to describe the electromagnetic propagation characteristics. The propagation model provides an estimate of the basic path-loss for each pair of units in the network. These estimates of the basic path-loss provide the possibility to estimate and simulate the effects of real terrain on network design problems. Also, different proposed protocols can be evaluated in a digitized real terrain environment.

In the following, we present an overview of accomplishments and ongoing work with the proposed network model, including:

- Design and evaluation of terrain adaptive spatial reuse access schedules.
- Modeling mobility in digitized terrain.
- Simulating the effects of mobility in network protocols.

The first topic, considers a "snapshot" view of the network, and describes methods of obtaining spatial

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

reuse access schedules in a multi-hop network. The approach here is to reuse channel resource whenever the terrain separates units so that little interference is caused. The second topic, describes a mobility model that takes the terrain into account. At last, the described setup is used to estimate the effects of mobility on an access protocol.

2 Propagation model

An essential part of modeling an on ground or near ground radio network is to model the electromagnetic propagation characteristics due to the terrain variation. A common approach then is to use the basic path-loss, L_b, between two radio units. The most simple assumption concerning the wave propagation is that no obstacles appear between the transmitting and receiving antennas, and no reflection or diffraction exist in the neighborhood of the direct path between the receiving and transmitting antennas. This model is the so called free-space assumption. For this case, the basic path-loss is given by the following well-know equation, e.g., see [1, Eq. (2.4), p. 18].

$$L_{bfs} = \frac{(4\pi df)^2}{c^2}.$$

Here we use d as the distance in meters between the transmitter and receiver antennas, and f to denote the center frequency of the transmitted narrow-band signal, and c is the speed of light. This means that if the efficient isotropic radiated power of the transmitting antenna is P_{EIRP}, then the efficient isotropic collectible power is given by $P_{EICP} = P_{EIRP}/L_b$, which for the free-space assumption yields $P_{EICP} = P_{EIRP}/L_{bfs}$.

To obtain a finer model of near or on ground propagation, a single ground reflection is taken into account. Doing so, we use the two-ray plane earth assumption. In its simplest form, perfect lossless reflection is assumed. When the distance d is large relative the antenna heights, h_1, h_2, we have the following estimate of the basic path-loss, see [1, Eq. (2.22), p. 25].

$$L_{bpe} = \frac{d^4}{(h_1 h_2)^2}.$$

Refined assumptions of wave propagation conditions near ground include terrain height information and terrain type information to estimate the basic path-loss. In the digital terrain database we use, the height is represented as terrain height samples at equidistant square lattice intersection points, and the terrain type is one of maximum 256 terrain types such as fresh water, salt water, forest, wet ground, open rocks, etc. We assign electro-magnetic ground constants such as relative dielectric constant ε_r and conductivity σ for each terrain type and we also assign surface roughness for terrain types. See Parsons' [1, Sec. 2.3] for an introductory description of the problem. Our database has a 50 meter grid for the terrain height samples and a 25 meter by 25 meter terrain type area.

All our calculations are carried out using the ground wave propagation library DetVag-90®, [2]. Here, we use a multiple knife-edge model of Vogler [3, 4] with five knife edges.

3 Network model

We consider a radio network consisting of N radio units spread out in some terrain. We use the notation suggested by Jönsson in [5], where the units are described as *nodes*. A *node* v is characterized by a set of attributes, a pair of coordinates describing the location in some coordinate system x, y, the used transmitter power P, the transmitting antenna gain diagram $G_t(\theta, \varphi)$, the broadband equivalent single sided receiver noise power spectral density N_r, the receiving antenna gain diagram $G_r(\theta, \varphi)$, the antenna height h, and the polarization of the transmitted electrical field *pol*. Mainly refer to a node v_i as $v_i = (x_i, y_i)$ and denote the set of nodes by V.

Our model of a radio network is the set of nodes V and the basic path-loss $L_b(i,j)$ between any two nodes.

For any two nodes, (v_i, v_j) where v_i is the *transmitting* node and $v_j \neq v_i$, we define the signal to noise power ratio (SNR), Γ_{ij}, as follows

$$\Gamma_{ij} = \frac{P_i\, G_t(i,j)\, G_r(i,j)}{L_b(i,j)\, N_r}, \qquad (1)$$

where P_i denotes the power of the transmitting node v_i, $G_t(i,j)$ the antenna gain of node v_i in the direction of node v_j, $G_r(i,j)$ the antenna gain of v_j in the direction of v_i, and $L_b(i,j)$ is the basic transmission path-loss between nodes v_i and v_j. For convenience, we define $\Gamma_{ii} = 0$ corresponding to the physical situation of a node not being able to transmit to itself.

Unless stated otherwise, we assume isotropic antennas, that is

$$G_t(i,j) = G_r(i,j) = 1.$$

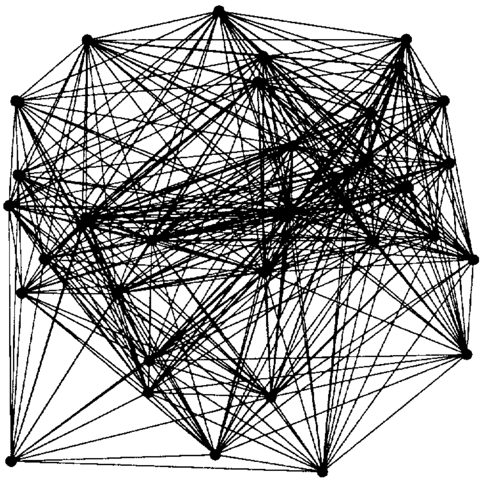


Figure 1: The figure illustrates the set of links, $\mathcal{L}$, obtained with a plane earth model, for $\gamma_0 = 1$ and $P/N_r = 10^{13.5}$.

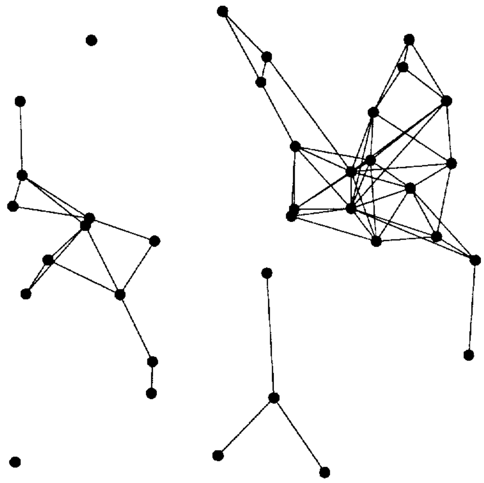


Figure 2: The figure illustrates the set links, $\mathcal{L}$, obtained with a terrain dependent propagation model, for $\gamma_0 = 1$ and $P/N_r = 10^{13.5}$.

The model we propose is based on the assumption that the SNR is a good measure of the communication quality of a radio link. This is a reasonable assumption provided that the radio link technology employed takes appropriate measures to handle multipath propagation. For the radio link, we assume that the link provides perfect transmission for SNR values down to a certain threshold. We say that a pair of nodes (v_i, v_j) form a *link*, (i, j), if the signal to noise power ratio (SNR) at the receiver is not less than a threshold, γ_0. That is, the set of links, $\mathcal{L}$, is defined as follows:

$$\mathcal{L} = \{(i, j) \text{ such that } \Gamma_{ij} \geq \gamma_0\} \ . \qquad (2)$$

3.1 Network examples

The examples given in Figures (1) and (2) illustrate the difference in network topology obtained when using two different propagation models. In Figure (1), a plane earth model, as implemented in DetVag-90® [6, 7], is used and in Figure (2), Vogler's multiple knife edge model. The node positions are randomly chosen according to a uniform distribution. The terrain is hilly with mixed forest and meadows.

In both Figures, the center frequency is chosen to be 450 MHz, antenna heights 3 m, P_i and N_r constant such that $P/N_r = 10^{13.5}$, and $\gamma_0 = 1$. The examples show that there is a great difference in network connectivity obtained for the same transmitter power and link threshold.

3.2 Connectivity criterion

In multi-hop networks, the choice of transmitter powers, antenna gains etc. lead to a variety of different topologies and link capacities. In network design, the problem of choosing appropriate transmitter powers and antenna gains with respect to optimizing some cost function is a very complex procedure, involving assumptions on access and routing procedures. One common approach, is to separate this design problem and choose the topology with respect some connectivity criterion. Here, we describe the connectivity criterion we use to choose transmitter power and give some examples. Again, we compare a plane earth model with Vogler's multiple knife edge model.

We use a normalized 1-connectivity measure, Φ. It estimates the probability that two arbitrarily chosen radio units can communicate. This does not mean that the two units necessarily must be able to establish a link. Units that can reach each other through a relaying procedure are considered as well.

The normalized connectivity measure relies on a path indicator function. This function, denoted p_{ij}, is one if there exists a path from the source node v_i to the destination node v_j. Otherwise, it is zero. As a normalized measure of the network's connectivity we define

$$\Phi = \frac{1}{N(N-1)} \sum_{\text{for all } i,j} p_{ij} \ ,$$

where N is the number of nodes in the model. This

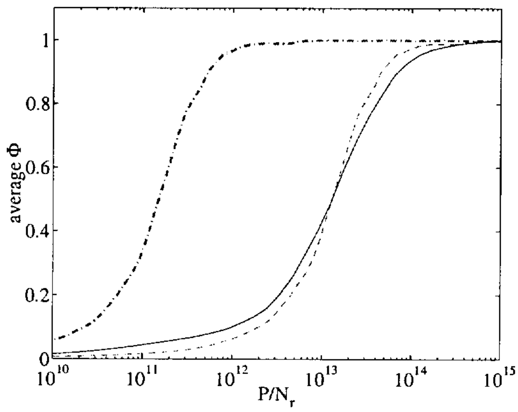


Figure 3: The plot shows normalized connectivity, Φ, in average. The solid line is obtained with the multiple knife edge model and the dash-dot line is obtained with the plane earth model. The dashed line is a translation of the dash-dot line to coincide with the solid line for average $\Phi = 0.5$.

measure can be seen as the "probability" that two randomly chosen nodes can communicate, i.e., there exists a path between them.

Figure (3) shows our connectivity measure, Φ, in average, over a sequence of 1000 networks, versus P/N_r for link threshold of $\gamma_0 = 1$. The Figure shows that there is a great difference in transmitter power needed in order to achieve a certain normalized connectivity. Compare the dash-dot line with the solid line in Figure (3).

In Figure (3), the dotted line is a translation of the curve obtained for the plane earth assumption (dash-dot line) to coincide with the curve for the terrain dependent model (solid line), at connectivity 0.5. Comparing this dashed line with the solid line we can see that an increase in transmitter power does not totally describe the impact of the terrain on the connectivity measure.

4 Interference based reuse schedules

One problem in a radio network is the interferences caused by simultaneously transmitting nodes. These conflicts occur if the received signal is too weak compared with the interfering signals. An important issue is therefore to design access schedules that control the use of the channel. An often used access schedule is time division multiple access (TDMA). For multi-hop connected networks, however, this is usually inefficient. To increase capacity one can instead use spatial reuse TDMA (STDMA), which is an extension of TDMA where the capacity is increased by spatial reuse of the time slots.

The STDMA channel access schedule for multi-hop packet radio networks was introduced in [8]. Different algorithms for generating STDMA schedules have been proposed. Two of these are described in [9, 10].

The multi-hop properties of the network introduces another problem. The traffic load on the links in the network can vary considerably, even if the incoming traffic to the network is uniform. Therefore traffic sensitive, or traffic controlled access schedules, have been proposed [11, 12].

Here, we give a brief description of algorithms and results developed by Grönkvist in [13, 14]. The work presented here is based on the signal-to-interference ratio (SIR) in each link, estimating the basic path-loss. This approach differs from previous work where the network graph is used to decide if interferences occur. That approach does not capture the total interference in the network. The use of SIR gives a possibility to take the terrain into account in a more realistic way.

One of the two algorithms is traffic controlled. We investigate by simulations how the traffic controlled schedules increase the capacity of the network. Another advantage of using SIR is that power control and the use of directional antennas can be straightforwardly included. However, these issues will not be investigated here.

4.1 Notation

For a set of links, $L \subseteq \mathcal{L}$, and for any link, (i,j) in L, we define the *interference* as follows

$$I_L(i,j) = \sum_{(k,l)\in L\setminus(i,j)} \frac{P_k\, G_t(k,j)\, G_r(k,j)}{L_b(k,j)} .$$

Furthermore we define the receiver *signal to interference power ratio* (SIR) as follows

$$\Pi_L(i,j) = \frac{P_i\, G_t(i,j)\, G_r(i,j)}{L_b(i,j)(N_r + I_L(i,j))} .$$

We assume that any two radio units can communicate a packet without error if the SIR is not less than a threshold, γ_1. More precisely, the following condition must hold

$$\Pi_L(i,j) \geq \gamma_1 \quad \text{for all} \quad (i,j) \text{ in } L. \qquad (3)$$

Further we assume that a node cannot transmit more than one packet in a time slot and that a node cannot receive and transmit simultaneously in a time slot. This gives us the following condition

$$\text{If } (i, j) \in L \text{ then no other link containing either } i \text{ or } j \text{ can be in } L. \quad (4)$$

If the above two conditions, (3) and (4), hold for a set of links L we say that the set of links can *transmit simultaneously*.

4.2 The Algorithms

Here we give an outline of the algorithms used, for more details see [13]. Roughly, the basic idea is to loop through the time slots. For each new loop a time slot, numbered k, is created. The algorithm first checks to see if the links that have not yet received a time slot can be assigned to this slot without violating the interference requirements. In the next step the same check is performed on the rest of the links. Eventually, all the links will have received at least one time slot and the procedure terminates.

The output of the algorithm is the sets L_k for all slots $k = 1, 2, \ldots T$ where T is the period of the schedule. When the algorithm terminates the sets L_k contain the links that are assigned to time slot k.

Algorithm I, is a refined version of the basic idea, described above, where links are assigned time slots according to a priority list. The priority of a link is based both on the relative traffic load and on the number of time slots passed since the link previously was assigned a slot. With this procedure the slots assigned to each link will be spread out evenly over the period, resulting in a decreased average delay.

To make a traffic sensitive schedule, some links with heavy traffic are allocated more than one time slot. We guarantee the links a number of time slots depending on the traffic on the link. This is done in Algorithm II. All links will have at least this many time slots in the final schedule. In Algorithm I, which is not traffic controlled, the guaranteed number of time slots is one for each link.

4.3 Simulation results

Simulations have been used to evaluate the average delay, D, of a message in two different networks, networks 1 and 2 described in Figures (4) and (5), respectively.

In the simulations, additional assumptions have been made:

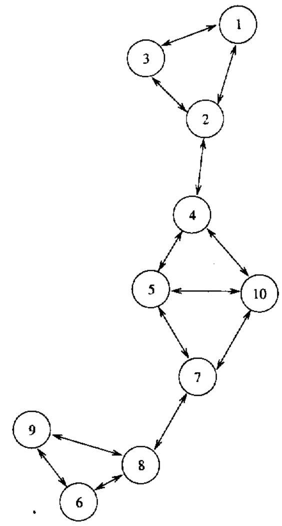


Figure 4: Network 1.

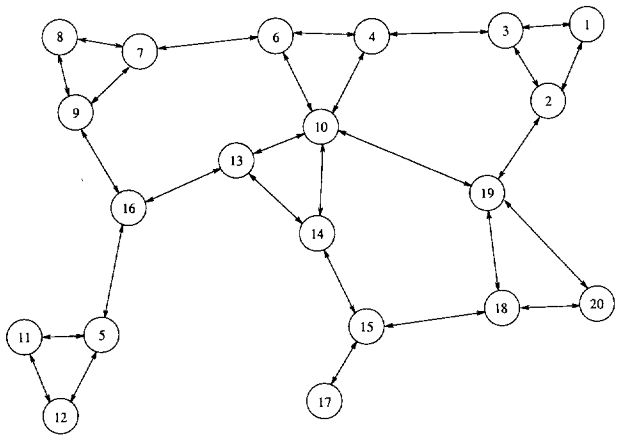


Figure 5: Network 2.

Routing has been made by shortest path route, i.e., a message between two nodes will always use the path which requires the least number of transmissions. If several routes of the same length exist, all messages between two specific nodes will always use the same route. The relative traffic of each link is then defined to be the number of routes where it is included.

The messages are assumed to be generated according to a Poisson process with intensity λ. Since the schedule is collision free, all messages are assumed to be perfectly received, i.e., no retransmissions in that sense are considered. The average delay of a message is measured in time slots from arrival to the network until it reaches its destination. The two STDMA schedules are also compared with a TDMA access schedule. The average delay for networks 1

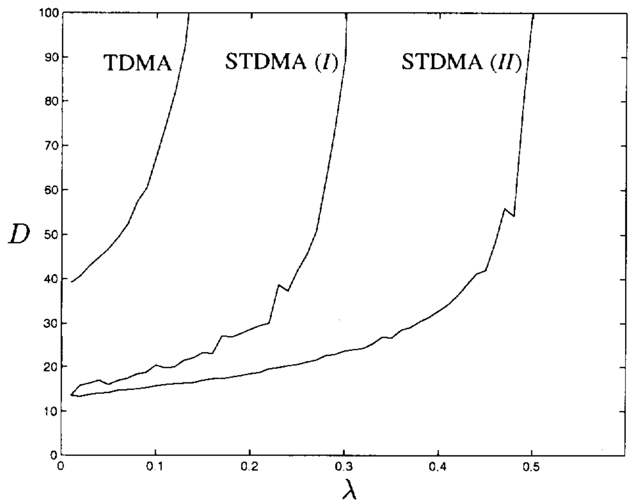


Figure 6: The result from the simulation on Network 1. We have plotted the average delay D measured in time slots as a function of the throughput λ in packets per time slot.

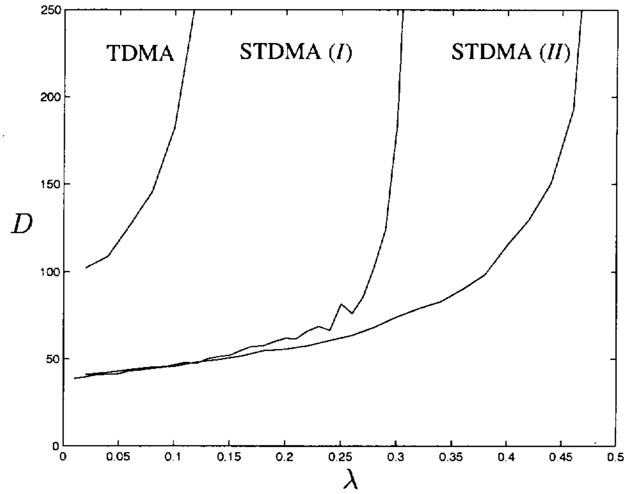


Figure 7: The result from the simulation on Network 2. We have plotted the average delay D measured in time slots as a function of the throughput λ in packets per time slot.

and 2 can be studied in Figures (6) and (7), respectively. As can be seen, the average delay of a message is considerably decreased for both algorithms compared to a TDMA schedule, though algorithm II can achieve a much higher throughput than algorithm I.

5 Modeling mobility in tactical radio networks

Modeling the movements of the units in a tactical radio network is another issue where the terrain has an essential impact. Here, we outline the work of Sterner given in [15] where a digitized terrain is used when modeling the movements of nodes.

The main idea is to search for the shortest path, according to some criterion, between two given positions. The path chosen between the points is obtain by assigning different weights to different types of terrain. The weights can be assigned according to the difficulty of moving through the terrain, perhaps combined with tactical considerations on movements of units. An algorithm finding the "shortest path", given these weights, is then applied.

The terrain is, in our case transformed into a network description, thereafter the shortest path is found with traditional search methods. The procedure can be described by the following three steps:

Step 1: Assign weights Divide the terrain representation into squares and assign each square a weight according to the characteristics of the terrain.

Step 2: Compose network Let each square represent a "node". Neighboring squares are connected with "edges" where the weight of an edge is the average of the connected squares. The weight of the diagonal edges are increased by the square root of two.

Step 3: Search path The path can now be found by applying a "shortest path" algorithm in the network description obtained in step two.

This method gives a good path, in "shortest path" sense, however there are problems in implementing the method. To obtain accuracy enough the terrain representation must be divided into small segments and the problem of finding shortest path becomes infeasible, even with efficient algorithms.

An approach to this problem, is to relax the demand for finding the overall shortest path. Here, we use efficient suboptimal searches described in [16, 15].

The search method uses a sequence of terrain descriptions, with increasing resolution. First a path is found according to some shortest path criterion in the terrain description with least resolution. Second, this path is refined in the next terrain description, and so on according to the following steps:

Step 1 Create a sequence of terrain descriptions with increasing resolution.

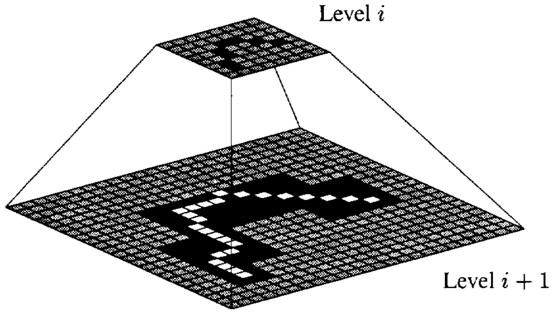


Figure 8: Two levels of resolution in the map search procedure.

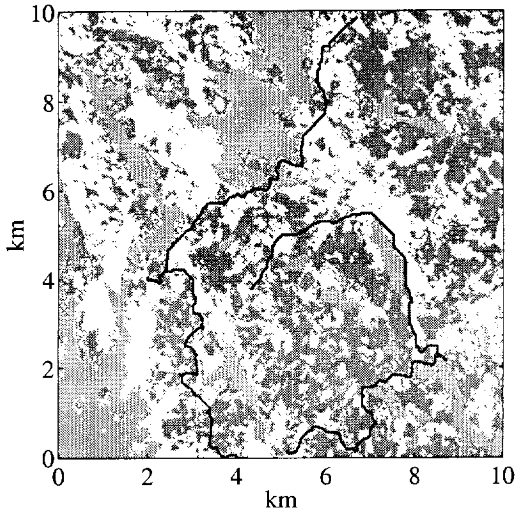


Figure 9: Examples of paths obtain with map search methods. The dark gray areas correspond to more difficult terrain, the light grey areas are lakes and watercourse and white areas correspond to light weighted terrain.

Step 2 Find the shortest path in terrain description with least resolution level i, $i = 1$.

Step 3 Place the path found, on the next resolution level $i + 1$ in the sequence. Apply a new search within the vicinity of this path. A new, more detailed path, is obtained.

Step 4 Repeat step 3 until desired resolution is obtained.

This procedure is visualized in Figure (8).

Two examples on paths obtained with the search described above are given in Figure (9). For comparison, Figure (10) shows some examples of paths generated from a random walk model, see [17]. The procedure described, can be improved further by performing an iterated search. For more details see [15].

6 Mobility in reuse schedules

In stationary networks, an STDMA protocol results in a conflict free access scheme. As the radio units move, the interference environment changes and eventually the protocol must be updated. To estimate the duration of an STDMA protocol we need a model of the interference environment. Again, we use two different models for the basic path-loss to illustrate the influence of the terrain on the performance of a mobile multi-hop network. The duration of an STDMA protocol is here defined as the time elapsed until the interferences exceed a certain threshold.

The reuse schedule, L_k, for $k = 1, 2, \ldots, T$ is assumed to be fixed under its duration. No new links are introduced in the protocol so condition (4) will automatically hold as time elapses. However, condition (3) will soon be violated. One simple approach to handle this problem is to design the schedule with a certain margin, called mobility margin γ_m. This margin will guarantee that conflicts in terms of SIR do not appear at slightest change in interference environment. To attain a mobility margin we demand a higher value on SIR when the protocol is designed, $\gamma_1 = \gamma_g$, than what we demand for operating the protocol γ_{op}. The margin obtained is $\gamma_m = \frac{\gamma_g}{\gamma_{op}}$.

The mobility margin will result in an STDMA protocol which is not as *tightly packed*, i.e. the number of links that can use the same time slot will decrease with increasing mobility margin. This means that the reuse efficiency is decreased.

As a measure of the interference environment we study the minimum SIR for the time slots in the access protocol. We say that an access protocol is conflict free, in terms of SIR, if this minimum exceeds the threshold γ_{op}.

Further we define the duration of an STDMA protocol as the time between generation of the protocol and the first time a conflict in terms of SIR appear in the protocol. The duration of an STDMA protocol can then be seen as the time it takes for the minimum SIR of the access protocol to decrease from γ_g to γ_{op}.

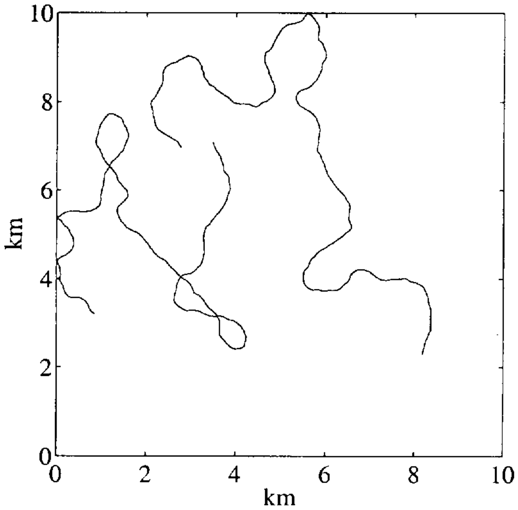


Figure 10: Two examples of paths generated by a random walk model.

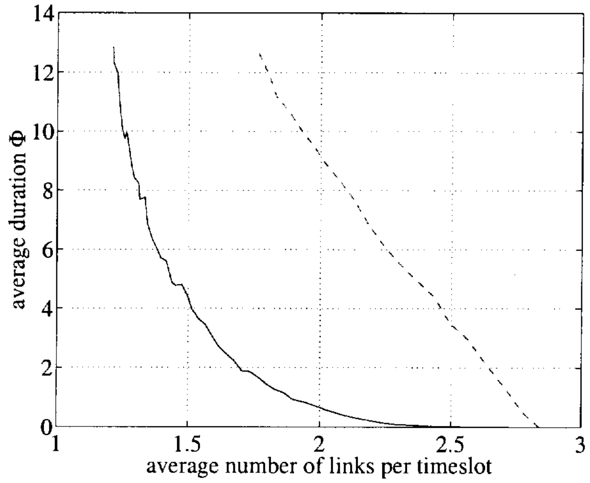


Figure 11: The plot shows the average duration of an STDMA protocol. The dashed line shows the result where the plane earth model is used and the solid line shows the result for Vogler's model.

6.1 Performance evaluation

In our simulations, we estimate the average duration for a STDMA protocol with respect to the reuse efficiency.

The simulations are run on a network with 36 nodes that move according to the mobility model described in Section 5. Furthermore, the nodes movements are constrained to an area of 10 km $\times$ 10 km. To calculate the basic path-loss L_b between nodes we use a plane earth model and the terrain dependent multiple knife edge model of Vogler.

6.2 Simulation Results

To get a measure of the network's connectivity for a simulation run, we use the time average of the connectivity measure suggested in section 3.2. We choose transmitter power such that this average connectivity is 0.9.

We present the simulation results in Figure (11). As a measure of the reuse efficiency we use the average number of links that are assigned to the same time slot. The results from the simulation with the plane earth model are represented by the dashed line. The results from Vogler's model is represented by the solid line. Studying Figure (11), we can see that for a given reuse efficiency the average duration is dramatically decreased when the terrain is considered while estimating the basic path-loss.

7 Conclusion and Comments

Simulation results show that traffic controlled spatial reuse of time slots, provides a substantial decrease in delay and increase in throughput. Sometimes as much as a factor of five in increased throughput.

Also, our simulations show that how frequently an STDMA protocol must be updated is highly dependent on the characteristics of the terrain.

It can be concluded that the method suggested for handling variations in interference environment may very well be used in a flat and simple terrain, where the loss in reuse efficiency is moderate. However, it is not sufficient for handling mobility in a more difficult terrain, the loss in reuse efficiency is not acceptable and other solutions must be sought. Probably, the demand for a strictly conflict free scheme is inefficient in this case. In our future work, the proposed method will be combined with conflict resolution.

In general, our experience while working with our models and methods suggest that it is important to use terrain dependent propagation models to estimate the performance of network protocols. We feel that this type of study, bridging over traditionally separated disciplines, are fruitful an offers mutual insight and understanding.

We hope that this presentation has given a flavor of the type of results and accomplishments we strive to achieve in our project.

References

[1] J.D. Parsons, *The Mobile Radio Propagation Channel*, Pentech Press, London, 1992.

[2] B. Asp, G. Eriksson, and P. Holm, "Detvag-90® — Final Report," Scientific Report FOA-R–97-00566-504–SE, Defence Research Est., Div. of Command and Control Warfare Tech. Linköping, Sweden, Sept. 1997.

[3] L.E. Vogler, "An attenuation function for multiple knife-edge diffraction," *Radio Science*, vol. 17, no. 6, pp. 1541–1546, Nov.-Dec. 1982.

[4] J.H. Whitteker, "Fresnel-kirchhoff theory applied to terrain diffraction problems," *Radio Science*, vol. 25, no. 5, pp. 837–851, September-October 1990.

[5] C.C. Jönsson, "Mobile tactical radio networks as directed graphs," Scientific Report FOA-R–97-00532-504–SE, Defence Research Est., Div. of Command and Control Warfare Tech. Linköping, Sweden, Nov. 1997, In Swedish.

[6] S. Rotheram, "Ground-wave propagation part 1," in *IEE Proc. F, Commun., Radar & Signal Process.*, Oct. 1981, vol. 128, pp. 275–284.

[7] S. Rotheram, "Ground-wave propagation part 2," in *IEE Proc. F, Commun., Radar & Signal Process.*, Oct. 1981, vol. 128, pp. 285–295.

[8] R. Nelson and L. Kleinrock, "Spatial-TDMA: A collision-free multihop channel access protocol," *IEEE Trans. Commun.*, Sept. 1985.

[9] A. Ephremides and T. Truong, "Scheduling broadcasts in multihop radio networks," *IEEE Trans. Commun.*, Apr. 1990.

[10] I. Chlamtac and A. Lerner, "A link allocation protocol for mobile multi-hop radio networks," in *GLOBECOM '85, IEEE Global Telecommunications Conference, Conference Record*, 1985, vol. 1, pp. 238–242.

[11] J. Shor and T. Robertazzi, "Traffic sensitive algorithms and performance measures for the generation of self-organizing radio network schedules," *IEEE Trans. Commun.*, Jan. 1993.

[12] O. Somarriba, *Multihop Packet Radio Systems in Rough Terrain*, Tech.lic. thesis, Radio Communication Systems, Department of S3, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, Oct. 1995.

[13] J. Grönkvist, "Traffic controlled spatial reuse TDMA for multhop radio networks," in *Personal, indoor and Mobile Radio Communications*, 1998, pp. 1203–1207.

[14] J. Grönkvist, "Spatial time multiple access - algorithms and evaluation," Tech. Rep. FOA-R–98-00781-504–SE, Defence Research Establishment, Division of Command and Control Warfare Technology, P.O. Box 1165, S-581 11 LINKÖPING, SWEDEN, June 1998, In Swedish.

[15] U. Sterner, "Mobility models for mobile radio networks," M.S. thesis, Linköping University, S-581 83 Linköping, Sweden, Dec. 1998, In Swedish.

[16] F. Markus Jönsson, "An optimal pathfinder for vehicles in real-world digital terrain maps," M.S. thesis, KTH, Royal Institute of Technology, SE-100 44 Stockholm, Sweden, 1997.

[17] U. Sterner, "Mobility in spatial reuse multihop networks," in *SNRV's Radiovetenskapliga konferens, RVK*, RVK 99, Högskolan i Karlskrona/Ronneby, SE-371 79 Karlskrona, rvk99@hk-r.se, June 1999, Accepted.

Performance Analysis of Direct Sequence Spread Spectrum and Non-Spread Spectrum Techniques in Frequency Selective Fading Channels

Erdal TORUN
Turkish Land Forces Command
Technical & Project Management Department
Yücetepe 06100 Ankara, Turkey
etorun@kkk.tsk.mil.tr

H.Orkun ZORBA
Aydın Yazılım ve Elektronik San. A.Ş.
R & D Department 13.Cad.No.10
Beysukent 06530 Ankara, Turkey
orkunz@ayesas.com

Abstract

In a frequency selective fading channel where symbol rate is higher than the coherence bandwidth, both fading and delay spread effects are the severe impairments that should be combated. In this work, spread spectrum (SS) and non-spread spectrum (NSS) techniques together with different modulation schemes are investigated to combat these impairments for the high-speed data transmission in mobile environments. Simulation results of the techniques proposed are presented and alternative design parameters are given.

1. Introduction

In a mobile radio environment the received signal arrives by several paths bounced from large objects such as buildings and local paths scattered from objects close to the receiver such as ground or trees. In these conditions multipath delay spread that limits performance of the system occurs. A channel exhibits intersymbol interference (ISI) due to multipath delay spread if the symbol rate of the transmitted signal is greater than the coherence bandwidth of the channel. This channel is said to be frequency-selective and suffers from ISI. In frequency-selective fading a portion of the band is affected by a deep null in the channel frequency response and the Signal-to-Noise Ratio (SNR) in that part of the band is degraded [4, 6, 7].

The effects of frequency-selective fading can be controlled by several means including multiple antennas, adaptive equalization and multirate modems. However, some of the techniques such as spread spectrum (SS) have their inherent resistance to these deeply affected channels.

In this paper two different schemes are considered to implement a physical layer of a high speed outdoor Wireless Local Area Network (WLAN). The first one uses Spread Spectrum (SS) techniques holding some necessary facilities like security and multiple access and a Rake receiver that optimally combines multipath components as part of the decision process. The second scheme is a Non-Spread Spectrum (NSS) technique which uses Quaternary Phase Shift Keying (QPSK) as the modulation technique, equalization and diversity combining techniques to combat the effects of fading and ISI.

The remainder of this paper is organized as follows. In Section 2, some general information about different SS techniques and performance of them over Rayleigh fading channel structure are given. Then simulation results for SS and NSS techniques are presented to combat frequency selective channel impairments in Section 4 and 5, respectively. In the last section, conclusions of this study are drawn.

2. Spread Spectrum

In recent years there has been a great interest in using spread spectrum techniques for commercial applications in addition to their use in military communications. New opportunities like wireless local area networks, personal communication networks and digital cellular radios have created the need for research on how spread spectrum systems can be optimised for the most efficient use in these

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

environments [1, 2, 3, 4, 5, 9].

The main characteristic of spread spectrum techniques is to spread the spectrum of the information signal to be transmitted using spreading codes. In Figure 1, a general model of the spread spectrum communication system is given.

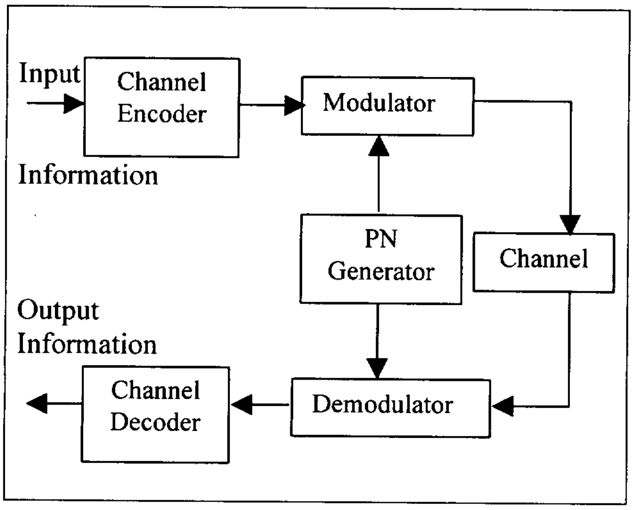


Figure 1. A general model of spread spectrum communication system.

Two different methods are used widely for spread spectrum systems:

- Direct sequence spread spectrum (DS-SS)
- Frequency hopping spread spectrum (FH-SS)

Time hopping, chirp and hybrid spread spectrum techniques are other methods for implementing spread spectrum systems. In the DS-SS technique the data bits are modulated second time with a PN-code generated by a maximal-length PN-code generator. The length of the code generator is changed from 10 to 100. The results of this technique for Rayleigh fading channel are shown in Figure 2. For the simulation, slow fading rate fading channel is considered, i.e. $B_fT_s = 0.1$ where B_f is fading bandwidth and T_s is symbol duration in the order of microseconds.

Three different modulation techniques are tested with DS-SS in a Rayleigh fading channels: BPSK, QPSK and DPSK. Performances of these modulation techniques in a Rayleigh fading channel without and with direct sequence spread spectrum are given in Figure 3 and 4, respectively. As power is increased, these 3 modulation techniques will give almost equal performances.

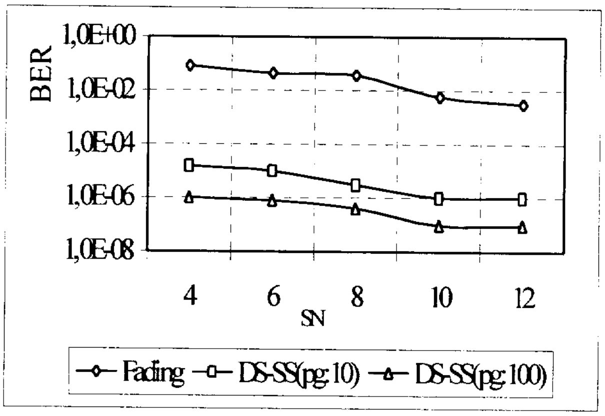


Figure 2. The performances of a direct sequence spread spectrum with processing gains of 10 and 100 in a Rayleigh fading channel with BPSK modulation.

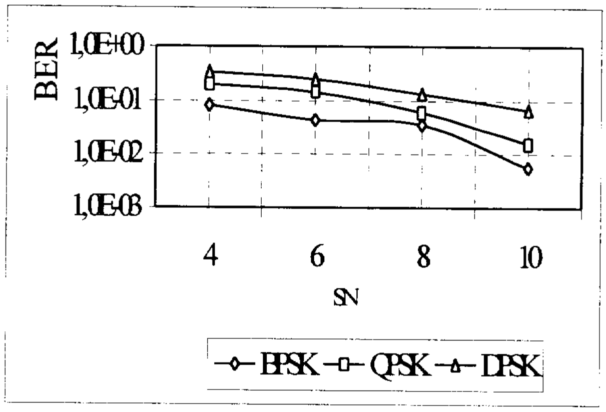


Figure 3. The probability of bit error for BPSK, QPSK and DPSK modulation techniques in Rayleigh fading channels.

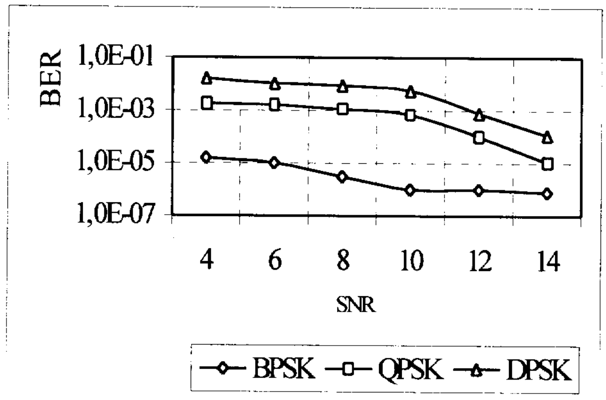


Figure 4. The performances of BPSK, QPSK and DPSK modulation techniques in Rayleigh fading channels with direct sequence spread spectrum.

In hybrid spread spectrum technique which is a combination of direct sequence and frequency hopping, a data bits sequence is divided over N_1 frequency hopping channels to obtain frequency hopping spread spectrum. Then a direct sequence spread spectrum is applied by a complete PN-code of length N_2 sequence in each frequency-hopped channel (see Figure 5). Using the fast frequency hopping scheme instead of the slow frequency hopping scheme causes an increase of the bandwidth. However this can be neglected regarding to the enormous bandwidth already in use.

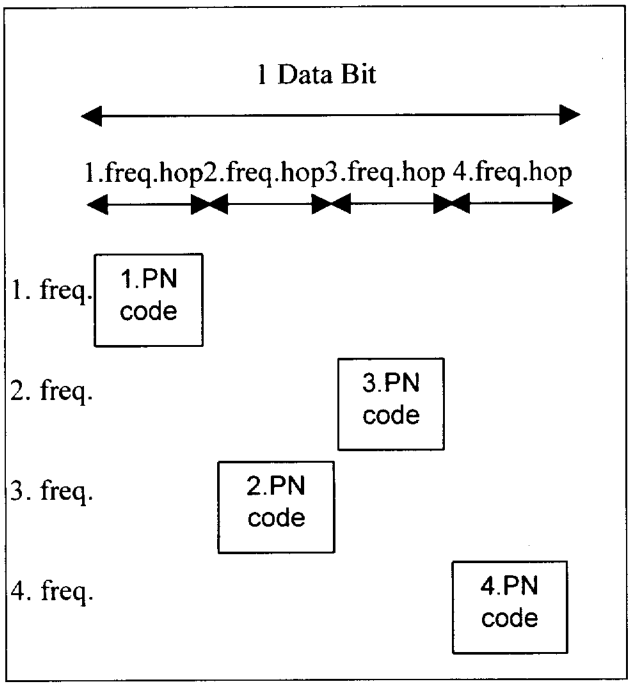


Figure 5. A hybrid direct sequence/frequency hopping spread spectrum model.

The number of frequency hops and the length of PN-codes are supposed to be short for simplicity. In applications, as this number and length are increased, system will provide a better performance to combat multipath fading, jamming, interference, etc. The results of a hybrid DS-SS/FH-SS system are given in Figure 6.

The performances of DS-SS and hybrid DS-SS/FHSS are almost the same in Rayleigh fading channels with BPSK modulation. But the hybrid model can be preferred with its advantages like better spreading property gained by frequency hopping and better multipath rejection property gained by direct sequence.

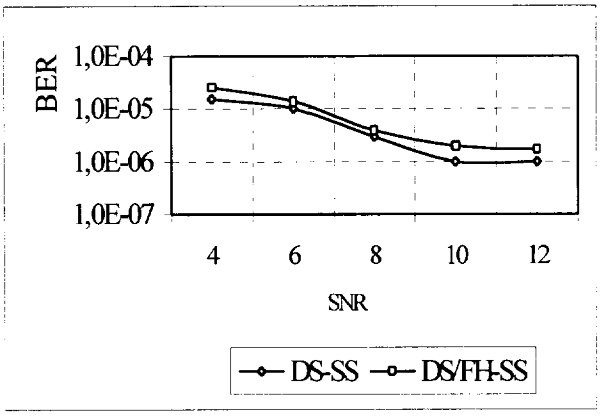


Figure 6. The performance of BPSK in a Rayleigh fading channel with DS-SS and hybrid DSSS/FH-SS.

3. Frequency Selective Channel

In most mobile radio communications, transmitted signals arrive at the receiver from various directions over a multiplicity of paths. Except for the LOS path, all paths are going through at least one order of reflection, transmission or diffraction before arriving at the receiver. In a multipath environment, the composite received signal is the sum of these signals [4, 6].

When the coherence bandwidth of the channel, $(\Delta f)_c \cong 1/T_m$, is greater than the bandwidth of the transmitted signal, the channel is frequency non-selective. In this case, intersymbol interference is prevented. On the other hand, when the coherence bandwidth of the channel is smaller than the signal bandwidth, the channel is said to be frequency selective and intersymbol interference occurs.

The impulse response of the channel must be determined correctly to remove this interference in the receiver. The complex envelope of this channel can be modelled as finite impulse response (FIR) filter by [6]:

$$h(t) = \sum_i \beta_i e^{-j\theta_i} \delta(t - \tau_i) \qquad (1)$$

where $\beta_i = P_i \alpha_i$, i is the path index, P_i is the amplitude of the ith path, α_i is Rayleigh or Rician fading envelope, θ_i is associated phase shift and τ_i is the delay. These parameters are randomly time varying functions due to the characteristics of the wireless channels. However, they can be considered as time-invariant random variables when they are compared with high signalling rates.

Some of the distributions used for modelling the channel are Rayleigh, Rician, Nakagami and Log-normal distributions. If the received signal is zero-mean due to the fading caused by randomly scattering objects in the environment, channel is modelled as Rayleigh fading channel. If there is also line-of-sight (LOS) or fixed objects in the environment, the channel may be modelled as Rician fading channel.

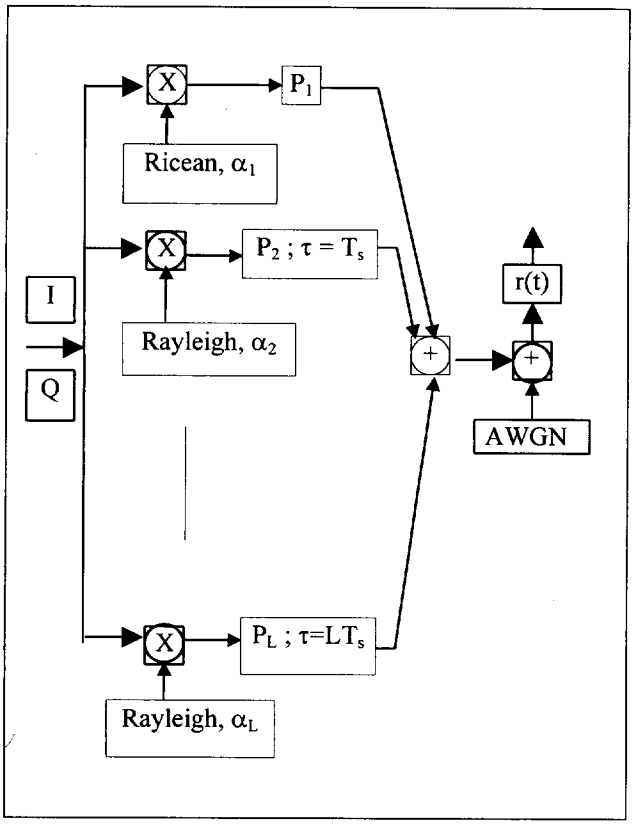


Figure 7. A frequency selective channel model with L paths.

Frequency selective fading channel in our case consists of L discrete multipaths in which the first path of the channel consists of a direct component (non-faded component) and several non-selective fading components with the delays approximately equal to the delay of the symbol duration T_s. Thus the fading on the first path is modeled as Rician fading. Each of the remaining paths consists of several fading components with small differential delays. Thus the remaining L-1 paths are modeled as Rayleigh fading. The maximum number of paths which depends on the multipath delay spread and symbol rate can be approximately by $L = T_m/T_s+1$.

As an example, with QPSK modulation for the data rate of R_b = 10 Mbps (the symbol rate $R_s = R_b/2 = 5$ Msym/s or the symbol duration $T_s = 0.2$ s) and with maximum delay spread $T_m = 0.7$ s (for the rural area), the number of paths, L can be taken as 4 [7].

For other types of modulation schemes such as 16-QAM and 64-QAM the same approach is used to find path numbers. For a general case, channel model is given in Figure 7.

4. Spread Spectrum Techniques to Combat Multipath Fading and Other Impairments of Frequency Selective Channels

The following analysis is done to see the effects of multipath fading on spread signal in mobile radio communication environment.

The transmitted baseband signal using direct sequence spread spectrum is given by:

$$R_{xs}(\tau) = \sum_n a_n R_{ff}(\tau - nT_b) \tag{2}$$

where a_n is the information digit.

$$R_{ff}(\tau) = \int_{-\infty}^{\infty} f(t) f(t+\tau) dt \tag{3}$$

is the autocorrelation function of the spreading signal.

The cross-correlation function of the transmitted signal and the periodic PN spreading signal is

$$R_{rs}(t) = \sum_n a_n \sum_{i=1}^{L} \beta_i R_{ff}(t - iT_b - \tau_i) e^{j\phi_i} \tag{4}$$

and is given in Figure 8.a.

The effects of the multipath on the receiver correlator output can be seen by comparing the figures in Figure 8.a and b. The first figure (a) represents a single-path channel whereas the second one (b) represents four-path multipath channel. The delay spread is smaller than the information bit interval T_n.

Therefore the coherence bandwidth of the channel is greater than the bandwidth of the transmitted signal and the channel is frequency non-selective. On the other hand, if the channel is frequency selective, intersymbol interference occurs as depicted in Section 3.

To overcome intersymbol interference, one solution is generating PN sequences with the chip intervals short enough. If we operate with a chip duration short enough to resolve individual paths, we can design a receiver to take advantage of the multiple paths to provide diversity and enhance the reliability of the decision on each received information symbol [10].

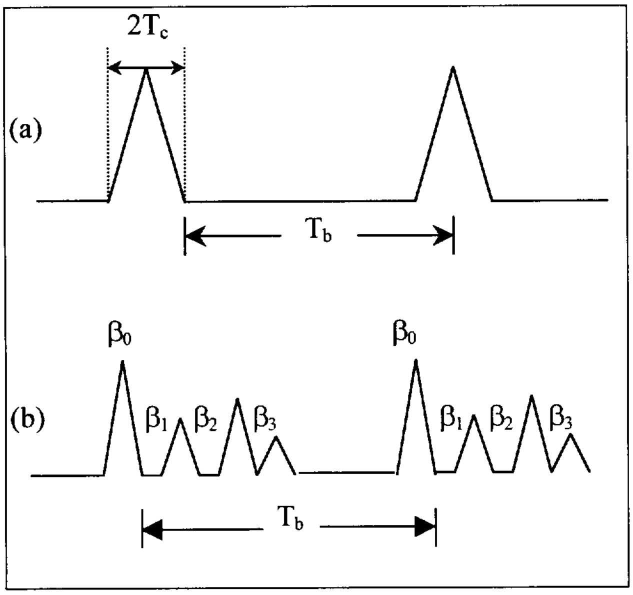


Figure 8. The cross correlation function of the transmitted direct sequence spread spectrum signal for (a) single-path channel, b) four-path multipath channel

Another step to overcome intersymbol interference is to use diversity combining techniques and Rake receivers with spread spectrum techniques. In a diversity system, the receiver obtains several copies of the same information signal through several different channels. It is the task of the receiver to utilize each received copy of the signal in its determination of the transmitted signal. Maximal-ratio, square-law and equal-gain diversity combining techniques are tested and maximal-ratio diversity combining is chosen for its optimum performance for coherent detection in this work [3, 8, 10].

Rake receiver employs a single delay line through which the received signal is passed. The signal on each tap is demodulated and combined in order to increase the equivalent signal to noise ratio. Rake receiver will have a tap for each chip in the pseudonoise sequence with each delay on the tapped delay line being equal to one chip time. Each independent path can be received by including a despreader on each tap.

Since most taps will contain only noise, it is important to eliminate these taps from contributing to the output. All information from the channel in our work can be received on 5 or 7 adjacent taps.

Various numbers of paths and taps are used in this work. The Rake receiver is used in AWGN channel and frequency selective/non-selective Rayleigh channels; in one-path, three-path and five-path channels with five-taps and direct-sequence spread spectrum. The results are given in Figure 9 and 10. As seen in figures lower degree of BER values can be reached by using Rake receivers.

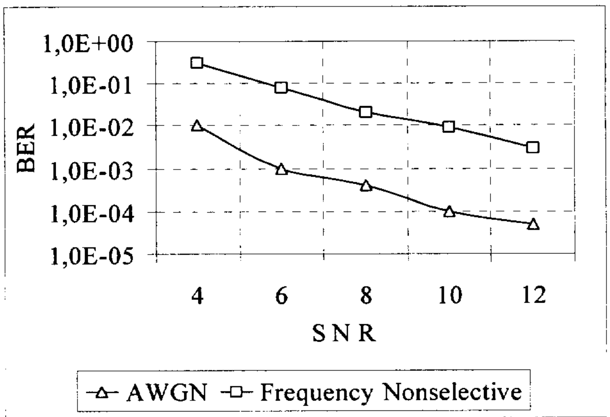


Figure 9. The performance of the Rake receiver in AWGN channel and frequency non-selective channel.

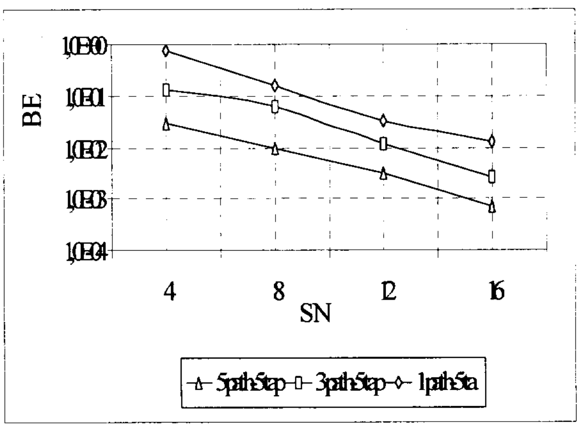


Figure 10. The performance of the five-tap Rake receiver in single-path, three-path and 5-path frequency selective channel.

5. Non-Spread Spectrum Techniques to Combat Multipath Fading and Other Impairments of Frequency Selective Channels

To combat the fading, delay spread and other effects in frequency selective channel non-spread spectrum technique (NSS) is also considered in this paper. Both diversity combining and equalization systems are used to remove these impairments. In a diversity system, several copies of the transmitted signal through several different channels are obtained in the receiver. Space diversity from several diversity techniques and maximal ratio combining from different types of diversity combining techniques are chosen as their best performances in this work [7].

Adaptive equalization techniques have been widely used in mobile radio channels to counteract the intersymbol interference produced by a form of delay spread. Since the frequency selective channel has severe amplitude distortion, nonlinear equalizers, which are less susceptible to the noise enhancement, should be used at the receiver. Decision feedback equalization (DFE) is the most suitable nonlinear equalization technique for this work since it does not require the knowledge of the statistical characteristics of the additive noise.

To improve BER performances over multipath fading channel, the combination of an equalization technique and a diversity combining technique should be used. This combination is performed over the wireless channel model given in Figure 7.

Different modulation techniques like BPSK, QPSK, 16-QAM and 64-QAM are investigated for this system. The performance of QPSK or 4-QAM in a frequency selective channel, which exhibits intersymbol interference, is better than the performance of 16-QAM or 64-QAM and the performance of 16-QAM is better than 64-QAM.

However, our simulation results show that 16-QAM and 64-QAM systems can also achieve low BER if appropriate techniques to combat the fading and intersymbol interference are used. Since these modulation schemes have high spectral efficiency, they can be used in a system that suffers from the bandwidth limitation. For instance, in the case where spread spectrum techniques have to be used in an application with a speed of at least 10 Mbps, 16-QAM and 64-QAM can give better results.

The BER performance for 4-QAM with two and four-branch diversities and equalization (DFE) over a mobile channel are given in Figure 11 and 12.

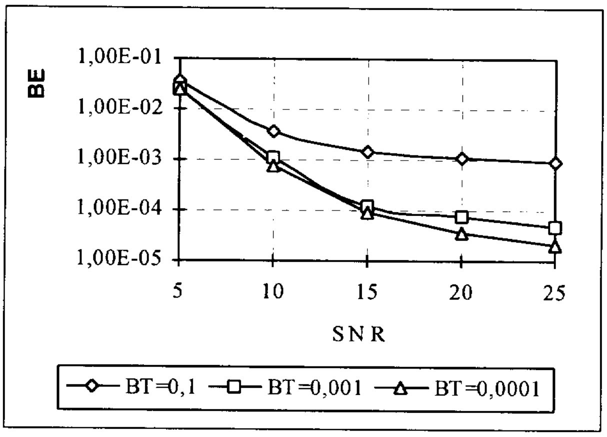


Figure 11. BER performance for 4-QAM with two branch diversity and DFE (2,3) over the frequency selective channel with different B_fT_s

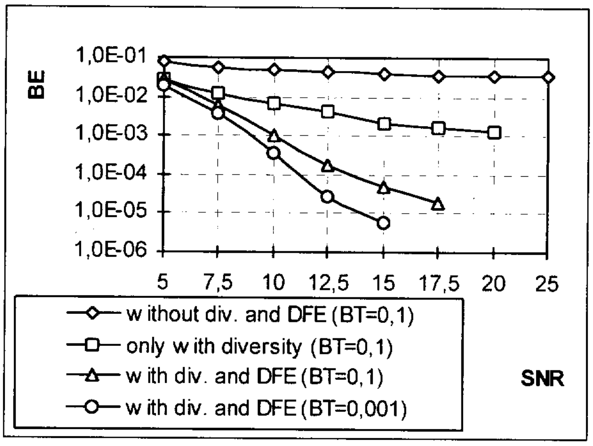


Figure 12. BER performance for 4-QAM with four branch diversity and DFE (2,3) over the frequency selective channel with different B_fT_s

6. Conclusion

Performance analysis of spread spectrum and non-spread spectrum in frequency selective fading is investigated.

Direct sequence spread spectrum techniques can be used in frequency selective channels with properly designed PN codes, Rake receivers and PSK modulation techniques.

The key parameter of designing PN code generator is to adjust the chip duration short enough to prevent the effects of delay spread. Using a Rake receiver with appropriate number of taps will also prevent these effect of multipath. Varios kinds of spread spectrum techniques can be used in fading channels. But a hybrid solution using both frequency hopping and direct sequence will provide some advantages like better spreading property and better multipath rejection property.

Effects of multipath fading and other impairments of frequency selective channels can also be removed or degraded by non-spread spectrum techniques with a combination of diversity and equalization techniques. Space diversity with maximum ratio combining is preferred for diversity combining technique and decision feedback equalizer is prefered for equalization technique in this work. 4-QAM gives the best results as the modulation scheme for non-spread spectrum systems.

Simulation results presented here shows that both spread spectrum and non-spread spectrum techniques can be considered to have a low BER for the high speed data transmission in a mobile communication environment

References

1. Bantz, David F. and Bauuchot, Frederic J.; *Wireless LAN Design Alternatives*, IEEE Network, March/April 1994.
2. Cooper, George R. and McGillem, Clare D.; *Modern Communications and Spread Spectrum*, McGraw-Hill, 1986.
3. Flikkema, Paul G.; *Spread Spectrum Techniques for Wireless Communication*, IEEE Signal Processing Magazine, May 1997.
4. Pahlavan, Kaveh and Levesque, Allen H.; *Wireless Information Networks*, John Wiley &Sons, 1995.
5. Pickholtz, Raymond L. and Schilling, Laurence B.; *Spread Spectrum for Mobile Communications*, IEEE Transactions on Vehicular Technology, Vol.40, No.2, May 1991.
6. Proakis, John G.; *Digital Communications*, McGraw-Hill, Third Edition, 1995.
7. Trabelsi, Chokri and Torun, Erdal; *Physical Layer Design Alternatives for Outdoor Wireless Local Area Networks*, Submitted to ICC, Communication Research Center, September 1995.
8. Turin, George L.; *The Effects of Multipath and Fading on the Performance of Direct-Sequence CDMA Systems*, IEEE Journal on Selected Areas in Communicaitons, Vol. SAC-2, No.4, July 1984.
9. Zorba, H.Orkun and Torun, Erdal; *The Use of Spread Spectrum Techniques in Mobile Radio Communication*, 6th IEEE Signal Processing and Application Conference, Kızılcahamam, May 1998.
10. Zorba, H.Orkun and Torun, Erdal; *Combating to Impairments of Frequency Selective Fading Channels by Using Spread Spectrum and Rake Receivers,* 7th IEEE Signal Processing and Application Conference, Ankara, June 1999.

High latitude off-great circle propagation effects with reference to HF communication systems and radiolocation

E.M. Warrington, T.B. Jones, N.C. Rogers[1] and A.J. Stocker
Department of Engineering
University of Leicester
Leicester, LE1 7RH
UK.

INTRODUCTION

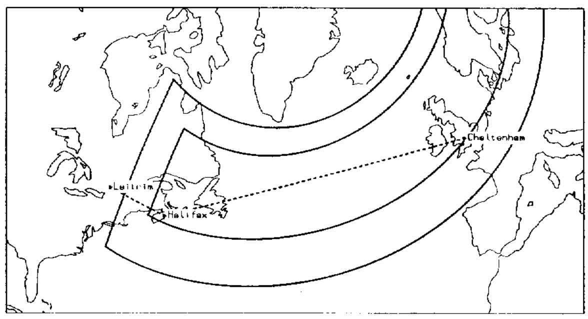
Owing to the presence of tilts and gradients in the high latitude electron density distribution, HF radio signals often arrive at the receiver over paths well displaced from the great circle direction. Deviations of a few degrees are associated with tilts due, for example, to the solar terminator and to travelling ionospheric disturbances (TIDs) (see, for example, Jones and Reynolds [1]). Very large deviations are particularly prevalent in the high latitude regions where signals often arrive at the receiver with bearings displaced from the great circle direction by up to ±100° or more. These large deviations from the great circle path are due to electron density depletion and the associated ionospheric tilts within the mid-latitude trough at sub-auroral latitudes, whereas in the polar cap they are attributed to the presence of convecting patches and arcs of enhanced electron density.

In addition to the large scale tilts which cause gross deviations of the signal from the great circle direction, irregularities in the electron density distribution may be considered as providing a rough reflecting surface for HF radio waves. As a result of this roughness, signals associated with each propagation mode arrive at the receiver over a range of angles in both azimuth and elevation. Accordingly, a single ionospheric mode is often modelled as a single ray specularly reflected from a smooth ionosphere surrounded by a cone of rays produced by the roughness of the ionosphere - the former is referred to as the specular component and the latter as the scattered or diffracted component (see Gething [2]). The hypothesis of a specular component of constant amplitude and constant direction of arrival surrounded by a cone of scattered energy is an over-simplification in practice since the model carries an implication of a smooth ionosphere of infinite extent upon which are superimposed the localised irregularities leading to the scattered energy. No account is taken of larger scale, possibly time varying, tilts in the model.

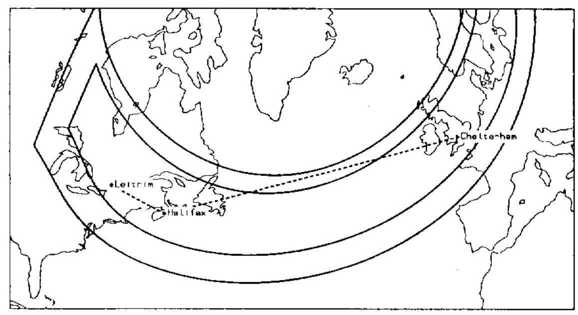
Various experimental measurements and their interpretation by researchers at the University of Leicester of off great-circle propagation over a range of high latitude paths are summarised in this paper.

PATHS ALONG THE SUB-AURORAL TROUGH

The trough is a region of depleted electron density in the night-time F-region ionosphere in which the critical frequencies drop by a factor of at least 2 and the altitude of the electron density peak rises by 100 km or more (see Moffett and Quegan [3]). During the winter and equinoctial months the trough takes the form of a band a few degrees wide in latitude on the equatorward edge of the auroral oval, stretching in local time from dusk to dawn. In summer the trough is much less pronounced and is confined to the hours around midnight. The trough has been modelled on a statistical basis and consequently it is possible to predict the periods during which, on average, propagation well displaced from the great circle direction is likely to occur for signals propagating through this region. Moreover, it is also possible to estimate the direction of the bearing deviations and for how long these disturbed conditions are likely to persist.



(a) $A_p = 0$



(b) $A_p = 48$

Figure 1. The Halifax to Cheltenham and Halifax to Leitrim paths in relation to the mid-latitude trough for (a) quiet and (b) moderate geomagnetic activity from the model of Halcrow and Nisbet [4] for a time of 0000 UT during the equinox months.

Presented in this paper are measurements obtained from two propagation paths whose locations are illustrated in Figure 1. The modelled extent the mid-latitude trough (Halcrow and Nisbet [4]) is also indicated for midnight

[1] N.C. Rogers is now with DERA, Malvern, UK.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

UT for geomagnetically quiet conditions ($A_p = 0$) in the upper frame of this figure and for geomagnetically active conditions ($A_p = 48$) in the lower frame on an equinox day. These figures illustrate the extension of the trough into earlier local times in the evening sector and also how the trough region moves equatorwards as geomagnetic activity increases.

Examples of measurements on various sub auroral propagation paths which clearly illustrate the features outlined above are presented in the following sections. The features have, wherever possible, been related to other parameters, such as the geomagnetic index A_p, in order to indicate how the possible presence of large deviations from the great circle path may be estimated.

Halifax, Canada to Cheltenham, UK (4490 km, 286° bearing)

VOACAP [5] predictions for this path indicate that the dominant mode is likely to be 2F on each of the three signal frequencies monitored (the exact modal structure will, of course, vary with time of day, season, etc.). Figure 2 illustrates the solar control of the three frequencies observed for this path. At night, only the 5.097 MHz signal is received as the maximum usable frequency (MUF) is too low to support the higher frequencies (10.945 MHz and 15.920 MHz). During the daytime, when solar ionisation occurs and the ionosphere is fully formed, the higher frequencies propagate due to an increased MUF but the lower frequency is not received as a consequence of the enhanced D-region absorption, i.e. an increase in the lowest usable frequency (LUF) (a good discussion of ionospheric radiowave propagation is given by Davies [6] and the reader is referred to this text for background information). Note the comparatively large bearing fluctuations on the 5.097 MHz signal during the night.

The occurrence and nature of propagation well displaced from the great circle path is well correlated with the geomagnetic activity index A_p. Typical examples of the bearings measured for three frequencies during March 1994 are reproduced in Figure 3. Large deviations (>10°) from the great circle path are a consistent feature on all of these frequencies. The effect of the magnetic disturbances is particularly noticeable on the 5.097 MHz signal. During the more disturbed days (i.e. when $A_p \geq 15$), a characteristic high-to-low bearing angle swing is evident, as for example on the 12 March 1994. On all of these days the bearings suddenly jump to about +30° relative to the true great circle bearing (GCB) at about 0000 UT. The bearing angle then slowly decreases as time progresses, passing through the GCB at about 1000 UT after which the signal is strongly attenuated due to the storm enhanced daytime absorption. However, when the signal returns there is a large negative offset which increases to about -25° before the sudden jump back to a positive value. Figure 4 illustrates the seasonal variation in the occurrence of this phenomenon. During the winter and equinoctial months (October to March) the number of days on which the characteristic bearing changes are observed (light bars) is closely correlated with the number of geomagnetically active ($A_p \geq 15$) days in that month (dark bars). In the summer months however, there is a disproportionately low occurrence of the characteristic bearing changes which is consistent with the well documented absence of the trough during this period. Note also that there is some evidence for a secondary minimum in the occurrence of the large deviations around the 1994 winter solstice.

A correlation with Doppler spread (see Warrington *et al* [7]) is also evident in these data. The increase in Doppler spread for the days when strong bearing swings are observed on the 5.097 MHz signal is illustrated in Figure 5 which indicates greatly increased Doppler spread on those days when large offsets from the great circle path occur.

A systematic dependence on A_p has also been observed on the 10.945 MHz and 15.920 MHz transmissions. For example, Figure 3 indicates that the time of occurrence of the change to negative bearings is correlated with A_p as is the time when the bearing reaches its minimum value. This dependence is clearly illustrated in Figure 6 in which the time of minimum bearing on each day has been plotted against the A_p value for that day. The trend to earlier UT for increasing A_p is very clear and is consistent with the expected extension of the sub-auroral trough into earlier local times in the evening sector during enhanced geomagnetic activity.

During periods of low geomagnetic activity (see, for example, Figure 3, 20-29 March 1994) propagation at 10.945 MHz often continues throughout the night at bearings equatorwards of the great circle path (GCP). It is also interesting to note that following the onset of a geomagnetic storm (e.g. see Figure 3, 7 March 1994), propagation during the night is often observed to the north of the GCP.

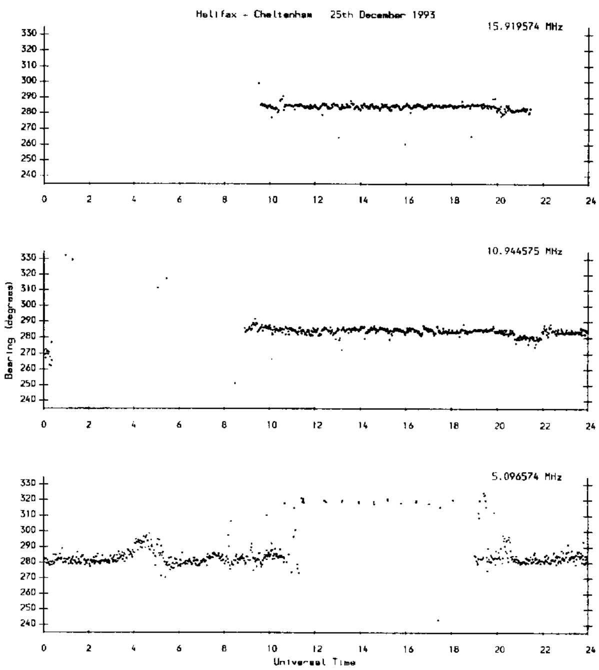
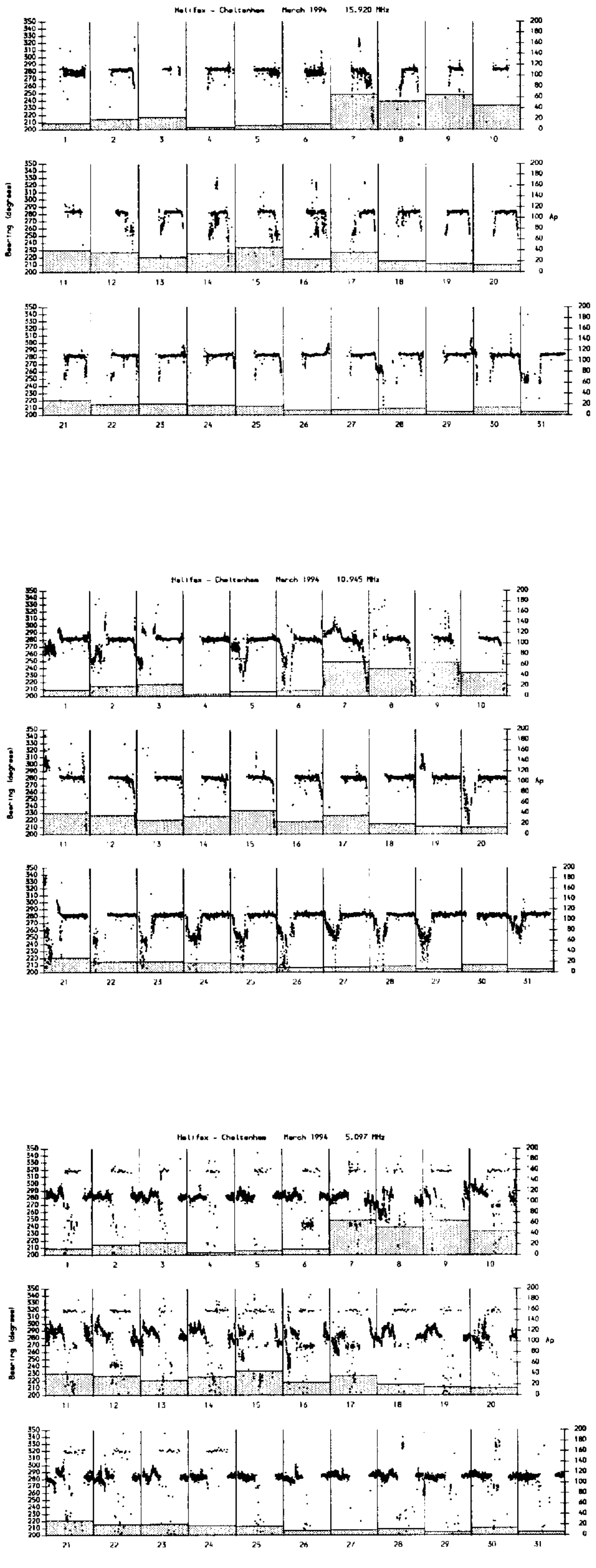


Figure 2. Solar control of the three frequencies transmitted from Halifax, Canada and received at Cheltenham, UK. Note the reduction in spread of the bearings with increasing frequency.



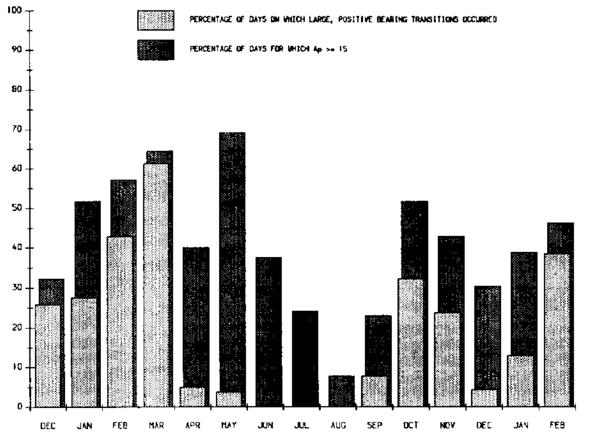


Figure 4. Percentage of days in which large northerly deviations occur for the Halifax - Cheltenham path. December 1993 - February 1995. 5.097 MHz.

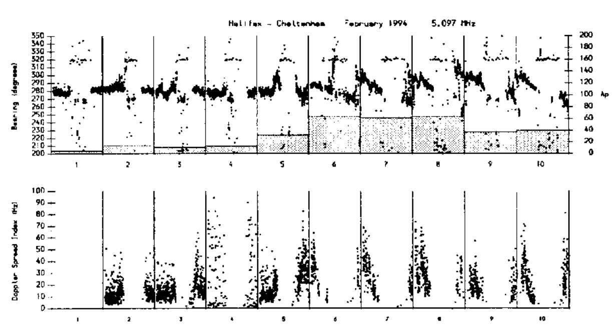


Figure 5. Bearings and Doppler spread index values recorded for the 5.097 MHz transmissions from Halifax to Cheltenham, February 1994. Note the increase in Doppler spread at times of large northerly bearing errors.

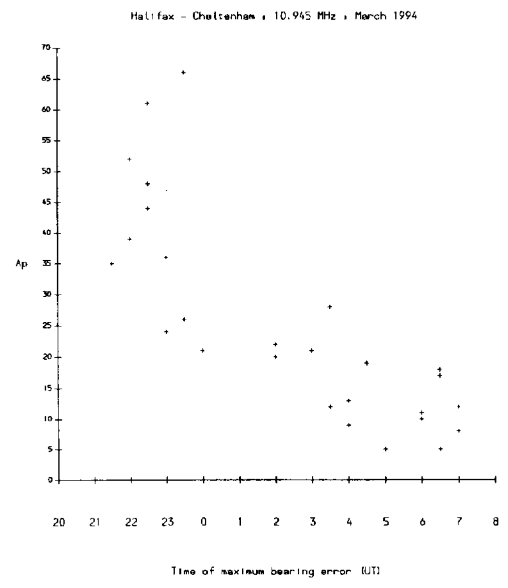


Figure 3. Bearings measured over the Halifax to Cheltenham path at 15.920 MHz (upper frame), 10.945 MHz (middle frame) and 5.097 MHz (lower frame) during March 1994 (the dates on these plots are for the UT day). Daily A_p values are also indicated.

Figure 6. Dependence of time of occurrence of maximum bearing error on the geomagnetic index A_p. Halifax to Cheltenham path, 10.945 MHz, March 1994.

Ray tracing simulation

Observations of the bearing of the 10.945 MHz signal during the period from 24 March 1994 at 12 UT to 25 March 1994 at 12 UT are presented in Figure 7. Prior to about 0 UT, propagation is predicted to be via a 2F mode and is on the great circle path. However, after 0 UT the bearing begins to turn southwards reaching a peak deviation from the great circle path (GCP) of around 40-50° by about 6 UT. After this time, the southward deviation in bearing reduced, returning to the GCP by 9 UT. The geomagnetic activity during this interval was relatively low (daily A_p was 18 and 17 for the 24 March and 25 March, respectively). Note that similar behaviour in bearing was also observed on a number of days immediately before and after this particular interval for which the geomagnetic activity was comparable (see the middle panel of Figure 3). The interval when the trough is present, i.e. from 21 UT (24 March) to 9 UT (25 March), has been simulated in a ray tracing model.

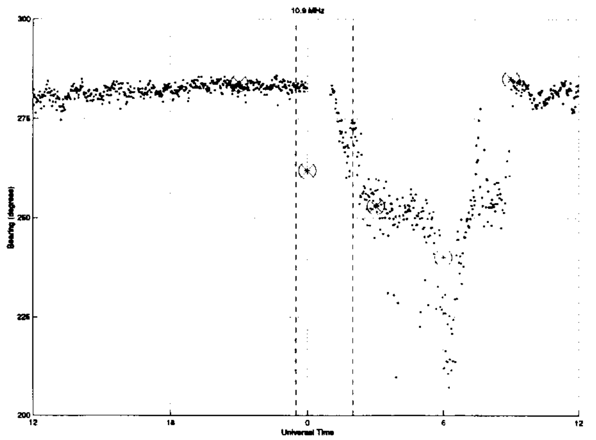


Figure 7. Observed and simulated bearings as a function of time between 24 March 1994, 12 UT and 25 March 1994, 12 UT. Note that there is a data gap just after 0 UT. The values obtained from the ray tracing simulation are given by the symbol ⊗. The vertical dashed lines indicate the earliest and latest times of trough opening for days of low geomagnetic activity at the end of March, 1994.

A numerical ray tracing code [8] was employed in this study with a background electron density model consisting of two Chapman layers, representing the E and F regions, with a gradient in electron density from the geographic equator. The key parameters of the electron density model (critical frequency, critical height and scale height of each layer) were based on values obtained from the International Reference Ionosphere (IRI) [9] for the selected interval. The location of the trough was calculated according to the model of Halcrow and Nisbet [4]. However, in the ray tracing code a simplified version of this trough model was implemented in which it is assumed that the trough walls lie at constant values of geomagnetic latitude (for a given UT and A_p) and the trough ends are at constant geomagnetic longitude (again for a given UT and A_p). The electron density in the main trough is assumed to be 70% below the ambient (i.e. troughless) value, while this perturbation decreases to zero over a few degrees of geomagnetic latitude once outside of the main trough.

At the beginning of the interval chosen for the simulation (21 UT) the trough covers about the last third of the GCP but the depleted electron density does not prevent propagation along on the great circle direction. By 0 UT, and similarly for 3 UT, the trough covers the entire length of the GCP and so prevents GCP propagation since the maximum usable frequency (MUF) within the trough is reduced to below the signal frequency. At 6 UT, although the trough closes just to the west of the receiver, the signals are still off great circle since the first ionospheric reflection of the 2F mode remains within the trough. At 9 UT, less than the first quarter of the GCP is covered by the trough and therefore the 2F mode reflection points are outside of the trough and the signals are received along the great circle path.

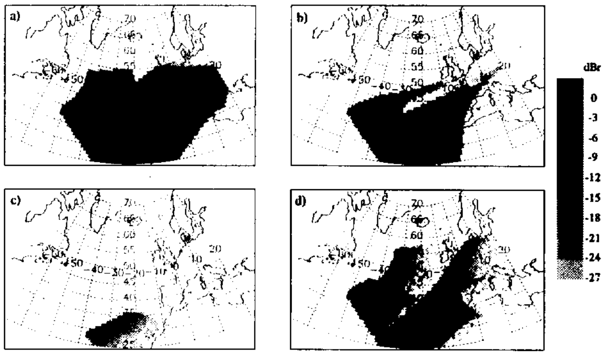


Figure 8. Map of relative power at ground level for the following conditions, a) 0 UT (no trough present), b), c), and d) 0 UT, 6 UT and 9 UT, respectively (with trough present).

In the simulation, rays are launched from Halifax at azimuths between 50° and 120° (for great circle path propagation the launch azimuth is 57°) and elevations between 4° and 22°. If the trough is absent, then the propagation between Halifax and Cheltenham is via a 2F mode. With the trough present, great circle propagation is no longer supported since the MUF within the trough is considerably reduced. However, there are directions in which propagation is supported to the south of the trough (the locations where rays are present at the earth's surface after both one and two hops are given in Figure 8). When calculating the second hop the ray tracing program assumes a specular reflection from the ground. However, since the sea is a rough surface then the signal is likely to be scattered in many directions. Therefore, in order to model this, the extreme positions of the regions of highest power (i.e. ray density) in Figure 8 have been used as the starting points for further ray tracing. The results of ray tracing from these regions of seascatter indicate that a signal would often be received at Cheltenham. The simulated direction of arrival at Cheltenham of these seascattered rays has been plotted as a function of time, together with the experimental observations, in Figure 7. Generally, there appears to be an excellent agreement between the observed bearings and those derived from the ray tracing simulation. The exception to this is at 0 UT, where the simulated bearing is about 20° lower than the observed unperturbed great circle direction. It should be noted, however, that the trough model employed [4] is based on the results of a statistical study and that there is some variation in the opening times of the trough under given conditions. Marked on Figure 7 are the earliest and latest times of opening of the trough inferred from the observed southward turning in bearing during the period 22 March 1994 to 28 March 1994 (where the A_p values are roughly similar). If the simulated opening time of the trough were to be delayed by about an hour then a good fit to the data would be achieved.

Further experimental observations remain to be investigated using the ray tracing simulation. In particular, account will be taken of the changes in trough position with geomagnetic activity.

Halifax to Leitrim, Canada (910 km, 90° bearing)

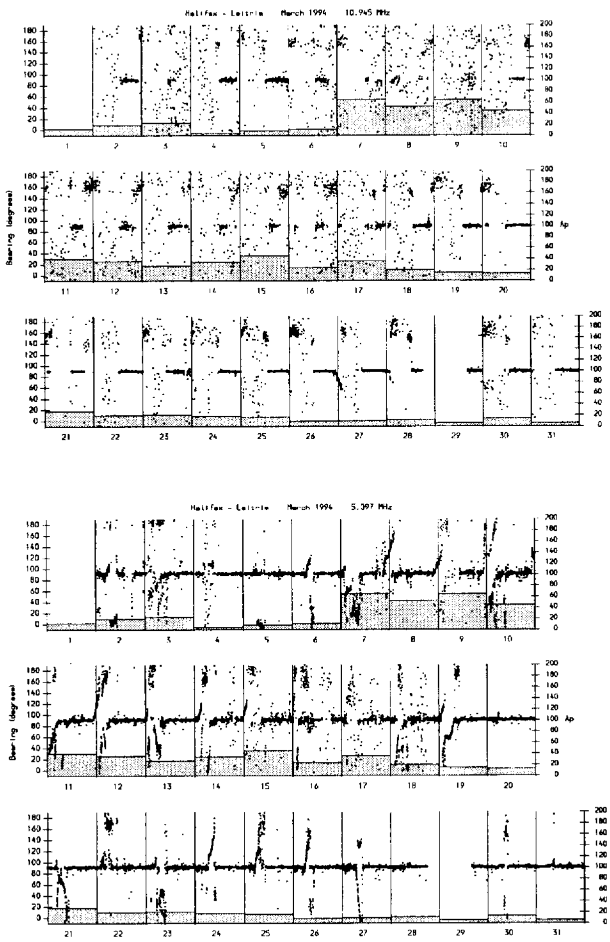


Figure 9. Large bearing deviations (~ 100°) observed at 10.945 MHz (upper frame) and 5.097 MHz (lower frame) on the Halifax to Leitrim path, March 1994. A_p values are also shown.

For the much shorter path from Halifax to Leitrim (near Ottawa), 1F propagation is expected to be the dominant mode based on VOACAP predictions. This path is at most times sub-auroral and strong effects due to the mid-latitude trough are expected. The bearings recorded for the 5.097 MHz and 10.945 MHz signals during March 1994 are reproduced in Figure 9 together with the corresponding A_p values. At 5.097 MHz, very large bearing deviations of up to ± 100° are a common feature. The period during which large bearing deviations occur usually begins about two or three hours after midnight UT. It is interesting to note that even on the geomagnetically quiet days the deviations can be either positive (e.g. 24th March), negative (e.g. 27th March) or include both directions (e.g. 19th March). During disturbed times, the period of large bearing deviations occurs earlier, starting at about 2200 UT. Large bearing swings are much less of a common feature on the 10.945 MHz transmission than at 5.097 MHz. The signal is usually lost at 0000 UT due to the low MUF and is not reacquired until about 1200 UT. During the disturbed days (7th to 12th March) only a limited bearing sample could be collected for a short period following 1200 UT (see Figure 9), presumably due to a storm-related reduction in critical frequencies throughout this region. During the quiet days, the most accurate bearings were obtained during the period 1200 to 2400 UT.

The strong dependence of the bearing error period on the A_p geomagnetic index is illustrated in Figure 10 in which the dependence of the start time of the large bearing errors on A_p is clearly evident. The entire period of occurrence of these large bearing errors (> ± 50°) can be related to the A_p value.

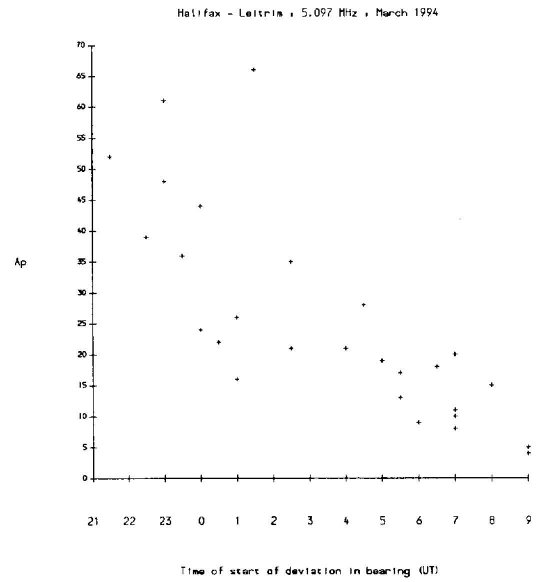


Figure 10. Dependence of the start time of the disturbed bearing periods on A_p. 5.097 MHz, Halifax to Leitrim, March 1994.

Discussion

Previous studies (Moffett and Quegan [3] and references therein) have determined that the trough forms at high geomagnetic latitudes near dusk and descends to mid-latitudes throughout the night before returning to higher latitudes near dawn. The results of these investigations show that during increased geomagnetic activity the propagation ceases at an earlier local time period on the Halifax - Leitrim (5.097 MHz) and the Halifax - Cheltenham (10.945 MHz) paths thus demonstrating the practical consequences of the proximity of the mid-latitude trough to the path reflection points. This emphasises the importance to HF predictions of catagorising the trough location as a function of geomagnetic index. The time at which the trough closes depends upon solar illumination re-forming the ionosphere and this does not vary with geomagnetic activity.

POLAR CAP PATHS

Observations over recent years have established that large scale electron density structures are a common feature of the polar cap F-region ionosphere. During periods of southward directed Interplanetary Magnetic Field (IMF) ($B_z < 0$) and the associated high levels of geomagnetic activity, patches of plasma 100-1000 km across with electron density enhancements of up to a factor of 10 above the background densities have been observed in the high latitude F-region ionosphere. These drift antisunwards across the central polar cap at velocities of a few kilometres per second in the high latitude convection current flows [10], [11]. When geomagnetic activity is low and the IMF is directed northward (approximately 50% of the time), Sun-Earth aligned arcs of plasma with electron density enhancements of a factor of 2-3 above the background can occur. These plasma striations are elongated for thousands of kilometres in the trans-polar noon-midnight direction but are much narrower (around 100 km) in the dawn-dusk direction. These features can persist for periods often in excess of one hour in the background F-region ionosphere [12] and have been found to be approximately twice as prevalent in the morning sector than in the evening sector [13]. They drift across the polar cap at velocities of a few hundred metres per second, generally in a duskward direction [11].

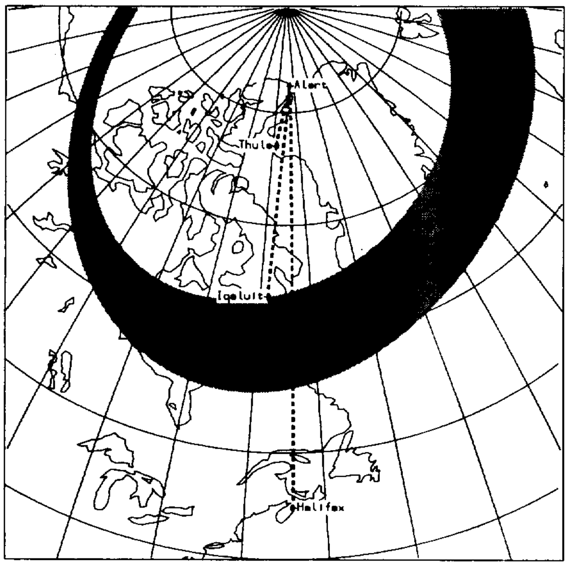


Figure 11. Location of the transmitting and receiver sites. A modelled position of the auroral zone at 0000 UT and A_p = 15 is also shown [14].

The electron density gradients associated with these large scale electron density structures form tilted reflection surfaces for HF radio waves which allow off great circle propagation paths to be established between the transmitter and the receiver. In order to investigate this type of propagation, a series of experiments have been undertaken in which the bearings and signal characteristics of a number of HF transmissions were measured by means of a wide aperture goniometric DF system located at Alert in the Canadian North West Territories. Large periodic bearing variations of up to ±100° from the great circle direction were observed. Measurements are presented in this paper for signals received from Thule, Greenland (670 km), from Halifax, Nova Scotia (4180 km) and from Iqaluit, North West Territories (2100 km). These paths are illustrated in Figure 11.

Halifax and Thule to Alert, November 1990

During the period 11th to 20th November 1990 bearing measurements were made at Alert for an 8.697 MHz signal from Halifax and on the 18th November for an 8.050 MHz signal from Thule. The bearings measured for the 8.697 MHz transmission are presented in the upper frame of Figure 12. The most striking feature of these data are the very large (±70°) bearing swings observed on the 16th, 17th and 18th November. This period is associated with a southward turning of the IMF on the 16th November measured by the IMP-8 satellite and a corresponding increased level of geomagnetic activity ($A_p > 10$) (see the lower frame of Figure 12). The bearing deviations are not random, but exhibit a periodic structure (see Figure 13 for the 18th November) and occur during the period 0500-1100 UT (0000-0600 LT).

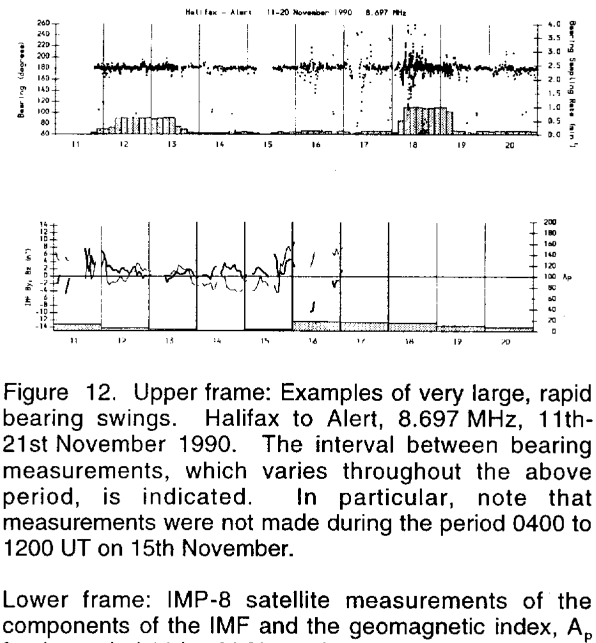


Figure 12. Upper frame: Examples of very large, rapid bearing swings. Halifax to Alert, 8.697 MHz, 11th-21st November 1990. The interval between bearing measurements, which varies throughout the above period, is indicated. In particular, note that measurements were not made during the period 0400 to 1200 UT on 15th November.

Lower frame: IMP-8 satellite measurements of the components of the IMF and the geomagnetic index, A_p for the period 11th - 20 November 1990. B_y is shown by the thin line and B_z by the thick line. IMF data are not available for the later part of this interval.

Measurements of the bearing of the signal received on 8.050 MHz over the short range (670 km) path from Thule on 18th November are presented in Figure 14. Very large periodic bearing variations are observed during the period 0530-1300 UT, which is approximately the same period during which the large periodic variations were observed on the path from Halifax. Much slower but equally large magnitude variations are present during the remainder of the day for this path. Only during the interval 1300-1600 UT was the measured bearing approximately along the great circle path (GCP).

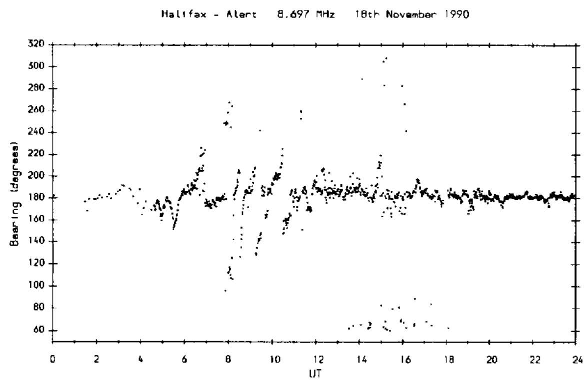


Figure 13. Examples of very large, rapid bearing swings. Halifax to Alert, 8.697 MHz, 18th November 1990.

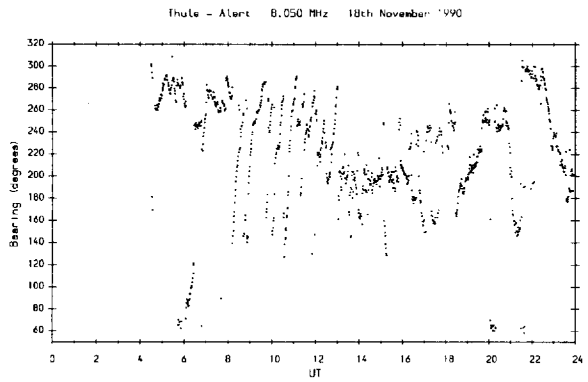


Figure 14. Examples of very large, rapid bearing swings. Thule to Alert, 8.050 MHz, 18th November 1990.

During the morning period for the data presented in Figure 14, the bearings swing from a low bearing angle (reflection point to the east of the GCP) through the true great circle position to a high bearing angle (reflection point to the west of the GCP) in both cases. The period between 1300 and 1600 UT is relatively undisturbed but after 1600 UT large bearing swings are again observed. At this time, in contrast to earlier in the day, the bearing changes from a high bearing angle, through the great circle position to a low bearing angle as time progresses. Note that the number of bearing swings is fewer in the post-noon period than in the pre-noon period.

Measurements of the critical frequencies throughout this period have been obtained from the Digisonde operated by the Phillips Laboratory, Hanscom AFB at Thule [15]. These measurements are reproduced in Figure 15 for the 17th, 18th and 19th November 1990. There is a very clear difference between the F-region behaviour on the geomagnetically active and quiet days. The measurements of foF2 indicate that large enhancements in electron density occurred on the geomagnetically active days of 17th and 18th November lasting for periods of about 30 minutes whereas, in contrast, no such changes were present on the 19th November, a geomagnetically quiet day. Comparison with the bearing measurements indicates that large bearing swings are only present on those days when the ionosonde records large transient fluctuations in foF2. Propagation is always close to the great circle direction during the geomagnetically quiet days when there are no disturbances in the ionosonde records. Detailed comparison between the times of occurrence of the foF2 peaks and the bearing fluctuations do not indicate any correlation between individual events, probably because the ionosonde is not located close to the propagation path mid point.

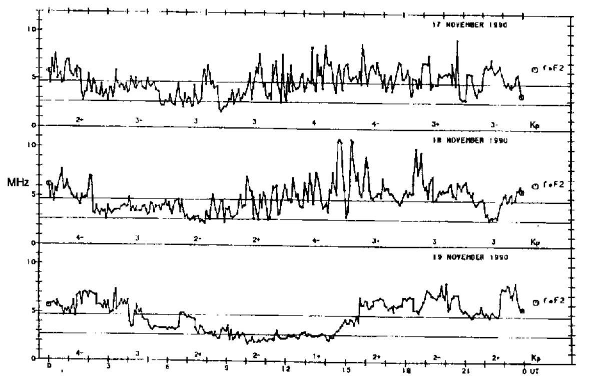


Figure 15. Changes in F-region critical frequency observed at Thule during periods of blob activity (17, 18 November 1990) when large bearing errors were observed at Alert. An undisturbed day (19 November 1990) is included for comparison. The approximate equivalent vertical frequencies for the Thule path (4.7 MHz) and the Halifax (2.7 MHz) are indicated by the horizontal lines.

Iqaluit to Alert, 1994

The 2100 km path from Iqaluit to Alert is always contained within the polar cap ionosphere. In winter, the path remains in darkness for long periods and the reverse is true in summer. There is, consequently, a marked seasonal dependence of the signal behaviour on this path, with large (~100°) bearing deviations observed during winter and equinoctial months (see, for example, Figure 16) and only small (<10°) fluctuations for most of the time during the summer months.

Figure 16 shows typical examples of winter measurements for 5.832 MHz and 9.292 MHz signals. A very large spread in measured bearings is present on both frequencies, even during geomagnetically quiet periods. The night time period 0000-1000 UT is characterised by very large bearing swings, sometimes in excess of 100°. The effect of a magnetic storm during the period 5th to 12th February 1994 is clearly evident. During this period, large positive bearing errors are observed on the 5.832 MHz signal during the night and the 9.292 MHz transmission is only received consistently after about 1200 UT when very large and random bearing errors are observed.

There is an underlying diurnal trend for propagation to deviate to the west of the GCP (high bearing angles) in the evening sector (local midnight at the GCP mid-point is 0430 UT) with propagation returning from the east of the GCP (low bearing angles) in the morning. This may arise from very large scale ionospheric gradients in the polar cap associated with the solar terminator. More rapid bearing swings with periods of about 30 minutes are often superimposed upon these trends and are thought to result from the presence of convecting patches of enhanced ionisation.

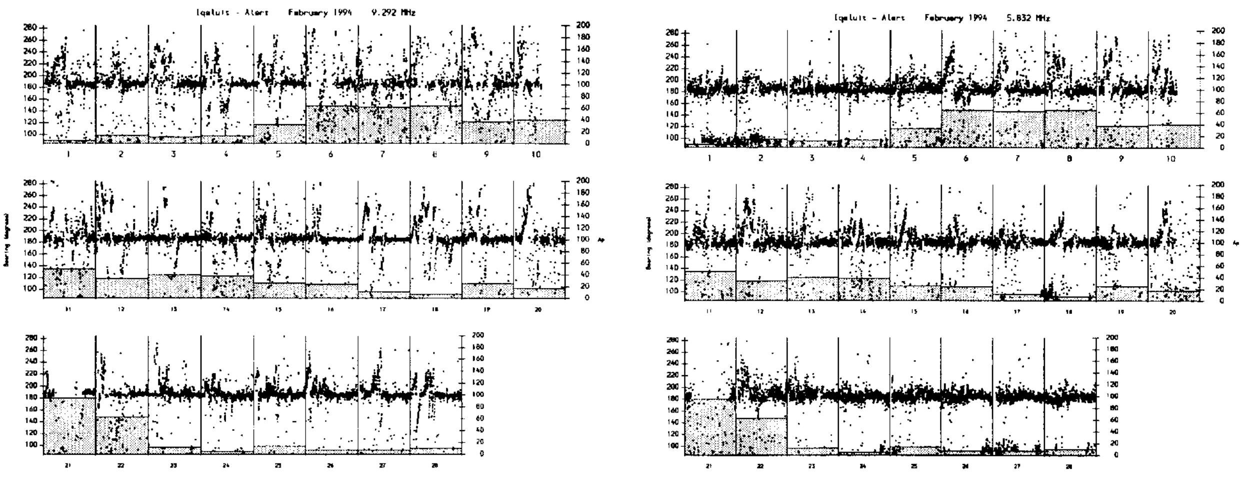


Figure 16. Bearing measurements for the 9.292 MHz (left hand frame) and 5.832 MHz (right hand frame) transmissions from Iqaluit received at Alert, February 1994. A_p values are also shown. Note that there are no data for the latter part of 10th February.

A further geomagnetic storm is evident on 21st-22nd February and examples of the rapid bearing swings observed at 9.292 MHz for the period 21st-24th February 1994 are presented in Figure 17 together with values of the 3-hourly a_p index and the B_y and B_z IMF parameters. The principal bearing swings on the night of the 21st-22nd - a period of southward IMF and high a_p values - have a decreasing bearing angle and occur in the six hour period before local midnight (0430 UT) whereas the principal bearing swings on the following night - a period of northward IMF and low a_p - have an increasing bearing and occur principally in the hours following local midnight.

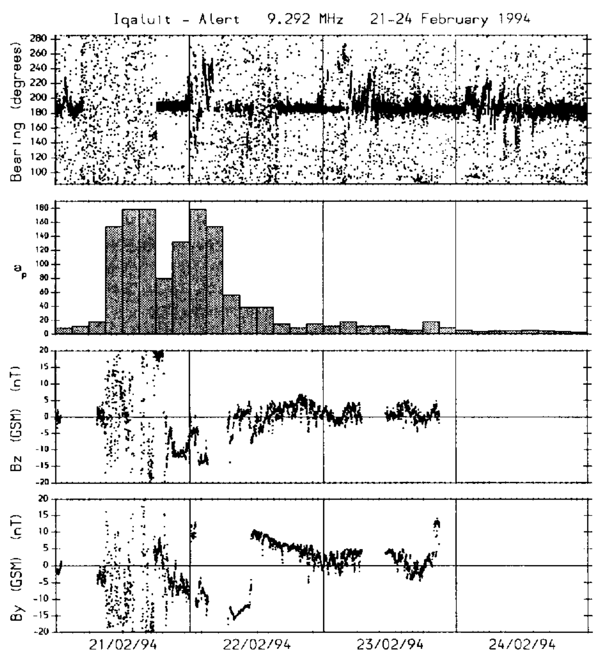


Figure 17. Bearings measurements for the 9.292 MHz transmission from Iqaluit received at Alert for the period 21st-24th February 1994. Three-hourly a_p values and the IMF B_y and B_z values are also shown (middle and bottom panels).

In order to correlate the occurrence of bearing swings with geomagnetic activity, the start and end times and the maximum and minimum bearings of recognisable swings in the 9.292 MHz bearing measurements were manually scaled and recorded for the period 25th January - 30th April 1994. The average number of bearing swings recorded for each local time (LT) hour during this period are presented in Figure 18 after categorising data by the current 3-hourly a_p value. The decreasing bearing angle swings are found to occur predominantly during periods of high a_p and preferentially during the hours before local midnight (0430 UT) whereas the increasing bearing angle swings peak in the post-midnight hours (at about 0200 LT) during periods of low a_p.

Discussion

One of the interesting features of the large bearing swings observed on the Thule-Alert path is that the direction of rotation of the bearing angle is different in the pre- and post-noon periods. The bearing angle increases in the period between local midnight and local midday and decreases in the period between local noon and local midnight. At around local noon there is little change in bearing which is near to its great circle value.

Sketches of the polar cap convection flow are reproduced in Figure 19 (reproduced from Lockwood [16]) for six orientations of the IMF. On the 18th November 1990 the IMF was predominantly southward ($B_z < 0$) and duskward ($B_y > 0$); under these conditions there is a strong two-cell convection pattern with flows over the North magnetic pole directed towards about 2100 CGT (see the upper right hand sketch of Figure 19). This pattern is fixed in space relative to the sun and the earth rotates beneath these flows. For the Thule to Alert path (mid-point is at 87.1° N CGLat), in the morning the flow is almost orthogonal to the path (T-R) as indicated in the upper frame of Figure 20. Patches of enhanced ionosation detached from the dayside ionosphere (~1200 LT) will follow the direction of the convection flow over the polar cap as indicated by the arrow. Reflections from a patch will thus be received first when it is in position (1) and the measured bearing will be smaller than the great circle value. As the patch moves with the flow, the bearing angle will increase to its true great circle position and continue to increase as the patch travels away to position (2).

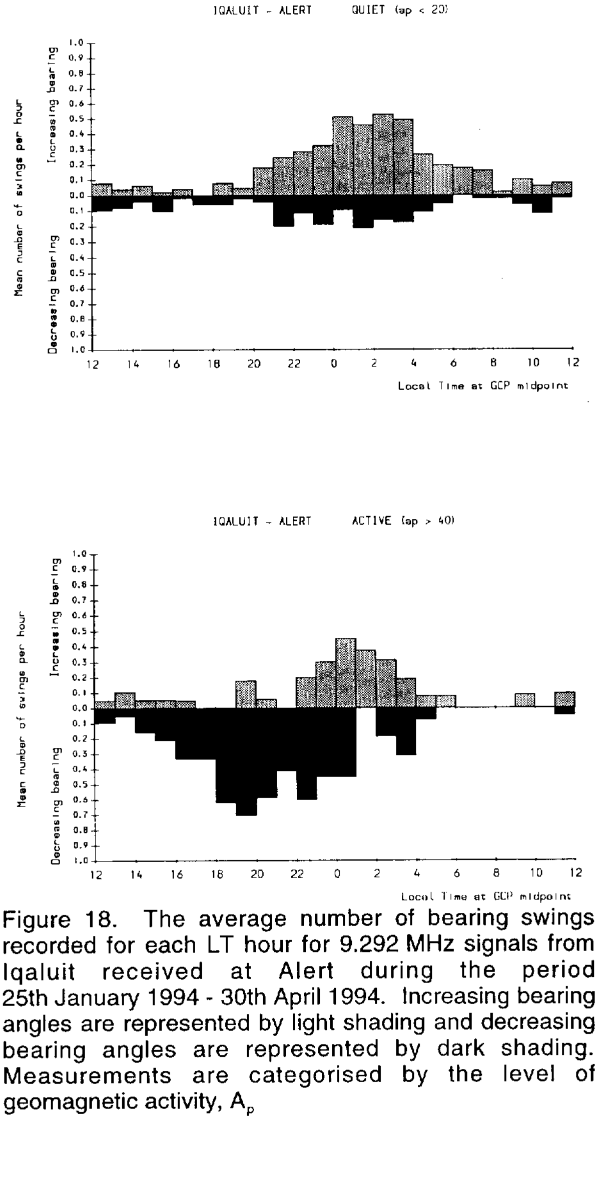


Figure 18. The average number of bearing swings recorded for each LT hour for 9.292 MHz signals from Iqaluit received at Alert during the period 25th January 1994 - 30th April 1994. Increasing bearing angles are represented by light shading and decreasing bearing angles are represented by dark shading. Measurements are categorised by the level of geomagnetic activity, A_p

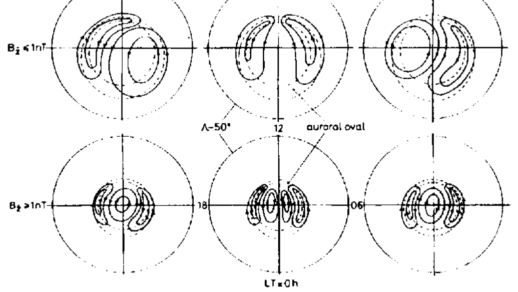


Figure 19. Model of the cross polar cap convection flow (after Lockwood [16]).

Near noon, (see the middle frame of Figure 20) the flow is approximately parallel to the propagation path, consequently the presence of patches of enhanced ionisation is unlikely to produce large swings in the measured bearing. In the post noon period, the flow is again almost perpendicular to the path but this time the earth has rotated so that the path is in the position shown in the lower frame of Figure 20. When reflections from a patch are first obtained (position (1)) the bearing angle will be larger than the great circle value. The bearing angle will decrease as the patch moves with the convection flow, the smallest bearing angle occurring at position (2), when reflection from the patch ceases.

The change in position of the propagation path relative to the ionospheric convection flow as the earth rotates can account for the change in the direction of rotation of the measured bearings between the pre and post noon periods. The number of bearing swings will depend on the number and spatial structure of the patches present which can provide reflection points between the transmitter and receiver. The period of the bearing swing will be related to the velocity of the patch relative to the transmission path. This will depend on several factors including the velocity of the flow which, in turn, depends on the magnitude and direction of the IMF and the cross-cap potential. It is interesting to note that in the morning the path is moving towards the flow (as a consequence of the earth's rotation beneath the convection pattern) whereas in the afternoon the path is moving with the flow. This difference in relative motions may account, to some extent, for the difference in swing rate between the morning and afternoon periods on the Thule-Alert path.

The equivalent vertical frequencies are about 4.7 MHz for the signal from Thule and about 2.7 MHz for the signal from Halifax. These frequencies have been indicated on the plots of foF2 measured at Thule and presented in Figure 15. It is interesting to note that the large bearing swings were observed on the Halifax signal during the period 0500-1100 UT when the signal frequency was close to the path MUF but not during the afternoon period when the MUF was well above the signal frequency. In contrast, bearing swings were observed on the Thule signal during both the morning and afternoon periods for which the signal frequency was close to the path MUF throughout the day. This demonstrates the well known increase in sensitivity of HF propagation characteristics to electron density changes for frequencies close to the MUF (or to the critical frequency at vertical incidence).

Similar swings in bearing were observed on the Iqaluit - Alert path during 1994. For these measurements, both positive-going and negative-going swings were observed dependent upon the level of geomagnetic activity (see Figure 18) and upon the direction of the IMF. The underlying offsets observed in the bearing measurements (for example, see the data presented in Figure 16 for 20th February 1994, 0400-1000 UT) are most probably a consequence of very large scale (>1000 km) ionospheric gradients across the polar cap. Reference to the Parameterized Ionospheric Model (PIM) [17] indicates that the critical frequency at the latitude of the path mid-point usually ranges from less than 2.0 MHz near local midnight to over 4.5 MHz near local noon during days in February and March 1994. The equivalent vertical frequency of the 9.292 MHz signal for the Iqaluit to Alert path is about 2.9 MHz. This implies that in the hours around local midnight the 9.292 MHz signals will not propagate along the GCP and must therefore be reflected from regions of enhanced electron density to the side of the path in the direction of the solar terminator. It is for this reason that the underlying trend is for signals to deviate to the west of the GCP (higher bearing angles) before local midnight and return from the east of the GCP (lower bearing angles) after local midnight. In the hours around local midnight localised regions of electron density enhancements may drift across the path and these would be observed as swings in the direction of arrival of the signals, the direction of swing dependent upon the drift direction relative to the GCP.

When the IMF is directed southward ($B_z < 0$), patches of ionisation drifting anti-sunwards would lead to a preponderance of decreasing bearing angle swings in the pre-midnight hours and increasing bearing angle swings in the hours after midnight. Figure 12 shows this to be the case where $B_y < 0$, although where $B_y > 0$ there is a distinct lack of decreasing swings in the pre-midnight sector. Reference to the convection flow diagrams of Figure 19 indicates that for these conditions the flow direction is skewed towards the evening sector and patches of solar produced ionisation will drift across the dawn side of the polar cap, thus producing the likelihood of a decreasing bearing angle swing.

When the IMF is directed northwards ($B_z > 0$), the principal large scale electron density structures within the polar cap ionosphere are sun-aligned arcs. Buchau *et al* [11] have reported F-region ionosonde returns in excess of 10 MHz from such arcs and conclude that the reflections are specular in nature. Various studies (e.g. [11] and [13]) have established that the polar cap ionosphere is a region of highly disordered flows during periods of northward IMF. Arcs drift across the polar cap at up to 250 m s^{-1}, usually in a duskward direction, although they have been observed to remain stationary or drift dawnward. They can also break up and blobs of plasma may detach from the arcs and move in the convection flows. A series of arcs drifting steadily across the polar cap from dawn to dusk would lead to the expectation that increasing bearing swings would be observed during the time sector 1800 to 0600 LT, with the largest swings expected in the midnight sector. Decreasing bearing swings would be observed in the local time sector 0600 to 1800 with the largest swings in the noon sector (see Figure 22). The preponderance of swings with increasing bearing in the 1800-0600 LT sector for conditions of positive B_z (see Figure 21, lower panels) tends to support this theory. Figure 21 indicates that the distribution of increasing bearing angle swings is a maximum in the post-midnight hours at about 0200 LT. This may be explained by the greater abundance of arcs on the morning side and by the fact that the path is offset to the east of the magnetic pole. The differences in the distributions for $B_y > 0$ and $B_y < 0$ might be explained by the differences in the flow patterns at the mid-point latitude. Where $B_y > 0$ the flow is predominantly westward throughout the night and would lead to increases in bearing angles whereas when $B_y < 0$ strong eastward flows exist in the pre-midnight sector which would lead to decreasing bearing angles.

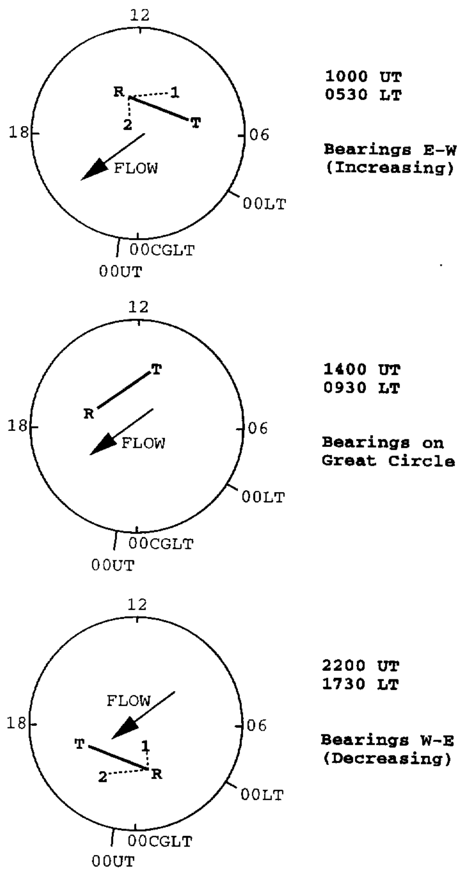


Figure 20. Schematic representation of the Thule to Alert path during the morning, noon and afternoon periods relative to the convection flow. T and R indicate the locations of the transmitter (Thule) and receiver (Alert) respectively. The bearing to the reflection point when the signal is first acquired is indicated by the dashed line to the point labelled 1 and the bearing to the reflection point when the signal is last heard indicated by the dashed line to the point labelled 2.

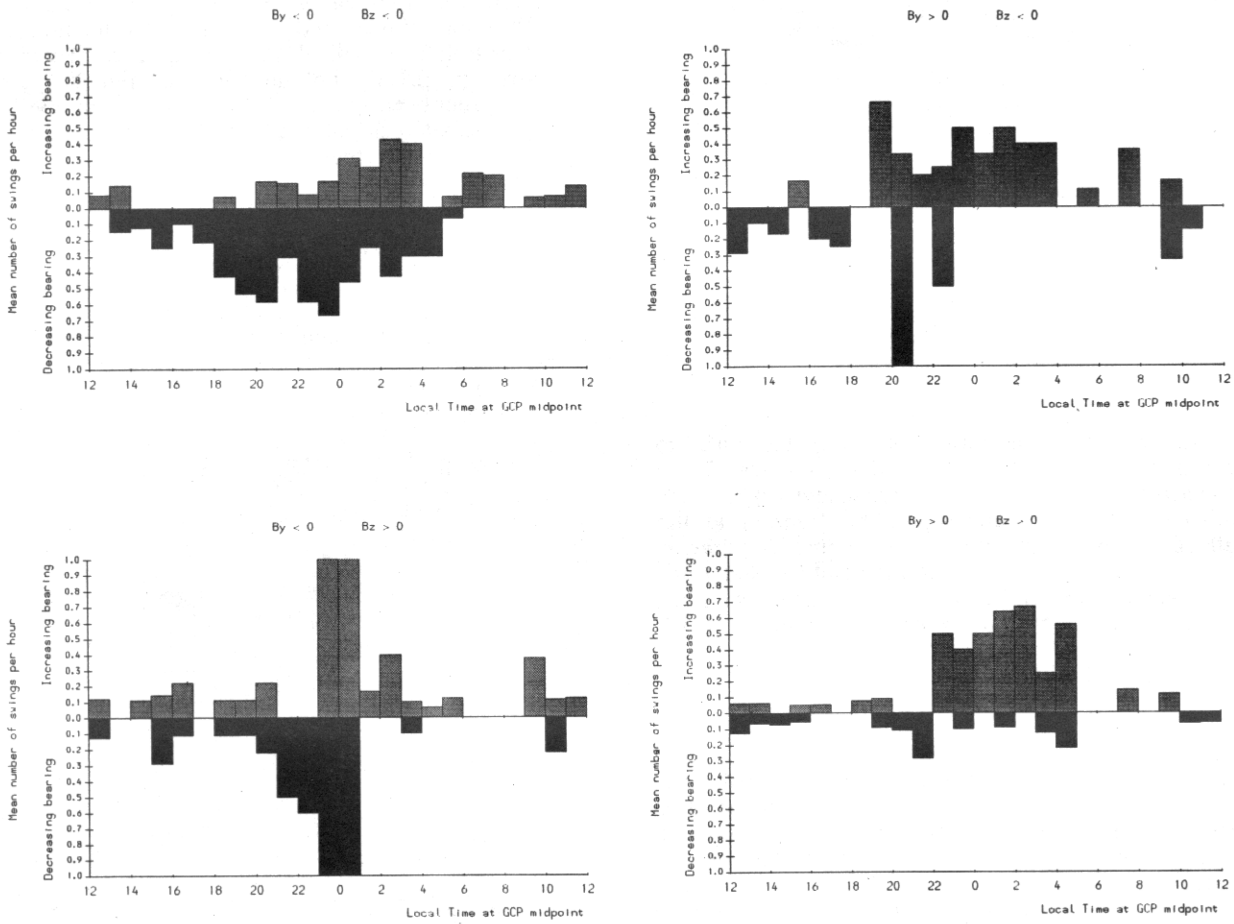


Figure 21. The average number of bearing swings recorded for each LT hour for 9.292 MHz signals from Iqaluit received at Alert during the period 25th January 1994 - 30th April 1994. Increasing bearing angles are represented by light shading and decreasing bearing angles are represented by dark shading. Measurements are categorised by the B_y and B_z IMF directions.

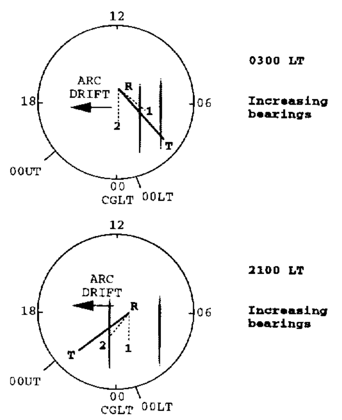


Figure 22. Schematic representation of the Iqaluit to Alert path relative to the drift of sun-aligned arcs. T and R indicate the locations of the transmitter (Iqaluit) and receiver (Alert) respectively. The bearing to the reflection point when the signal is first acquired is indicated by the dashed line to the point labelled 1 and the bearing to the reflection point when the signal is last heard indicated by the dashed line to the point labelled 2.

DIFFUSE REFLECTIONS

The directional spread in the received signal energy is an important parameter to be considered in the design of multi-element receiving arrays and the associated signal processing methods used, for example, in direction finding or adaptive reception systems. In this paper, observations made with two super-resolution DF systems of the direction of arrival of a narrow band pulsed channel sounding signal received over two high latitude paths are reported. The signals were processed in such a way that the directional characteristics of the received signals were determined as a function the relative times of flight of the various propagation modes and of Doppler frequency.

Experimental configuration

The signals employed for these experiments were radiated by a DAMSON (Doppler And Multipath SOunding Network) transmitter (see Davies and Cannon [18]). This system characterises the propagation path using a number of sounding waveforms which can be flexibly scheduled. Reported in this paper are measurements of the delay-Doppler waveform which was designed to measure the channel scattering function (frequency / arrival time dispersion). For this waveform, the radiated signals comprise sequences of 13-bit Barker coded BPSK pulses modulated at 2400 baud with a repetition rate of 80 coded pulses per second. The codes were sent 128 times per data analysis interval giving an integration time of 1.6 seconds and a frequency resolution of 0.63 Hz.

Measurements are presented for signals received over two paths:

1. A 3073 km high latitude path from Isfjord, Svalbard to Cricklade, UK. This path is illustrated in Figure 23 together with the average statistical positions of the auroral oval (Holzworth and Meng [14]) for a time of 15:00 UT with moderate (Kp = 3) geomagnetic activity. Also shown in this figure are the outermost beams of the CUTLASS HF radar (Jones and Thomas [19]) located at Hankasalmi in Finland, the viewing area of which covers the first third of the Isfjord to Cricklade path. The true bearing of Isfjord from Cricklade is approximately 7°, although it should be noted that there is a small uncertainty in the precise orientation of the antenna array.

2. A 1383 km polar cap path from Isfjord to Alert in the Canadian North West Territories. This path is also illustrated in Figure 23 and the great circle bearing of the transmitter from the receiver is 68.3°. This path is also within the field of view of the CUTLASS radar at Hankasalmi, although in this case the mid-point of the path is approximately 2500 km from the radar site.

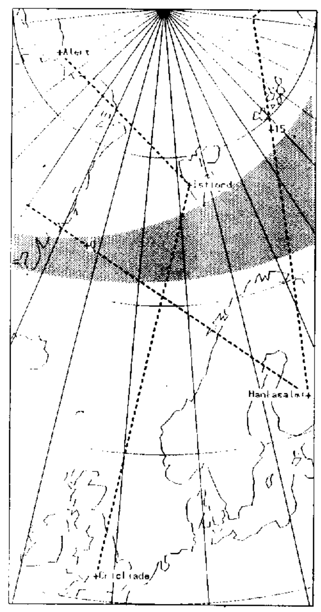


Figure 23. Map showing the path from Isfjord to Cricklade together with the mean position of the auroral oval at 15:00 UT for Kp = 3. Also shown are the two outside beams from the Finland CUTLASS radar and the path from Isfjord to Alert.

Owing to the long range of the paths and the relatively narrow bandwidth of the sounding signal, resolution of multiple propagation modes was not always possible since the despread pulse width was comparable with the difference in the times of flight of the various propagation modes. Multiple modes therefore usually resulted in pulse broadening rather than the presence of several discrete peaks. Consequently, where possible, particular attention is given to those periods where the propagation appears to be single moded (i.e. to periods with no significant pulse broadening). It should be noted that the DF receivers were not time-synchronised with the DAMSON transmitter and hence whilst it was possible to measure relative propagation delays, it was not possible to measure absolute times of flight.

The signals were received with large sampled aperture antenna arrays, each element of which was connected to a separate receiver. The complex amplitudes of the signals received on each antenna within the array were sampled simultaneously many times per second and the data processed to provide a measure of the relative times of flight of the propagating modes and their associated Doppler spectra (see Davies and Cannon [18]). In this way, the signal was split into components distinguished by time of flight, Doppler frequency and by antenna position in the receiving array. A direction finding algorithm (a modified version of an iterative null steering super-resolution direction finding algorithm [20] based upon the IMP [21] and DOSE [22] algorithms) was then applied to each signal component in turn in order to estimate the directional characteristics of the received signal. Note that when several signal components closely separated in direction of arrival are present in a single delay-Doppler cell, the DF algorithm may produce a single estimate of the direction of arrival. This arises from the limited resolving capability of the DF algorithm and it is expected that the estimated direction of arrival will usually be bounded by the limits of the direction of arrival spread within the cell. The precise value of the estimated direction of arrival will vary with time as the relative phases and amplitudes of the constituent components within the delay-Doppler cell change.

Isfjord - Cricklade

Measurements were made over this path on 7 days in late 1995 and early 1996 during the 11:00 - 16:00 UT time interval. A variety of signal characteristics were observed. In particular, at times the signal did not exhibit significant Doppler spreading, whereas at other times large Doppler frequency spreads were apparent.

A typical example of a single moded measurement made at 14.4 MHz on 31 October 1995 which exhibits Doppler spreading is presented in Figure 24. The upper frame of this figure shows the received pulse amplitude (after despreading) on a normalised scale and on a time delay scale with an arbitrary zero (the receiver and transmitter were not synchronised). The Doppler spectrum of the received signal, also on a normalised amplitude scale, is given in the right hand frame of the figure. The time / frequency dispersion characteristics (delay-Doppler) are indicated in the main part of the figure on a grey scale on which the amplitude of the peak value is represented as 0 dBr. On the delay-Doppler plot, the sidelobes of the pulse compression along the time axis have been suppressed (the despread pulse from a Barker-13 coded sounding pulse has sidelobes at best approximately -22 dB relative to the peak). The measured elevation angles of arrival indicate that propagation was probably by a 1F mode. Note, however, that owing to the limited aperture of the antenna array, poor accuracy was achieved in the elevation angle measurements. For this reason, only the azimuthal measurements are considered here in further detail. It should also be noted that owing to differences in the reference frequencies at the transmitter and receiver sites there was a small apparent frequency offset of the received signals (approximately +5 Hz for the 14.4 MHz signal of Figures 24 and 25).

A marked relationship is evident between the Doppler frequency and the measured bearing (see Figure 25). Signal components arriving at the receiver from directions to the east of the great circle path (high bearing angles) have positive Doppler shifts imposed, whereas signals arriving from the west of the great circle direction (low bearing angles) have negative Doppler shifts imposed. In this instance, the relationship between the bearing and the Doppler frequency is approximately linear with a gradient of about 3 Hz / degree. It is important to note that although a bearing spread of around 10° is observed in these data, the majority of the power is concentrated in a much narrower bearing range. In this case, the distribution of received power as a function of bearing has a mean of 6.2° and a standard deviation of 0.84°.

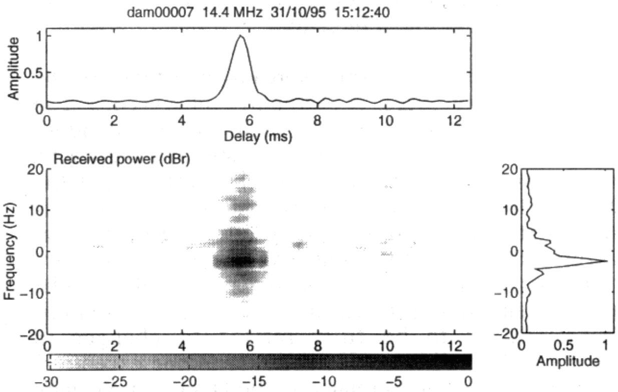


Figure 24. Delay-Doppler plot for the 14.4 MHz signal from Isfjord. 15:12 UT, 31 October 1995.

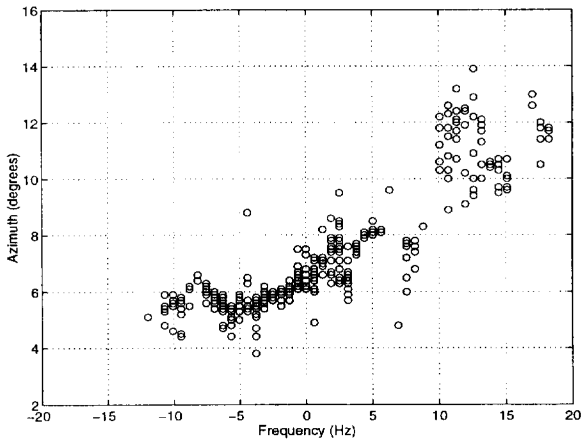


Figure 25. Variation of bearing with Doppler frequency for the 14.4 MHz signal from Isfjord. 15:12 UT, 31 October 1995.

The geomagnetic activity on 31 October 1995 was moderate (Kp = 3) and the statistical average position of the auroral oval relative to the path is indicated in Figure 23 for a time of 15:00 UT. The expected position of this disturbed region of the ionosphere is well to the north of the path mid-point (the expected reflection point for a 1F mode). However, during the DF data collection period, the CUTLASS radar detected the presence of ionospheric irregularities moving with the convection flow at velocities of several hundred metres per second in an east to west direction at latitudes similar to the mid-point of the Isfjord to Cricklade path (note that the CUTLASS viewing area at these latitudes is to the east of the Isfjord to Cricklade path mid-point). For irregularities drifting in a westerly direction, positive Doppler shifts would be imposed on any signal components scattered from irregularities to the east of the great circle direction where the motion of the scatterers was in a direction tending to shorten the path and negative Doppler shifts imposed on those scattered components to the west of the great circle path direction where the motion of the scatterers was in a direction tending to lengthen the path. The sense of the observed relationship between bearing and Doppler frequency is in good agreement with the CUTLASS radar observations of the flow direction.

Larger Doppler and bearing spreads than those of the above example were occasionally observed. The maximum azimuthal power distribution standard deviation observed for this path was approximately 2.5°. Again, the relationship between the bearing and the Doppler frequency was approximately linear, in this case with a gradient of about 1 Hz / degree. Unfortunately, insufficient experimental data are available for this path for a full statistical analysis to be undertaken and for variations in the observed characteristics to be fully related to changes in geophysical conditions. However, with one exception, where bearing changes related to the Doppler frequency were observed, these were consistent with ionospheric irregularities moving in a westerly direction and were in agreement with the radar observations (as far as is possible with the offset of the path mid-point from the CUTLASS viewing area). It should, however, be noted that the flow direction changes with time of day and Interplanetary Magnetic Field (IMF) conditions (see, for example, Figure 19). It is therefore expected that the sign and the slope of the bearing / Doppler characteristic will vary with time of day and IMF direction. The fact that only westerly drifts were observed in these data is probably due to the limited UT period during which measurements were made.

Isfjord - Alert

Measurements were made over this path during the periods 01:46 - 08:40 UT and 10:26 - 13:42 UT on 22 January 1996 and also during the period 00:06 - 03:38 UT on 25 January 1996. A wide range of received propagation characteristics are evident in the received signals and, in general, the Doppler and delay spreads were much greater than those which had been observed on the Isfjord to Cricklade path. Several interesting features of the measurements made on 22 January 1996 are now discussed in detail.

Example measurements (01:46 - 01:55 UT, 22 January 1996)

At 2.8 MHz the signal exhibits large Doppler and delay dispersion (see the upper frame of Figure 26). In addition to the broad main pulse, which is probably associated with multiple unresolved propagation modes travelling close to the great circle direction, it is interesting to note the slow decay of received energy with path delay extending for approximately 4 ms after the main peak, corresponding to an additional group path length of around 1200 km. A marked feature of this example is the large spread in bearing of the various signal components (see Figure 27), the angular power distribution of the received energy having a mean bearing of 64.7° (cf. the true bearing of 68.3°) and a standard deviation of 26.3°.

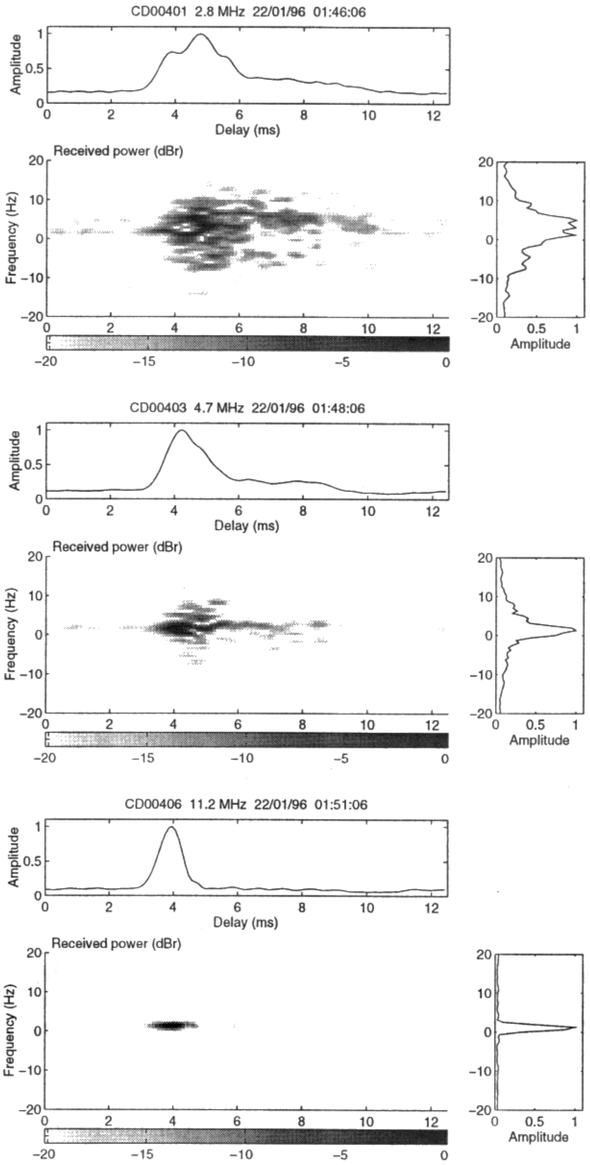


Figure 26. Delay-Doppler plots for the 2.8, 4.7 and 11.2 MHz signal from Isfjord. 01:46 - 01:51 UT, 22 January 1996. Note that the delay axis has an arbitrary zero.

The mean and standard deviations of the azimuthal power distributions for each signal frequency which was received during this period are summarised in Table 1. Also given in the table is an indication of the Doppler spread imposed on the signal by the propagation processes. This measure of Doppler spread is referred to as the Doppler Spread Index (DSI) and is calculated as the area under the normalised amplitude spectrum multiplied by a factor of 20, less a correction for the baseline noise level. Reference should be made to Warrington *et al* [7] for further details of this metric. It is emphasised that the DSI value is not a direct measure of the extent along the frequency axis.

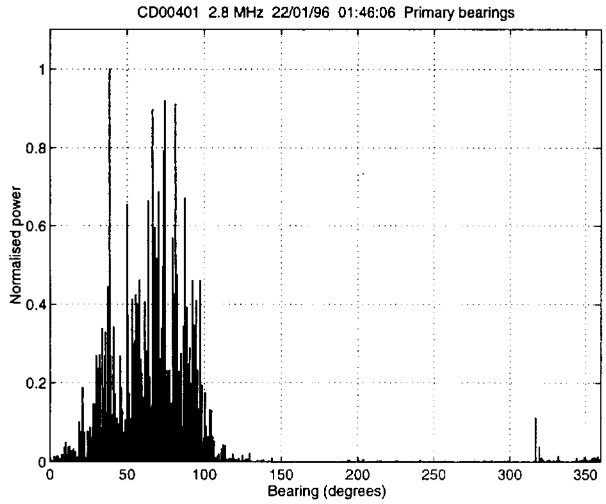


Figure 27. Distribution of received signal power as a function of bearing for the 2.8 MHz signal shown in the upper frame of Figure 26.

Table 1. Variation of the mean and standard deviation of the angular power distribution with sounding frequency. 01:46 UT, 22 January 1996 for the Isfjord - Alert path

Frequency	DSI	Mean	Standard deviation
2.8 MHz	87 Hz	64.7°	26.3°
4.0 MHz	95 Hz	67.3°	33.9°
4.7 MHz	45 Hz	66.9°	32.3°
6.8 MHz	1 Hz	74.7°	1.1°
9.0 MHz	0 Hz	69.8°	2.2°
11.2 MHz	1 Hz	71.2°	0.1°

Similar, but slightly less pronounced, characteristics are evident at 4.0 MHz and 4.7 MHz (see the middle frame of Figure 26). Of particular note at 4.7 MHz is the variation of the bearing vs Doppler frequency characteristic as the relative path delay changes (see Figure 28). The upper frame of this figure corresponds to the main delay peak and clearly illustrates that the signal components with a negative Doppler shift tend to have a bearing angle larger than the true bearing whilst components with a positive Doppler shift have bearing angles less than the true bearing. This effect was also clearly evident in the 2.8 MHz signal and is consistent with Doppler shifts imposed by reflections from irregularities in the ionospheric electron density moving with a component perpendicular to the propagation direction at the reflection points. The direction of the drift required to produce this Doppler / direction characteristic is from a low bearing angle direction to a high bearing angle direction. A similar effect is apparent at the start of the long 'tail' of the received pulse (6.2 ms on the delay axis) but with a larger bearing spread. At the end of the tail (8 ms on the delay axis), a large bearing spread is still evident but with a tendency for the bearing to increase with increasing Doppler frequency (i.e. in the opposite sense to to that observed earlier in the received pulse).

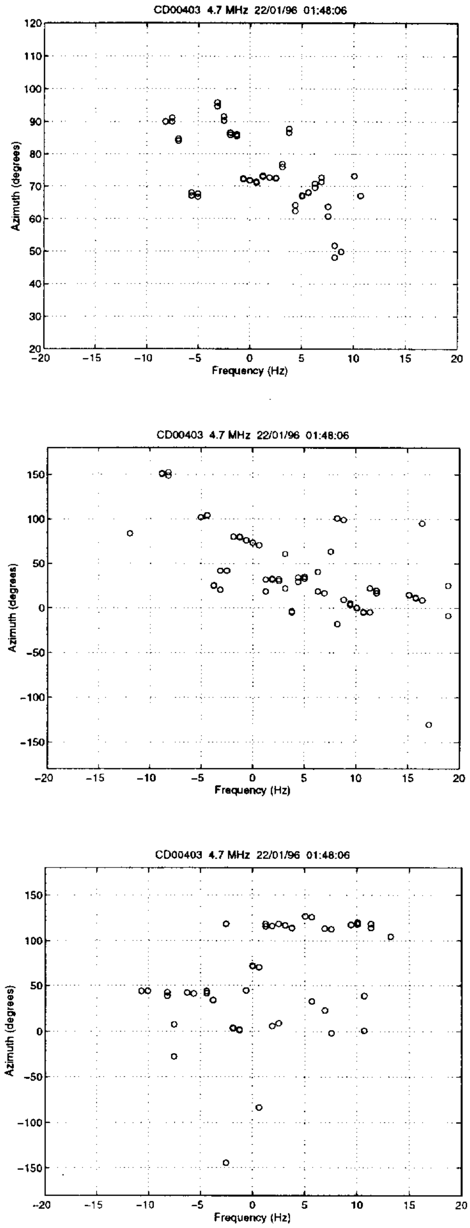


Figure 28. Variation of bearing with Doppler frequency for the 4.7 MHz sounding signal of Figure 26. The upper frame is for the main peak at 4.2 ms on the delay axis, the middle frame for the slight peak at 6.2 ms and the lower frame for 8 ms. (Note that the delay axes in Figure 26 have arbitrary zeros and also note the expanded scales on the middle and lower frames.)

As the frequency of the sounding signal is increased, the Doppler spread, the delay spread and the bearing spread decrease and a marked difference in characteristics is apparent between the 2.8, 4.0 and 4.7 MHz signals, which exhibit large bearing spreads with angular power distribution standard deviations of around 30°, and the 6.8, 9.0 and 11.2 MHz signals for which the spread is much less. Without the availability of ionosonde measurements for the path, it is not possible to identify the mode content of the received signals. However, a likely scenario is that propagation was multi-moded (probably sporadic E (Es) and various F modes) up to a frequency somewhere between 4.7 MHz and 6.8 MHz and that above this frequency propagation was only via sporadic E. This postulation is supported by the clear elevation angle measurement of about 8° for the 11.2 MHz signal, indicating a 1-hop reflection height of around 100 km. At the disturbed frequency of 4.7 MHz, elevation angles of around 8° are observed at the start of the received pulse with angles of between 20° and 30°, appropriate for F region propagation, later in the pulse. The long 'tail' following the main received pulse may result from scatter from ionospheric irregularities within the auroral zone located behind the transmitter. Such a postulation is consistent with the excess path length required to produce the time delays associated with the long tail.

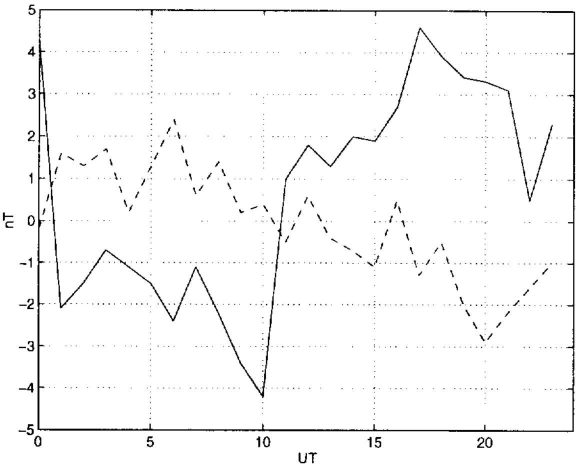


Figure 29. IMF By (solid curve) and Bz (dashed curve) parameters for 22 January 1996.

For most of the received energy, signal components with a negative Doppler shift tended to have a bearing angle larger than the true bearing whilst components with a positive Doppler shift had bearing angles less than the true bearing. The direction of the motion of irregularities required to impose such Doppler shifts is consistent with the expected anti-sunward cross polar cap convection flow. Various flow patterns for different Interplanetary Magnetic Field (IMF) conditions are reproduced in Figure 19 and the IMF B_y and B_z parameters for 22 January 1996 are given in Figure 29. At the end of the 'tail' following the pulse there was evidence of an increase of bearing angle with increasing Doppler shift. If the presence of the 'tail' is due to scatter from ionospheric irregularities from the auroral zone behind the transmitter, then it is possible that the imposed Doppler shifts would be in the opposite sense to those associated with the main pulse since the irregularities at the reflection points may be drifting with the sunward return flows (again, see Figure 19).

Propagation on paths well displaced from the great circle direction

During the period discussed in the previous section, the signals arrived at the receiver over angular distributions approximately centred on the great circle direction to the transmitter. This was not always the case, as may be illustrated by the signals received some four and a half hours later during the 06:16 - 06:25 UT interval. The presence of large deviations from the great circle direction was anticipated and some of the characteristics of such propagation, particularly the time varying nature over periods of more than tens of minutes, is discussed earlier in this paper.

The delay-Doppler plot for the 4.7 MHz sounding made during this interval is presented in Figure 30. The most noticable feature of this measurement is the presence of two distinct peaks of similar amplitude in the received pulse separated by approximately 4.3 ms, a delay much greater than, say, the 1 ms or so delay expected for 1 and 2 hop F mode. The signal components associated with each of these peaks are directionally well separated - the angular power distribution of the first mode has a mean of 68.2°, which is close to the great circle bearing, whereas the angular power distribution of the second mode has a mean of 349.0°, offset from the great circle direction by 79.3°. For the first mode, an overall trend in the relationship between bearing and Doppler frequency is not apparent, although some structure is evident. No overall variation of bearing with Doppler frequency is evident for the second pulse. There is evidence of the second mode in the 6.8 MHz sounding with a similar bearing and delay to that at 4.7 MHz. At the next higher frequency (9.0 MHz), two distinct directions of arrival are again evident. In this case, there is one component with little Doppler or bearing spread arriving close to the great circle direction together with a second, weaker mode with an excess delay of around 2 ms and a bearing some 23° from the great circle direction.

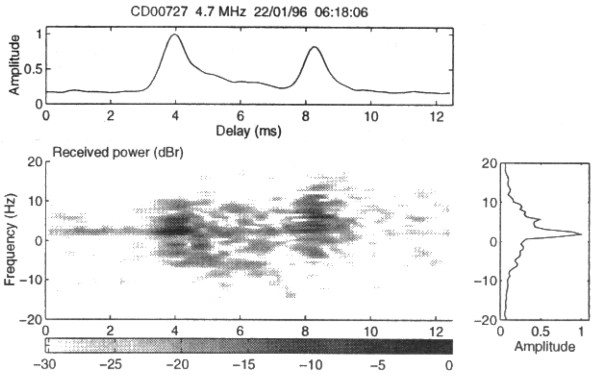


Figure 30. Delay-Doppler plot for the 4.7 MHz signal from Isfjord. 06:18 UT, 22 January 1996.

The mean bearings and bearing spreads for each received signal frequency for this period are summarised in Table 2. As with the earlier period, the lower frequencies exhibit Doppler and bearing spreads, whereas these effects are not as evident on the higher received frequencies. However, in contrast with the earlier period, both Doppler and bearing spreading is now evident on the 6.8 MHz signal. If the earlier postulation that the lower frequencies were propagated mainly by F modes and the higher frequencies by sporadic E, this observation is consistent with an increase in the maximum frequency supported by the F region during the daytime. The angular power distributions have a lower standard deviation than during the earlier interval. It is also interesting to note that the bearing and excess path delay of the secondary mode at 9.0 MHz is not the same as those of the secondary mode at 4.7 and 6.8 MHz.

Table 2. Variation of the mean and standard deviation of the angular power distribution with sounding frequency. 06:18 UT, 22 January 1996.

Frequency	DSI	Mean	Standard deviation
2.8 MHz	18 Hz	68.2°	5.7°
4.0 MHz	Not heard	-	-
4.7 MHz	30 Hz	68.2°	7.0°
	62 Hz	349.0°	7.2°
6.8 MHz	102 Hz	69.0°	8.8°
	128 Hz	351.3°	2.2°
9.0 MHz	4 Hz	69.1°	0.7°
	138 Hz	45.4°	1.6°
11.2 MHz	0 Hz	69.7°	0.3°

CONCLUDING REMARKS

Measurements have been presented of the azimuthal directions of arrival of various signals made at two mid latitude sites where the great circle paths were along the trough. Large (~50°) bearing deviations lasting for between 30 minutes and 5 hours were observed at both sites during both active and quiet geomagnetic periods. For the short Halifax to Leitrim path, these observations suggest deviations of the ray path both to the north and to the south of the GCP due to reflections from the walls of the mid-latitude trough and are consistent with the well documented changes in the morphology of this ionospheric feature with changes in geomagnetic activity. For the longer Halifax to Cheltenham path, the depleted electron density within the trough prevented great circle propagation. In these circumstances, ray tracing studies have indicated that propagation is via a 2 hop mode, the first hop being to the south of the trough and the GCP. The signal is then scattered from the sea surface and then propagates to the receiver site via a further single hop ionospheric mode.

Irregularities in the electron density distribution of the polar cap ionosphere can also result in propagation of HF radio signals well displaced from the great circle direction. Measurements have been presented of the bearings of various signals made at Alert, a very high latitude site close to the geomagnetic north pole. Large swings of up to ±100° were apparent in the bearing measurements and these have been attributed to reflections from large, drifting electron density structures such as patches of over-dense plasma and sun-aligned arcs.

Such large deviations of the direction of arrival of the signals from the GCP have serious implications for the operation of radiolocation systems operating within the HF band. These systems usually operate by measuring the direction of arrival at several receiving sites. The location of the transmitter is then estimated from the intersection of the individual lines of bearing. Deviations of individual bearings from the GCP will therefore adversely affect the estimate of the transmitter location. Although it is not thought possible to correct for the type of bearing errors reported here, it is possible to predict the periods during which the large deviations occur. These predictions could be used to assign ionospherically based weighting factors to the individual line of bearing measurements from a network of DF receiving stations in order to improve the accuracy of the position estimate.

It should be noted that for radio signals traversing this region of the ionosphere that the accuracy of bearing measurements is limited by the propagation conditions and not by the DF instrumental accuracy, which is typically of the order of 0.1°. An understanding of the propagation characteristics is therefore essential in order to achieve optimum DF performance for signals propagating via the sub-auroral ionosphere.

The directional spread in the received signal energy is also an important parameter to be considered in the design of multi-element receiving arrays and the associated signal processing methods used, for example, in direction finding or adaptive reception systems. It is often assumed in such systems that the signal environment comprises a limited number of specularly reflected signals arriving at the antenna array from well defined directions. This is clearly not the case in much of the data presented here. For the trans-auroral Isfjord to Cricklade path, for which the ionospheric reflection points were sub-auroral, standard deviations of up to around 2.5° were observed in the azimuthal power distribution of the received energy. Much more disturbed signals were received on the polar cap path from Isfjord to Alert for which azimuthal standard deviations of up to 35° were measured.

The measurements presented are consistent with the model that each propagation mode may be considered as a large number of rays reflected from an area of a roughly reflecting ionosphere, as opposed to a single specular reflection. Furthermore, it should be noted that the relative phases of the various components will change with time since each will have a slightly different frequency due to the Doppler effects. Consequently, the temporal stability of a received signal which has not been split into its various components defined by time of flight and Doppler frequency will be related to the Doppler spread. Large Doppler spreads were observed, particularly so over the polar cap path to Alert, indicating that at times such a composite received signal is unlikely to remain stable for more that several hundredths of a second.

ACKNOWLEDGEMENTS

The authors would like to thank the UK Defence Evaluation Research Agency, the Canadian Communications Research Centre, the Norwegian Defence Research Establishment, the Swedish Institute of Space Physics and the Swedish National Defence Research Establishment for their help in using the DAMSON transmissions.

REFERENCES

1. Jones, T.B. and Reynolds, J.S.B. Ionospheric perturbations and their effect on the accuracy of HF direction finders. *Radio and Electronic Engineer*, **45**, 1975, pp 63-73.

2. Gething, P.J.D. *Radio Direction Finding and Superresolution*. IEE Electromagnetic Series, Peter Peregrinus Ltd, London, 1991.

3. Moffett, R.J. and Quegan, S. The mid-latitude trough in the electron concentration of the ionospheric F-layer: a review of observations and modelling. *Journal of Atmospheric and Terrestrial Physics*, **45**, (5), 1983, pp 315-343

4. Halcrow, B.W. and Nisbet, J.S. A model of the F2 peak electron densities in the main trough region of the ionosphere. *Radio Science*, **12**, (5), 1977, pp 815-820.

5. Lane, G., Rhoads, F.J. and Deblasio, L. *Voice of America Coverage Analysis Program (VOACAP); A Program Guide*. VOA B/ESA Report 01-93, 1993, (PB93-155174).

6. Davies, K. *Ionospheric Radio.* Peter Peregrinus Ltd., London, 1990.

7. Warrington, E.M., Jones, T.B. and Dhanda, B.S. Observations of the Doppler spread on HF signals propagating over high latitude paths. *IEE Proceedings on Microwaves, Antennas and Propagation*, **144**, (4), 1997, pp 215-220.

8. Jones R.M. and Stephenson J.J. *A Versatile Three-Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere.* Office of Telecommunications, OT 75-76, U.S Department of Commerce, Washington, USA, 1975.

9. Bilitza D. (ed.). *International Reference Ionosphere 1990.* NSSDC 90-22, Greenbelt, Maryland, 1990.

10. Weber, E.J., Buchau, J., Moore, J.G., Sharber, J.R., Livingston, J.D., Winningham, J.D., and Reinische, B.W. F layer ionization patches in the polar cap. *Journal of Geophysical Research,* **89**, 1984, pp 1683-1694.

11. Buchau, J., Reinisch, B.W., Weber, E.J. and Moore, J.G. Structure and dynamics of the winter polar cap F region. *Radio Science*, **18**, (6), 1983, pp 995-1010.

12. Carlson, H.C. Jr., Wickwar, V.B., Weber, E.J., Buchau, J., Moore, J.G. and Whiting, W., Plasma characteristics of polar cap F layer arcs. *Geophysical Research Letters*, **11**, 1984, pp 895-898.

13. Gussenhoven, M.S. Extremely high latitude auroras. *Journal of Geophysical Research*, **87**, 1982, pp 2401-2412.

14. Holzworth, R.H. and Meng C.-I. Mathematical representation of the auroral oval. *Geophysical Research Letters,* **2**, (9), 1975, pp 377-380.

15. Buchau, J. Ionosonde data (Private communication), 1991.

16. Lockwood, M. Modelling the high latitude ionosphere for time varying plasma convection. *Proceedings of the IEE*, part H, 140, (2), 1993, pp 91-100.

17. Daniell, R.E. Jr., Brown, L.D., Anderson, D.N., Fox, M.W., Doherty, P.H., Decker, D.T., Sojka, J.J., Schunk, R.W., Parameterised Ionospheric Model: A global parameterization based on first principles models. *Radio Science*, **30**, (5), 1995, pp 1499-1510.

18. Davies, N.C. and Cannon, P.S. DAMSON - A system to measure multipath dispersion, Doppler spread and Doppler shift on a multi-mechanism communications channels. *In:* Proceedings of AGARD EPP Symposium on Multiple Mechanism Propagation Paths: Their Characterisation and Influence on System Design, Rotterdam, October 1993, pp 36/1-6.

19. Jones, T.B. and Thomas, E.C. CUTLASS: A new HF auroral radar. *In:* Proceedings of the IEE Seventh International Conference on HF Radio Systems and Techniques, IEE conference publication number 441, July 1997, pp 325-329.

20. Warrington, E.M. Measurements of the direction of arrival of HF sky wave signals by means of a wide aperture antenna array and two super-resolution direction finding algorithms. *IEE Proceedings on Microwaves, Antennas and Propagation*, **142**, (2), 1995, pp 136-144.

21. Mather, J. The incremental multi-parameter algorithm. Asilomar conference on circuits, systems and computers, Monterey, 1990. Published by Maple Press, San Jose, California, USA, **1**, 1991, pp 368-372.

22. Zatman, M., Strangeways, H.J. and Warrington, E.M. Resolution of multimoded HF transmissions using the DOSE superresolution direction finding algorithm. *In*: Proceedings of the IEE Eighth International Conference on Antennas and Propagation, IEE Conference Publication 370, 1993, pp 415-418.

UHF MILSATCOM Systems with Emphasis on Changes Made By the Recent Introduction of Automatic Control (AC) Mode Demand Assigned Multiple Access (DAMA)

(June 1999)

Gary Huckell, James Parsons
SPAWAR Systems Center D844
49258 Mills St Rm. 158
San Diego, California, 92152-5384, United States

ABSTRACT

Ultra high frequency (UHF) military satellite communication (MILSATCOM) has been providing service to mobile users for more than twenty years, and has become the common denominator for Allied communications. Prior to 1992 there were no formal interoperability standards governing the use of the UHF satellite system. To increase the efficiency of use of UHF satellite communication resources and to improve interoperability, a series of UHF satellite communication standards were developed. The publishing of the standards, along with a Joint Staff mandate requiring use of terminals certified to the new standards, has resulted in a tremendous surge of interest in UHF satellite communication. As many as twenty thousand new terminals, certified to these standards, will be built over the next few years. This paper begins by describing the previous and current UHF satellite constellations and summarizes how UHF satellite communications are being used. The history and capabilities of the three main UHF satellite communication standards are described along with problems that have delayed the move to demand assigned multiple access (DAMA). The paper also describes the initial development and future plans for a decentralized communications planning and management tool that will assist with creating, approving, allocating and maintaining networks of certified terminals. The paper concludes by describing work being performed to increase channel control reliability, improve the quality of secure voice, increase data rates, and enhance the ability to accommodate variable rate data protocols, including internet protocols.

BACKGROUND

The UHF spectrum allocated for US Military satellite communications is located at the boundary between the Very High Frequency (VHF) and UHF frequency bands. Uplink frequencies are located at the lower end of the UHF band (292 to 317 MHz) while downlink frequencies are located at the upper end of the VHF band (243 to 270 MHz). A pair of UHF Follow-On (UFO) satellites in geostationary orbits operates over each of four overlapping satellite coverage areas providing around-the-world coverage. Limited polar coverage is provided by two Package D satellites in Molniya polar orbits and a twenty year old Lincoln Experimental Satellite that has drifted into a highly inclined geosynchronous orbit.

Military UHF satellites contain a mixture of 5-kHz and 25-kHz bandwidth channels, each using an independent transponder. The transponders are unprocessed (they do not demodulate the data), simply filtering, frequency translating and amplifying the received signal. The use of unprocessed transponders has allowed UHF SATCOM users to take advantage of improved modulation techniques that have been developed since the original UHF satellites were launched. While twenty years ago a 25-kHz bandwidth transponder was often used at only 2400 bps, today they can be used at rates as high as 56,000 bps. The transponders hard limit received signals, providing maximum gain for weak signals, but also preventing the use of bandwidth-efficient modulation techniques that depend on amplitude modulation. Hard limiting makes it difficult for simultaneous signals to share a channel. UHF DAMA waveforms use Time Division Multiple Access (TDMA) to share channels since this technique doesn't require simultaneous access.

The UHF frequency spectrum has many characteristics that make it very suitable for mobile communications. The relatively low frequencies and data rates allow use of small, inexpensive terminals. Many current terminals actually began as 10-watt line-of-sight UHF radios. Non-directional antennas can be used since no other satellites share the frequency allocation. Directional antennas are required only when extra gain is needed to improve link margins. Because the satellite transponders are hard limiting, only minimum transmit power

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

control is required. The signal must be just strong enough to overcome channel noise while not so strong that out-of-band emissions interfere with other satellite channels. UHF frequencies operate through weather and under foliage much more reliably than SHF and EHF frequencies do.

The main problem with UHF SATCOM is that there just isn't enough of it. There are many more potential users of UHF SATCOM than there are available channels, and channel capacity is primarily limited by the UHF spectrum allocated for military use, not by an inability to build and launch satellites. The combination of earth coverage and non-processed transponders results in no anti-jam capability, and UHF is subject to ionospheric scintillation, multipath, and unintentional interference. By international agreements, satellite communication is the secondary user for the allocated UHF spectrum. The US has no legal recourse when interference is caused by other country's legitimate use of the frequencies.

UHF SATELLITE CONSTELLATIONS

The UFO satellites are 23,500 miles high in slightly elliptical geostationary orbits (e=.005), and begin life inclined 6 degrees off the equator. The four coverage areas are centered over the continental United States, the Atlantic Ocean, the Indian Ocean, and the Pacific Ocean. The UFOs are the fifth generation of UHF satellites.

Lincoln Experimental Satellites (LES)	1965Operational
MARISAT (Gapfiller)	1976
Fleet Satellite (FLTSAT)	1978Operational
Leasat Satellite (LEASAT)	1984
UHF Follow-on (UFO)	1994Operational

Other U.S. satellites, including Package D, DSCS and MILSTAR, carry UHF transponders, as do satellites owned or leased by our allies.

LES-3 was launched in 1965 on a Titan 3C launch vehicle. The purpose of this 35-pound satellite was to characterize the UHF band for military operations. LES-8&9 were launched together in 1976. These 1000-pound satellites were three-axis stabilized using momentum wheels, used pulsed plasma thrusters for station keeping, and were powered by a radio isotropic 238 Plutonium power generator. Both satellites are still operational. LES-8 is used to help locate interfering signals while LES-9 is providing polar communication services.

Three MARISAT satellites were launched in 1976 on Delta 2914 launch vehicles to provide interim operational capability while waiting for FLTSAT availability. Each satellite carried three transponders: two 25-kHz transponders and one 500-kHz transponder. The 500-kHz transponder was designed to accommodate a jam-resistant frequency-hop communication service. Designed for a service life of five years, the satellites were deactivated in 1996 after over 19 years of service.

Beginning in 1978, eight FLTSAT satellites were launched on Atlas Centaurs, but only six achieved proper orbits. Each satellite carried twenty-four transponders: twelve 5-kHz transponders, ten 25-kHz transponders, one 500-kHz transponder, and a 25-kHz transponder using an SHF uplink and a UHF downlink. Seven of the 5-kHz transponders contained onboard processing to accommodate a jam-resistant frequency-hop communication service used by the Air Force. The transponder with the SHF uplink was designed for use by the Navy Fleet Broadcasting service and contains onboard processing for a jam-resistant spread-spectrum uplink. FLTSATs 7 and 8 carry Fleet EHF Packages (FEPs). FLTSATs 1, 4, 7, and 8 are still in use providing operational communication services.

For a short time Congress mandated the use of leased rather than purchased satellites. While the Navy continued to work on a new generation of tactical/strategic tri-service satellites, five LEASAT satellites were launched using the Space Shuttle. One satellite failed early and one had to be repaired by a Shuttle crew while still in its initial parking orbit. These satellites provided a minimum capability for extending the service life of the FLTSAT constellation. Each satellite carried thirteen transponders: five 5-kHz transponders, six 25-kHz transponders, one 500-kHz transponder, and one 25-kHz transponder using an SHF uplink and a UHF downlink.

The newest constellation consists of a pair of UHF Follow-On satellites operating over each of the four satellite coverage areas. UFO-1 was launched in 1993 on the first commercially built Atlas but was placed into an improper orbit. The following eight launches were successful, providing a complete constellation. UFO-10 is scheduled for launch in 1999 to provide a spare satellite. Each satellite pair operating together provides seventy-eight transponders: forty-two 5-kHz transponders, thirty-four 25-kHz transponders, and two 25-kHz transponders using an SHF or EHF uplink and a UHF downlink. The portion of the frequency spectrum that was formally allocated to the 500-kHz transponders is now allocated to additional 5-kHz and 25-kHz transponders. UFOs 4-6 also carry Low Data Rate (LDR) EHF packages while UFOs 7-10 carry an enhanced EHF package. UFOs 8-10 carry a Ka-Band Global Broadcast System (GBS) broadcast package.

UHF MILSATCOM USERS

There is still considerable use of low rate message services (mostly 75 bps teletype networks), but secure voice, messaging networks, and information exchange networks operating at 2400 bps are the major users of UHF resources. UHF MILSATCOM is often used by early entry forces then replaced by large SHF and EHF communication systems as soon as possible. Early UHF satellites had special processed transponders to provide jam-resistant communications for a few critical communication systems but these functions are being moved to other satellites. The SHF uplink jam-resistant channel is still used for the Navy Fleet Broadcast. With thousands of new terminals becoming available, all services are expected to become large users of UHF MILSATCOM.

EVOLUTION OF DAMA

The first and still most prevalent use of UHF MILSATCOM channels is in dedicated access mode where the entire channel bandwidth is dedicated to a single communications requirement, regardless of the actual bandwidth required. There is a simple one-to-one correspondence between the number of UHF SATCOM channels available and the number of communications requirements that can be supported at any given time. The U.S. Navy developed an early version of DAMA to make more efficient use of the limited UHF SATCOM resources by providing multiple access to a UHF channel through the use of time-division multiple access (TDMA), increasing circuit availability and reducing radio requirements.

DAMA was introduced by the U.S. Navy more than 15 years ago with the development and fielding of the TD-1271 DAMA terminal. These terminals operated on a limited set of 25-kHz UHF satellite channels in DAMA Distributed Control (DC) mode, whereby users were provided access to predefined DAMA networks assigned to 25-kHz DAMA time slots on a relatively permanent basis.

The demand for UHF SATCOM resources to support the communications requirements of all the services has continued to increased at a rate that far outpaces the availability of these resources. As a result, the Joint Chiefs of Staff (JCS) mandated all UHF MILSATCOM users to transition to DAMA operation.

In 1996-1997 the Navy and the Air Force each fielded systems to provide Automatic Control (AC) mode access to DAMA channels. The Navy's DAMA Semi-Automatic Controller (SAC) provides user access to 25-kHz DAMA channels and the Air Force-developed Network Control System (NCS) supports user requirements on 5-kHz DAMA channels (with 25-kHz channel control functionality being added). These access modes are described in more detail in the following paragraph.

THE SWITCH TO AC MODE DAMA

Motorola, under contract to the Navy, developed the first Fleet Satellite Communication System (FSCS) UHF DAMA specification and the TD-1271 terminal. Although the FSCS DAMA specification was originally developed for operation on both 5-kHz and 25-kHz satellite channels, only the 25-kHz mode has been used. The FSCS standard defined two DAMA modes, Distributed Control (DC) mode and Automatic Control (AC) mode. DC mode operates as a distributed time division multiplexer, allowing multiple circuits to share a satellite channel in the same way that a multiplexer is used to combine circuits for transmission over a terrestrial link. AC mode added the capability for user terminals to control the setup of the multiplexer by requesting a circuit when needed and giving up the circuit when no longer needed. Both DC and AC modes require one terminal to act as a centralized channel controller. The main requirement for the DC mode channel controller is to establish frame timing. All TD-1271s are capable of serving as DC mode channel controllers. Under the direction of a controlling system, TD-1271s are also capable of serving as AC mode channel controllers. Motorola developed a Semi-Automatic Controller (SAC) to control TD-1271s in the AC mode. The Motorola SAC was developed when microprocessors were just emerging. Rather than use software to make the communication resource allocation decisions, SAC required a dedicated operator 24 hours a day to look at and respond to every request received from user terminals. Early in the life cycle of the SAC, this was determined to be not operationally feasible and all energy was directed toward DC mode. The Navy used DC mode exclusively until 1996.

In 1988 the Navy decided to take another look at AC mode operation. A small research and development project was funded to build a new SAC that could operate without a dedicated operator. Originally this project was intended to be nothing more than a demonstration of a capability. However, several events occurred during the development of the DAMA SAC that resulted in major changes and a major increase in scope for the project. The new UFO satellite constellation was being readied for launch and UHF MILSATCOM would now be available to all services. With the promise of more channel availability the Air Force started development of the UHF Satellite Terminal System (USTS), a DAMA system for operation over 5-kHz satellite channels. Also at this time the services were attempting to work closer together and found that most of their communication systems would not interoperate. To ensure that any new terminals developed to operate over the new UFO satellite constellation would be able to interoperate, the Defense Information Systems Agency (DISA) Joint Interoperability and Engineering Office (JIEO) Center for Standards developed a set of DAMA terminal waveform standards to provide terminal interoperability and efficient use of both 5-kHz and 25-kHz UHF satellite channels. The original DAMA mandate has been updated to require all UHF terminals being fielded to be certified to operate in accordance with MIL-STD-188-182[1] when operating on a 5-kHz DAMA channel and in accordance with MIL-STD-188-183[2] when operating on a 25-kHz DAMA channel. See Table I for a list of the MIL-STD-188 series.

Table I. UHF MILSATCOM STANDARDS

MIL-STD-188-181B
Interoperability Standard for Dedicated 5-kHz and 25-kHz UHF Satellite Communications Channels.
MIL-STD-188-182A
Interoperability Standard for 5-kHz UHF DAMA Terminal Waveform.
MIL-STD-188-183A
Interoperability Standard for 25-kHz UHF TDMA/DAMA Terminal Waveform Communications Channels.
MIL-STD-188-184
Interoperability and Performance Standard for the Data Control Waveform.
MIL-STD-188-185
Interoperability Interface Standard for 5-kHz and 25-kHz UHF SATCOM DAMA Control System.

MIL-STD-188-181[3] provides interoperability for non-DAMA users of UHF MILSATCOM. The standard is meant to minimize interference between users and to ensure that interoperability exists for all users. Only a few of the operating modes specified in the standard are mandatory. These include 2400 and 16,000 bps digital voice and 1200, 2400, and 16,000 bps data. MIL-STD-188-181A added convolutional error correction coding and Quaternary Phase Shift Keying (QPSK) modulation to the waveform. MIL-STD-188-181B, released in May 1999, added multi-h Continuous Phase Modulation (CPM), increasing the maximum data rate supported to 9.6 kbps operation over 5-kHz channels and 56 kbps operation over 25-kHz channels.

The two current DAMA standards that define operation over UHF satellite channels are MIL-STD-188-182A for operation on 5-kHz channels and MIL-STD-188-183A for operation on both 5-kHz and 25-kHz channels.

MIL-STD-188-182 protocols were originally developed to support an Air Force requirement for around-the-world messaging among Military Airlift Command (MAC) aircraft. MIL-STD-188-182 can support user data rates as high as 2400 bps and provides a messaging capability.

MIL-STD-188-183 protocols were based on the FSCS waveform introduced by the Navy with the fielding of the TD-1271. MIL-STD-188-183 can support user data rates as high as 16 kbps, but generally is used in a mode where the maximum data rate available is 2400 bps.

During the development of MIL-STD-188-182 it became apparent that the 5-kHz DAMA standard was not going to meet emerging user voice requirements. While the waveform was never intended to provide voice service, a voice capability was included. Because the waveform used an 8.96-second frame, circuit setup required 18-27 seconds and round trip voice delay was around 18 seconds, which was considered unacceptable by the user community. To reduce 5-kHz DAMA voice delays, and to address several other issues, DISA JIEO developed "A" revisions to both standards. MIL-STD-188-182A now allows a data transmission to begin before a full frame of data has been buffered. This can decrease the round trip voice delay to an average of around 5 seconds. MIL-STD-188-183A adds the ability to operate on 5-kHz time slots while being controlled by a 25-kHz channel controller. This decreases the round trip voice delay on a

5-kHz time slot to less than 2 seconds, reduces circuit setup to 4 seconds, and provides interoperability between users of 5-kHz and 25-kHz channels.

MIL-STD-188-184[4] defines an interoperable waveform allowing the error-free transmission of very long messages over 5-kHz and 25-kHz non-DAMA channels. Data compression, adaptive error-correction, and packet communications techniques are used to reliably control the flow of data over noisy communications channels at high-throughput rates.

MIL-STD-188-185[5] establishes mandatory requirements for equipment that control access to DAMA and demand assigned single access (DASA) UHF 5-kHz and 25-kHz channels. Since there will only be one manufacturer of UHF DAMA channel control equipments, this standard will no longer be maintained. Instead, the channel control requirements will be captured in system specification documentation.

DAMA TERMINALS

Towards the end of the development of the standards, the Army began a competitive development and procurement of a new man pack terminal certified to the new UHF standards. This contract was awarded to Magnavox (now Raytheon) with ViaSat, as a subcontractor to Magnavox, providing the DAMA modems. More than 5000 AN/PSC-5 "Spitfire" terminals are being built and all services have purchased substantial quantities. In 1989, prior to the development of the standards, the Navy contracted to Titan for the development of the Mini-DAMA, a FSCS-compliant terminal destined for aircraft and submarines. Mini-DAMA was started as a 25-kHz FSCS terminal, but during development, the DAMA mandate required all new terminals to be certified to all of the new DAMA standards or risk being denied satellite access. Mini-DAMA terminal fielding was delayed to allow implementation of and certification to the DAMA terminal waveform standards.

Many other companies have developed terminals using their own funds plus funding from various Government programs. Table II lists the terminals that have been submitted to the Joint Interoperability Test Command (JITC) for certification to the standards.

Table II. DAMA Terminal Certification

Nomenclature	Manufacturer	Current Certification
AN/PRC-117F Transceiver	Harris Corp.	181
AN/PSC-5 Spitfire	Raytheon Systems	181, 182, 183
AN/USC-54 VICS	E-Systems (now Raytheon)	181, 183
LST-5D Transceiver	Motorola Corp.	181, 182, 183
MD-1324 Modem	ViaSat Corp.	181, 182, 183
MD-1333/A (LSM-1000) Modem	Titan-Linkabit	181, 182, 183
MD-1293 USC-42 Mini-DAMA	Titan-Linkabit	181, 182, 183
MST ICOM/DAMA Radio (MIDR)	Cincinnati Electronic (now NOVA)	181
MXF-440 (Skyfire-440)	Raytheon Systems	181, 182, 183
MXF-460 (Skyfire-460)	Raytheon Systems	181, 182, 183
RT-1797/ARC-210A(V) Terminal	Rockwell	181, 182, 183

Currently there are three major terminal development programs, though there are many smaller programs building small lots of radios for specific platforms. The lead program is the Joint Tactical Radio System (JTRS) with a mission to develop a family of affordable, high-capacity tactical radios to provide both line-of-sight and satellite communications. This family of radios is planned to cover an operating spectrum from 2 to 2000 MHz, and be capable of transmitting voice, video and data. By building upon a common open architecture, JTRS's objective is to improve interoperability by providing the ability to share waveform software between radios.

The Air Force Airborne Integrated Terminal Group (AITG) program is under contract with Raytheon to develop an open system architecture, reprogrammable terminal. The terminal will be based on the Raytheon Multiple Output SATCOM Terminal (MOST) and will include Line-of-Sight (LOS) and satellite communications, Air Traffic Control protocols, digital secure voice, and communication security (COMSEC) for airborne applications.

The Navy Digital Modular Radio (DMR) program is under contract with both Motorola and Raytheon to develop an open system architecture, reprogrammable terminal. While similar to the AITG, the DMR will be designed for a shipboard environment, providing 8 data ports when operating on 25-kHz DAMA and include HAVEQUICK II and SINGARS capabilities. The Navy contract requires sample radios from both Motorola and Raytheon to be used to evaluate and select a single supplier of the DMR.

DAMA CHANNEL CONTROL

During the development of the DAMA standards the Navy accepted the responsibility for developing the 25-kHz DAMA channel control system required by the introduction of the new DAMA standards and the change over to AC mode operation. The Navy was already performing DC mode control for their own channels and was in the process of developing the DAMA SAC AC mode control system, although at the time DAMA SAC was being designed to control terminals designed to the FSCS specification, not the new DAMA standards. The Navy's DAMA SAC uses TD-1271 modems and WSC-5 radios installed at the three Naval Computer and Telecommunications Area Master Stations (NCTAMS) and the Naval Computer and Telecommunications Stations (NCTS) at Guam and Stockton. These control stations are all located in the overlap of two adjacent satellite coverage areas. Each control station can control satellite channels in two satellite coverage areas, providing redundant control capability for each coverage area. Equipments at these sites also control many of the baseband systems operating over the satellites and can provide data relay between adjacent satellite areas.

The DAMA SAC, which began only as a tool to demonstrate AC mode operation for FSCS terminals, was redefined to become the joint services 25-kHz DAMA control system, providing communications resources for FSCS and MIL-STD-188-183 terminals. DAMA SAC began operational service in 1996 and has performed very reliably. Just last month, the United Kingdom installed a DAMA SAC system to provide AC mode control of UK satellite resources.

The 5-kHz DAMA specification the Air Force developed as part of the USTS program was modified late in the program to implement the MIL-STD-188-182 waveform requirements. Since MIL-STD-188-182 has no DC mode equivalent, the original development contract included a 5-kHz DAMA control system. Late in 1994 the Air Force contracted with ViaSat to design, build, install, and test a 5-kHz DAMA control system, the NCS. The NCS was co-located with the DAMA SAC at the three NCTAMS and NCTS Guam (the fifth DAMA SAC site at NCTS Stockton is no longer active). In 1996 Air Force again contracted with ViaSat to add MIL-STD-188-183 AC mode channel control capability to the NCS. This will be completed in 1999.

Although developed and fielded by the individual services, the DAMA SAC and the NCS are both capable of supporting communication requirements for users operating terminals in accordance with MIL-STD-188-183. The NCS also supports 5-kHz DAMA services in accordance with MIL-STD-188-182 (and by the end of this year, MIL-STD-188-182A). These two DAMA control systems each provide user access to a separate set of satellite channels and the behavior of each control system, from the terminal user's perspective, is slightly different. From the control system operator's point of view, the operation of each of the two systems is completely different. The need for a single, integrated system became clear.

As these DAMA channel control systems were being developed, it also became evident that the ability to support more user communications requirements would significantly increase the complexity of communications planning and management. A UHF DAMA Concept of Operation (CONOP) and Operational Requirements Document (ORD) were developed and accepted by the joint services to define the requirements for the joint integrated DAMA channel control system. This system is under development and is known as the Joint (UHF) MILSATCOM Network Integrated (JMINI) Control System.

The JMINI Control System project is being led by the U.S. Navy (Space and Naval Warfare Systems Command), but is a joint-interest program and will support all U.S. and allied forces.

The JMINI Control System is the integrated 5-kHz/25-kHz DAMA and non-DAMA controller for all non-processed UHF MILSATCOM channels. The JMINI Control System consists of three main components, the Resource Controller (RC), the Channel Controller (CC), and the Network Management Subsystem (NMS). The RC does the real time processing of the DAMA orderwire, the CC consists of the modem, radio, power amplifier, and antenna system, and the NMS contains the database of UHF satellite, terminal, and network characteristics, and provides the user interface for communications planning and management.

Because of its already-fielded 5-kHz capability and its recent upgrades for MIL-STD-188-183 and MI-STD-188-182A operation, the Air Force-developed NCS was selected as the starting point for the JMINI RC subsystem. The CC requirement is to be able to operate every UHF channel in a DAMA mode, which adds up to over 300 channels globally. For the remainder of the CC requirement, the Navy will be procuring a next generation radio, the Digital Modular Radio (DMR). There was no direct product available to leverage the NMS; therefore a new product is being developed to provide the software tools to allow decentralized communications planning and management.

COMMUNICATION PLANNING

The UHF constellation has historically operated with only one network per channel. Some channels do have two networks, using frequency division to share a channel, and there are a few DAMA channels in each satellite coverage area. The process of planning which users get service, allocating networks to channels, and handling conflicts is done via manual coordination and manually with some rudimentary tools. For clarity, a *network* is defined as two or more user terminals communicating together with the same data type (voice, message, etc.) and encryption technique. As more and more channels transition to DAMA, and as the user population increases ten-fold, the process of planning and management will become overwhelming. In addition, the DAMA CONOP requires the management process itself to be distributed into the battlefield to provide necessary information required for re-planning in response to the real-time operational environment. These factors contribute to the need for automated communications planning and management tools to be provided by the JMINI Control System's NMS.

As stated above, whereas the RC subsystem is an enhancement to the existing NCS, and the CC subsystem will be provided through the DMR program, there was no product available that could be modified to become the NMS. A product that provides planning and management tools used for EHF MILSATCOM does exist, however. This system, known as the Automated Communications Management System (ACMS), did allow for some design and code reuse to begin the development of the NMS.

The starting point for the NMS was to understand the manual processes its tools will replace, leading to the development of a planning/management Concept of Operations (the NMS CONOP). As documented in this CONOP, it was highly desirable to utilize existing computer systems where possible since many facilities are space-constrained. It was also important to implement a familiar operator interface to simplify training since military departments are manpower constrained[f].

The NMS will enable decentralized communications planning and management by providing software tools to:

- Access and maintain databases of user terminal descriptions, capabilities, and configurations, network service definitions and membership, satellite ephemeris data, and MILSATCOM channel descriptions (e.g.. frequency codes, channel numbers, channel names, uplink and downlink frequencies, translation offsets, access mode);
- Apportion satellite resources among CINCs, services components and supporting echelons, and other agencies tasked to support operational missions and exercises;
- Allow these organizations and agencies to define and establish the relative priorities (ranking) of their own communications service requirements (CSRs);
- Determine best-fit communications service allocations and preassign and schedule resources for the most critical communications requirements;
- Monitor system performance to compare actual against planned operations, assisting the communications manager in finding ways to optimize the use of the limited UHF MILSATCOM resources; and
- Automate some of the planner's manual tasks.

---

[f] The Defense Information Infrastructure (DII) Common Operating Environment (COE) establishes a standardized set of applications, hardware, and User Interfaces to promote uniformity across DoD systems.

It was also understood that flexibility and modularity in software design would be extremely important since the requirements were not completely firm and the software would be developed incrementally. The software design objectives included reuse of existing modules from ACMS, portability, and to be able to add new functions simply. Therefore, the NMS is being developed using a true object oriented approach that uses a functional architecture as shown in Figure 1.

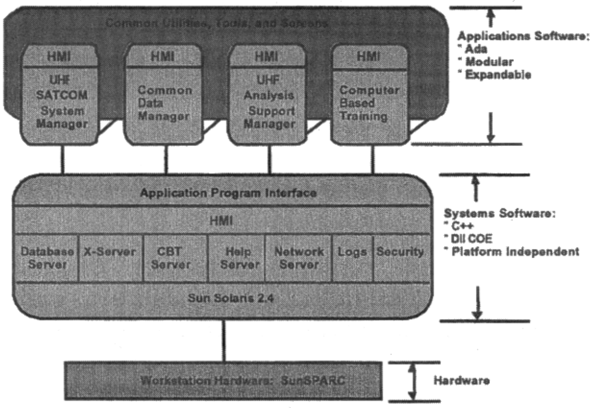


Figure 1. NMS Layered Architecture

This architecture will be fielded for IOC and is designed to be located at four main ground control points worldwide. This configuration will serve as the basis of a more elaborate distributed computing architecture for FOC. For IOC there are three main functional layers: the hardware – which is a Sun SPARC workstation, the system software – consisting of the Solaris operating system, database and other segments as an integrated portion of the DII COE, and the application software – customized modules designed for growth.

The NMS objective is to have a planning and management capability at many different levels, corresponding to "resource owners". A resource owner is typically an organizational entity that has authority to operate their network(s) over specific channels or DAMA time slots. The establishment of resource owners follows a well-documented process outlined in a Chairman of the Joint Chiefs of Staff Instruction. In accordance with NMS objectives, all users outside the four main ground control points (the JMINI Control Stations at the NCTAMS and NCTS Guam) will connect to these systems via a secure network using *any* existing computer that has access to the secure network and a Web browser. With modern client/server systems this method is very practical. The IOC configuration has been designed to allow for this expansion by using an object oriented design together with the Common Object Request Broker Architecture (CORBA). CORBA provides the "middleware" for distributed processing capabilities, promotes the maximum degree of machine and language independence and includes an interface to Java, thereby allowing easy scalability of capabilities through time. Objects representing satellites, terminals, or networks, for example, can be located anywhere within the distributed architecture and exchanged between processes using the Internet InterOrb Protocol (IIOP). This allows the Graphical User Interface (GUI) that runs on a Sun Workstation to be ported to other computers and still interface with the database for access and updates. Figure 2 illustrates the machine and language independence of CORBA.

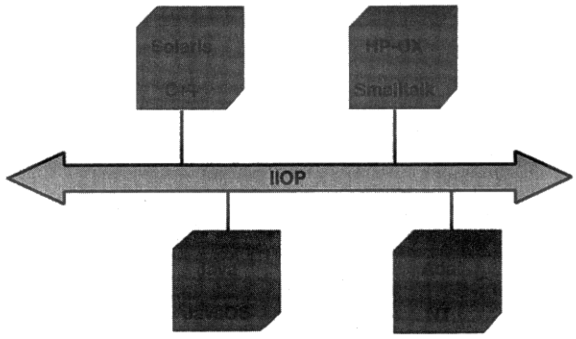


Figure 2. Machine and Language Independence

By FOC, there will be perhaps dozens of NMS users, most of which will be remote from the four control stations. They will use Web browsers to access and populate the database for the real-time resource controller. These NMS users will be able to handle the access requests, apportionment, allocation, and other processes electronically and rapidly, as well as perform their day-to-day mission even with as much as a ten-fold increase in DAMA users.

PROBLEMS WITH DAMA

While multiple access and demand assignment can greatly expand the number of users that UHF MILSATCOM can serve, nothing comes free. When a 25-kHz channel is used for a single "non-DAMA" voice network, the effective modulation rate over the satellite channel is equal to the baseband rate of the voice service. When DAMA is used, the modulation rate over the satellite channel has to exceed the sum of the data rates of all users. For example, when five users all operating at 2400 bps share a 25-kHz channel, each user is allocated 1/5 of the channel time and must operate at five times the normal modulation rate. The increased modulation rate, along with the requirement to break the signal into small bursts, requires a higher quality signal than required for single user operation. A channel that can provide good service for a single user communications requirement may operate poorly when DAMA-tized.

None of the DAMA terminals has been designed to provide the ease of use expected of a commercial cellular phone system. While the DAMA standards fully define how a terminal must communicate to the satellite, terminal communication to the user is not well defined or well understood. The requirement that the DAMA control orderwire be encrypted adds another level of complexity.

In DC mode the assignment of each time slot was fixed and the DAMA orderwire was not encrypted. A voice network was given a Circuit Identification Number (CIN) to identify the time slot it was allocated to and the user terminal operator only needed to acquire the orderwire and enter the CIN to join the network. With the introduction of AC mode, after acquiring the encrypted orderwire, the terminal operator must successfully transmit a return orderwire message to log in, transmit another return orderwire message to request to join a network, then wait for an orderwire message to direct connection to a network. If the operator has improperly set the terminal's address, port configuration code, or network address, the connection will be refused and an error message will be returned by the control system. The increased complexity of AC mode operation as well as limited user knowledge of the benefits and tradeoffs inherent in both 5-kHz and 25-kHz DAMA operation are two factors contributing to the hesitancy to move to DAMA.

The Navy was the first UHF MILSATCOM user to make the transition to AC mode. The transition was made in two steps. First, orderwire encryption was added to the existing DC mode controllers. Encryption of the DC mode orderwire allowed the Navy to identify and correct problems with orderwire encryption before attempting the switch to AC mode. AC mode operation requires orderwire encryption since orderwire requests can identify users and provide some indication of mission activity. The switch to AC mode was not problem-free. A few networks immediately experienced communication problems and AC mode was initially blamed and shut down. After extensive re-testing it was determined that the problems were not the fault of AC mode operation. Today there are 17 25-kHz channels operating in AC mode worldwide.

While technically we are operating these 17 channels in AC mode, there is currently little or no demand assignment being used. DAMA time slots are still being preassigned to each communication network. Many time slots are assigned to Navy messaging systems that require communications 24 hour a day, 7 days a week. Other time slots need to be preassigned because the baseband equipments being used are not capable of generating a DAMA request or responding to an incoming call. For example, voice networks are normally routed to a speaker. Hearing a voice from the speaker is the only indication that a "call" is coming in. Another major limitation to the use of demand assignment is Navy's use of single radios to operate in multiple networks. A single DAMA radio is used to connect to as many as eight separate networks. To operate on eight networks requires that no two time slots overlap in time. Since every platform must operate on a different set of networks, careful planning is required when assigning networks to time slots.

The transition to DAMA has progressed very slowly. In addition to the increased complexity that DAMA adds, and baseband equipment limitations mentioned above, other contributing factors are lack of terminals in the field and limited user training. Of these two problems, user training is probably the more serious. Despite the efforts of the DAMA Education team and service unique training programs, understanding of the most basic DAMA concepts seems to be lacking. For example, many users are not aware of the differences between the service quality on 5-kHz and 25-kHz DAMA channels and are, therefore, not anxious to move their secure voice requirements to DAMA channels.

Another problem that has kept many voice networks from transitioning to DAMA is user reluctance to accept the lower voice quality provided by 2400 bps LPC-10 narrowband vocoders. Some users currently use CVSD vocoders operating at 16 kbps on dedicated 25-kHz channels. This data rate is too high to be used efficiently on a DAMA channel. CVSD provides voice recognition and works well in the high background noise environment found in a helicopter and in the battlefield. LPC-10 works very poorly in a high background noise environment. MELP, a newly developed 2400 bps vocoder algorithm, solves both of these deficiencies but will not be generally available for several more years.

Users accustomed to operating over a dedicated channel are hesitant to accept the extra setup time required to operate over a DAMA channel. Users want guaranteed and immediate service when operating under battlefield conditions.

There are many UHF MILSATCOM systems using protocols that are incompatible with DAMA operation. The Navy has redesigned some of its systems to allow operation over DAMA, but many systems have requirements or funding problems that make it impossible to move them to DAMA channels.

Terminal fielding has also slowed the transition to DAMA operation. Terminals are being fielded by organization, not by network, however every member of a network must have a DAMA terminal before a network can be transitioned to DAMA operation. This problem will be resolved over time, if not by changes to fielding plans, then by the fielding of sufficient terminals. However, even if we solve the equipment and training problems, the severe interference problems we are experiencing on UHF MILSATCOM channels may become the limiting factor controlling the number of channels that can be switched to DAMA operation. As many as half of the channels currently have interference levels that are too high to allow DAMA operation.

NEW SATELLITES

The current UHF constellation will be complete after the launch of UFO-10 later this year. To prepare for the possible loss of one or more satellites during the planned lifetime of the UFO constellation, the Navy is in the process of procuring a single gapfiller satellite. UFO-11 will help ensure that adequate UHF communication resources will be retained until the next generation of satellites becomes available. While all previous UHF satellites have used analog filters and transponders, UFO-11 will use an all-digital design. Although this will make it possible to adjust the bandwidth and center frequencies of transponders after launch, the digital design may also reduce the dynamic range of the transponders. The Joint UHF MILSATCOM Technical Working Group (TWG) is investigating both the possible advantages and disadvantages of the new digital design.

The Mobile User Objective System (MUOS) is intended to be the next generation of satellites to provide narrowband communication services to mobile users. The MUOS requirements are performance-based and allow alternatives such as Low-Earth-Orbit (LEO) constellations to be considered. A key system objective is seamless operation with existing UHF systems and terminals, so the next generation of satellites is expected to provide services similar to those provided by the UFO constellation. The major MUOS goals are to greatly increase circuit capacity, increase the maximum data rate to 64-kbps, and to greatly improve the ability to communicate with very portable terminals having negative-gain antennas and very low transmit power. The most promising approach is the use of a mixture of earth coverage and spot beams. Spot beams allow frequency reuse within a satellite footprint while at the same time providing greatly increased link margins.

INTEGRATED WAVEFORM

The Joint UHF MILSATCOM TWG is working on a single integrated waveform standard that will provide a significant improvement over the two current DAMA waveforms now required for operation over 5-kHz and 25-kHz UHF military satellite channels. While the two current standards may have been sufficient for their intended purpose, they are not interoperable with each other and are not capable of satisfying all current and developing user requirements. This has contributed to the low user acceptance of DAMA, greatly slowing the transition of UHF MILSATCOM users from dedicated access to DAMA operation.

The TWG is presently developing a new layered protocol to be incorporated in revisions to the existing standards. The first revised standard will define the interoperable modulation, error-correction coding and multiple access protocols required to access the satellite channels. The second revised standard will define higher layer protocols, including demand assignment, required for full voice and data interoperability. MILSATCOM user systems not requiring interoperable voice and data would be required to implement only the lower layer protocols of the first standard. The integrated waveform standard will accommodate backward compatibility with the present standards.

The lower level multiple access protocol will operate much like 25-kHz DAMA DC mode, broadcasting stable channel and time slots assignments for all UHF SATCOM users. Setup and maintenance of these assignments will be performed by the NMS. At this protocol level terminals will have no ability to directly affect time slot assignments. A Forward Orderwire

Message (FOW) burst located within a time slot on a legacy 25-kHz controlled channel will be used to broadcast the assignment information for all channels. Navy messaging systems that require communications 24 hour a day, 7 days a week would operate on a stable time slot defined solely by this multiple access protocol, as would many voice networks that cannot tolerate the setup delays imposed by the higher layer demand assigned protocols. It is anticipated that many systems that cannot use DAMA today will be able to operate on top of this simplified multiple access protocol.

The higher layer protocols for voice and data interoperability will ride on top of the multiple access protocol layer, providing voice and data services similar to what is provided by commercial telephone systems. This protocol layer will use its own addressing plan, independent from any DAMA addressing plan, making it possible to interface to other secure voice systems.

The existing DAMA standards provide for fixed-rate "circuit" services, and fixed-length "message" services. Newer data protocols are generally designed to move data asynchronously, i.e. at variable data rates. Protocols that support the efficient handling of Internet Protocol (IP) datagrams will be developed to operate over the lower level multiple access protocols. A paper[6] presented last year at MILCOM'98 describes a proposed packet transfer mechanism, called the Variable Rate Data Packet (VRDP) transfer protocol, which could be used for this purpose.

CONCLUSION

In is anticipated that the increased circuit capacity, higher data rates, and improved quality of service that will become available with the launch of the next generation of satellites, along with the introduction of the integrated DAMA waveforms, will improve user acceptance of DAMA. The layered protocol design of the integrated DAMA waveform will allow the use of Commercial communication protocols over UHF satellite channels, providing simplified terminal operation and improved communication services similar to those provided by cellular phone systems.

REFERENCES

1. MIL-STD-188-182A, Department of Defense Interface Standard, "Interoperability Standard for 5-kHz UHF DAMA Terminal Waveform" 31 March 1997, U.S. DoD.

2. MIL-STD-188-183A, Department of Defense Interface Standard, "Interoperability Standard for 25-kHz TDMA/DAMA Terminal Waveform (Including 5-kHz and 25-kHz Slave Channels)," 20 March 1998, U.S. DoD.

3. MIL-STD-188-181B, Department of Defense Interface Standard, "Interoperability Standard for Single-Access 5-kHz and 25-kHz UHF Satellite Communications Channels," 20 March 1999, U.S. DoD.

4.MIL-STD-188-184, Department of Defense Interface Standard, "Interoperability and Performance Standard for the Data Control Waveform," 20 Aug 1993, U.S. DoD.

5. MIL-STD-188-185, Department of Defense Interface Standard, "Interoperability Interface Standard for 5-kHz and 25-kHz UHF SATCOM DAMA Control System," 29 May 1996, U.S. DoD.

6. Gary Huckell, Gary Clinesmith, Sid Graser SPAWAR Systems Center, "Integration of IP-Packet Data Transfers Within UHF DAMA (unclass)" MILCOM'98 Proceedings Classified, pp. 452-456, Nov. 1997.

Simulation of a Combat-Net Mobile Radio System with ARQ Feedback Errors

P. J. Vigneron , K. W. Moreland
Terrestrial Wireless Systems
Communications Research Centre Canada
Box 11490 Stn. H, Ottawa, Canada K2H 8S2
phil.vigneron@crc.ca, bill.moreland@crc.ca

Summary

Performance trends are reported from a simulation study of a tactical radio system for the Canadian Army, a mobile digital packet radio system for data transmission. The primary combat-net radio system uses a continuous-phase frequency-shift keying modulation to achieve a transmission rate of 1600 information bits per second, and can operate at a slow frequency hopping rate. The radio data-link protocol employs forward error-correction (FEC) coding, which is the combination of a parity bit, a (23,12) Golay code and a repetition code, as well as an interleaver to combat the effects of fades in the channel. An automatic repeat-request (ARQ) scheme is also employed for the retransmission of coded data that cannot be corrected by the coding scheme. A description of the channel simulation, the error correction and interleaving schemes, and the operation of the ARQ protocol under conditions of reverse channel errors are made. Performance trends are identified.

1 Introduction

Data transmission over wireless links is making significant communications and hence computational resources available to the soldier. While military communications systems designs are driven in part by the regulatory restrictions on the uses of spectrum, there are fundamental differences in approaches to military and civilian system design ([1], p.12) which are summarized in Table 1.

Military	Civilian
•Performance must comply with military specs.	•Performance requirements based on conventional service.
•Sparse density of users.	•High density of users.
•Noise levels are unique to tactical warfare situation	•Noise levels are predictable.
•Transportable and rugged base stations.	•Fixed-location base stations.
•Base station siting and antennas limited by tactical requirements.	•Base station siting and antennas carefully chosen.
•Interference from jamming as well as unintentional sources.	•Interference from unintentional sources.
•Performance-driven solutions.	•Cost-conscious solutions.

Table 1: Issues in military and civilian wireless communications systems design, adapted from ([1], p.12).

In this paper, a scheme for data transmission using a combat-net VHF digital mobile radio for use on land is described. Fixed-frequency and frequency hopped operation are considered. The performance of the system is evaluated by software simulation for typical terrestrial wireless fading channels.

In addition to the degradation of the received signal from the inherent multiplicative and additive distortion caused by the channel, intentional jamming can reduce performance. Typically the jammer will be a continuous-wave tone, a pseudo-noise jammer disrupting a portion of the hopping bandwidth, a partial-band jammer which disrupts some frequency content of the desired signal, or a follower-jammer which tracks and disrupts the frequency of the desired signal. In all cases the user can reduce the negative effects of jamming by using a frequency-hopping transmission scheme. For example, in frequency hopping operation, when a pseudo-noise jammer attempts to disrupt a large portion of the hopping bandwidth, the portion of that noise on any particular operating frequency will be small. More sophisticated partial-band or follower-jammers are then required, where the jamming effectiveness is enhanced by putting more of the jamming power at the user's frequency when that frequency is being used.

In both fixed-frequency and frequency-hopping

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

modes the performance of the transmission suffers due to the inherent effects of the fading channel. Deep fades in the received signal envelope are caused by the vectorial combining of the multipath signal components, and can cause runs of errors in the demodulator output. This can occur in a *Rayleigh* channel. Channels in which a fixed-path component propagates along with the multipath components, as is common in terrestrial mobile communications, are referred to as *Rician* channels. When either the transmitter or receiver is in motion, the relative physical configuration of the scatterers and reflectors is time-varying and so is the fading.

The burstiness of the errors can be reduced by interleaving the transmitted data. The remaining short runs of errors may be detected and possibly corrected if the data is encoded using block codes. When there are too many errors for correction, the detected errors can trigger the automatic repeat-request protocol to request a retransmission of data. This increases the chance of error-free reception at the cost of longer delays. These features are included in the present system design and the performance is evaluated by simulation.

The fading channel simulation used is based on one described by Hashemi [2], [3], where the output is shown to possess spectral properties predicted by R. H. Clarke ([4], p.44) for fixed-frequency operation and the statistical measures are shown to agree with theoretical values.

In Section 2 the narrowband fading channel simulator is described and the method of simulating frequency hopping is outlined. The communications system and the technique used for system simulation are described in Sections 3 and 4, respectively. The results are presented in Section 5, and the conclusions are provided in Section 6.

2 Channel Simulation

In this section, the simulation of a narrowband fixed-frequency channel excitation is described, and this is followed by an outline of the technique used for the channel simulation under conditions of frequency-hopping.

The simulation of a narrowband system subjected to multipath fading is accomplished using a channel impulse-response model with specified (but not necessarily equal) amplitudes. Angles of arrival of the multipath components are uniformly distributed, and arrival delays deterministically computed on the basis of the velocity, maximum Doppler shift and carrier frequency according to [3]. The received signal is of the form

$$r(t) = e^{j2\pi f_c t} \sum_{k=0}^{M} c_k e^{j\theta_k} e^{j\beta\nu t \cos\psi_k} \qquad (1)$$

and is available in the simulation as described in [3].

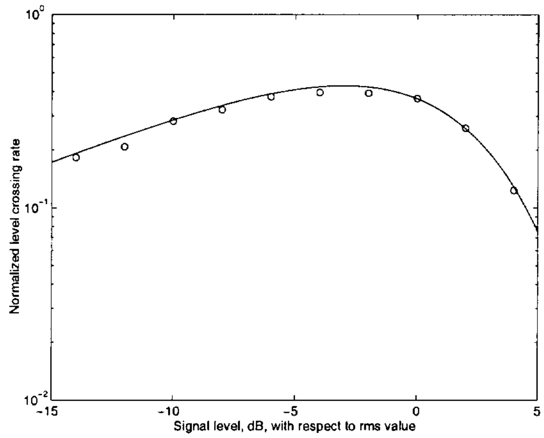


Figure 1: Normalized level crossing rate for both theoretical (solid) and simulated profiles (circles) at 312.5 MHz and 96 km/h.

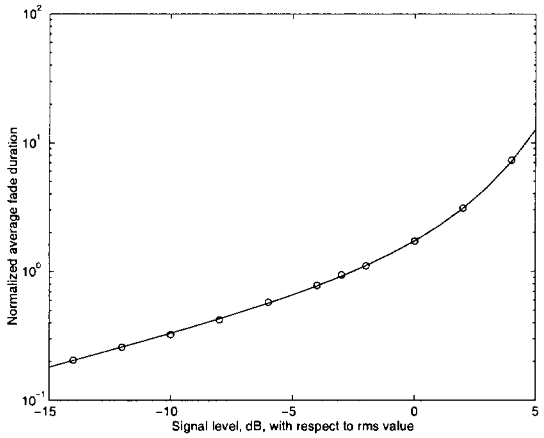


Figure 2: Normalized average fade duration for both theory (solid) and simulated profiles (circles) at 312.5 MHz and 96 km/h.

Here f_c is the carrier frequency with wavelength λ (and $\beta = \frac{2\pi}{\lambda}$), c_k is the amplitude of the k^{th} contributing multipath component with phase θ_k, ν is velocity, and ψ_k is a uniformly distributed random angle in the range $[0, 2\pi)$. This is easily implementable and is a slight modification of the simulation technique of Hashemi and Turin [2].

The theoretical normalized level crossing rate for Rayleigh fading is $N_{Ro} = Re^{-R^2}$, and the theoretical normalized average fade duration is $\bar{\tau}_o = \frac{1}{R}(e^{R^2} - 1)$, where R is the signal level normalized with respect to the root mean-squared (rms) signal envelope. Figures 1 and 2 show a comparison of theoretical and simulated level crossing rates and average fade duration results, respectively, as a function of fade depth. It is evident from the close agreement with theoretical values that the simulation is validated with respect to the second-order statistics.

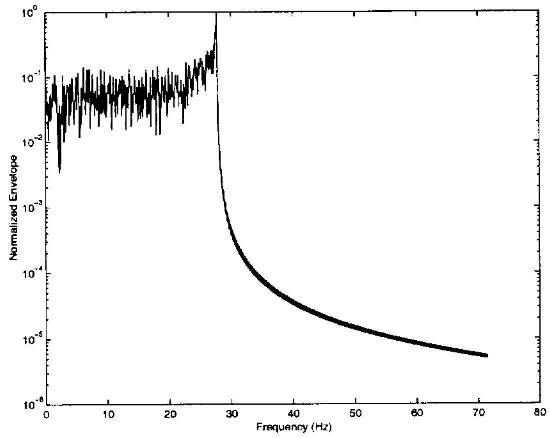


Figure 3: Spectrum of simulated response at 312.5 MHz and 96 km/h.

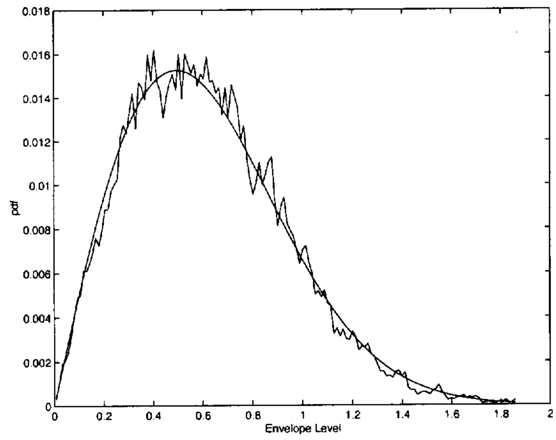


Figure 4: Rayleigh simulator envelope pdf at 312.5 MHz and 96 km/h and analytical pdf.

In Figure 3, a portion of the baseband spectrum around the maximum Doppler shift of $f_d = 27.77$ Hz is shown. The spectrum was calculated using the averaging of periodograms method, where the spectra from 200 non-overlapping segments of the simulator output were averaged. Each segment consisted of 40000 samples of the simulated waveform spaced at 266.67 μs and an FFT of size 131072. Each frequency bin was therefore 0.0286 Hz. In Figure 3 it is seen that at frequencies greater than f_d the spectral power decreases drastically. The spectrum possesses the characteristic shape of Clarke's two-dimensional isotropic scattering model ([4], p.41), described in normalized form by

$$S(f) = \begin{cases} \frac{A}{\sqrt{1-\left[\frac{f}{f_d}\right]^2}} & 0 \le f < f_d \\ 0 & \text{elsewhere} \end{cases} \tag{2}$$

where A is a normalization constant.

In Figure 4, the probability density function (pdf) computed from the simulation results is compared with the analytical Rayleigh pdf, where the Rayleigh parameter is obtained by matching the mean values of the simulation to the theoretical Rayleigh mean. The simulated pdf is demonstrated to resemble the theoretical pdf (*eq.*1.1-14, [5])

$$f_R(r) = \frac{r}{\sigma^2} e^{\left(\frac{-r^2}{2\sigma^2}\right)}. \tag{3}$$

In a communications system, r is the envelope of the received signal, *i.e.* the modulus of the complex baseband information-bearing signal.

The simulations used to generate Figures 1 - 4 were performed for a nominal carrier frequency of 312.5 MHz and a vehicle velocity of 96 km/h. Although these parameters have been chosen arbitrarily as representative for a mobile UHF channel, the specific values are not of absolute significance because the simulator results can be normalized and are applicable to any frequency and velocity combination. The basic propagation equation (1) used in the simulation has as the argument of the information-bearing portion of the signal an exponent equal to $j\beta\nu t \cos\psi_k$. This exponent term can equivalently be expressed as $j\frac{t\sqrt{2\pi}\cos\psi_k}{B}$, where $B = \frac{\sqrt{2\pi}}{\beta\nu}$, which is independent of frequency and velocity for a given B, and has time t scaled inversely by B. It is therefore possible to simulate a mobile channel such that it is appropriate for any frequency and velocity by appropriate choice of B.

It is on the basis of close agreement with second-order statistics, the resemblance to the Clarke spectral shape and a reasonable fit of the pdf of the envelope to the Rayleigh density function, that the simulation will be taken to be an accurate representation of short-term mobile radio channel behaviour.

2.1 Channel Simulation for Frequency Hopping

The simulation of frequency-hopped systems makes use of the envelope of the received signal versus time generated from the fixed-frequency simulation. First, correlation properties of the channel are noted.

The autocorrelation for the envelope of the received signal of the mobile-radio fixed-frequency narrowband simulation is given by $J_o(\beta v\tau)$ ([4], p.50) where τ is the time delay variable and $J_o(\cdot)$ is the Bessel function of the first kind with order 0. Correlations of the received signals at different frequencies are $\frac{1}{1+(\Delta w)^2 T_m^2}$ ([5], p.50), where Δw is the frequency separation in radians per second and T_m is the delay spread in seconds. For a frequency separation of Δw_o, the signal correlation is $\frac{1}{1+(\Delta w_o)^2 T_m^2}$, and this corresponds to a time shift of

$$\tau_o = \frac{1}{\beta v} J_o^{-1}\left(\frac{1}{1+(\Delta w_o)^2 T_m^2}\right) \tag{4}$$

where $J_o^{-1}(a)$ is taken to be the smallest value of the argument A at which $J_o(A)$ takes value a.

A frequency hop simulation which preserves correlation properties of the signal envelope is implemented using the simulation of a fixed-frequency system with the required signal wavelength and velocity. At the end of a dwell-time, the hop is simulated by skipping the channel simulator output corresponding to a time τ_o, which is derived from the size of the frequency hop Δw_o, using (4). The fixed-frequency simulation time series is then used for the duration of the dwell-time. This is repeated whenever a hop occurs, with a different random frequency change Δw_o and corresponding time τ_o (4). It should be noted that the hop bandwidth is typically narrow relative to the centre frequency, so wavelengths at the different frequencies are not appreciably different. When a system has centre frequency 40 MHz and hop bandwidth 5 MHz, the maximum difference in wavelengths is 11%. Therefore a simulated received signal envelope at the centre frequency is taken to be representative of the true signal at all frequencies in the 5 MHz band.

3 Description of System

In this section, the communications system under study is described. Issues and assumptions concerning the simulation are noted where relevant.

The VHF radio uses digital FM that is continuous-phase frequency-shift keyed (CPFSK) ([4], p.175). The radio is capable of transmitting in either fixed-frequency or slow frequency-hopped modes in the frequency range $30 - 108$ MHz. The data rate transmitted on air is 16 kbps, and the hop rate is 100 hops per second [6]. Figure 5 shows a block diagram of the coding and interleaving scheme [7] employed. Blocks of bits are used for generation of a parity bit. The forward error-correction code is a Golay (23,12) code ([1], p.457) which corrects three and detects four bit errors. A frame of data is encoded into 24 bits, 23 bit Golay codewords and the parity bit, and enters an interleaver row-wise in groups of 16 codewords. The 24 columns of 16 bits are sequentially read out, and are bit-wise repetition coded, where the repetition factor is an odd number, and is taken to be 5 in the present work. The combination of the repetition factor and the Golay-coding scheme causes 120 transmitted bits to contain 12 bits of information, which means that the on-air transmission rate of 16 kbps results in an information rate of is 1.6 kbps. When the radio operates in fixed-frequency mode the bit-stream enters the FM modulator and is transmitted. When frequency hopping mode is used, the bit stream is further interleaved (before modulation and transmission) in the *radio interleaver*, taking in 165 columns sequentially of 16 bits each. Contents of the 16 rows are then modulated and transmitted in a pseudo-random order.

The receiver performs signal demodulation and follows the pseudo-random hopping pattern of the transmitter to re-assemble the contents of radio interleaver blocks. The system overhead required for the synchronization is not considered in this simulation. The repetition decoding is performed by a majority-decision vote. When the Golay decoding process detects errors that cannot be corrected, a retransmission will be requested. Failure of the parity check also causes a retransmission request. The frame is then re-assembled from the received bit stream.

In the event that the FEC scheme is unable to correct bit errors caused by the channel, a selective-repeat ARQ scheme has been incorporated into the transmission system. The acknowledgement messages consist of messages encoded in the same way as in the forward channel. However, these acknowledgement messages are shorter, having fewer Golay codewords. The general operation is depicted in Figure 6. When the error detection capabilities of the FEC scheme identifies errors in the received frame that cannot be corrected, a NACK may be returned to the transmitter, which identifies the codeword in error, and it is assumed that the particular codeword is retransmitted. When the received signal is either error-free or has corrected errors, an ACK is returned to the transmitter. In both cases, the signals on the reverse channels are subject to errors caused by fading on the reverse channel, as with the forward channel.

4 Simulation Technique

The inner workings of the simulation techniques are now described. The simulation is used to obtain average frame error rates and histograms of system delays. The received signal level envelope versus time profile of the channel simulation, in either fixed or hopped frequency configuration, is at the heart of the simulation. The simulations are set up identically for fixed or hopped frequency operation; the respective received signal-level envelope versus time samples from the channel simulation are used.

A simulation is done for a given nominal carrier-to-noise ratio (CNR). The channel simulation envelope versus time data are denormalized so the mean power gives the desired CNR. Each sample of this time series is then converted to an instantaneous bit error rate (IBER) using a look-up table approach, from an IBER versus CNR curve. This IBER versus CNR curve describes the hardware operation of the radio, taking into account the transmitting and receiver filters, the receiver noise and the modulation scheme. The curve can be obtained either by a baseline radio performance simulation, or by laboratory measurement. In [8], the simulation of a CPFSK radio has been made, and the resulting IBER versus CNR curve is shown in Figure 7. In the present simulation a sixth order polynomial which has been fitted to the curve in Figure 7 is used.

Recognizing that the 16 kbps radios have symbol periods much shorter than typical coherence times of VHF channels, the channel simulation output does not have to be sampled at the baud rate of the radio, but at a rate consistent with the fading rate. Received signal levels at the inverse of the baud rate can be obtained by linear interpolation, thereby affording a significant savings in simulation time. In ([9], p.62) a typical Doppler spread of 4 Hz corresponds to coherence time of .25 sec, which is many times the $16000^{-1} = 62.5\ \mu s$ symbol period. Interpolated values of IBER for each bit in the received data stream are therefore obtained. An error in decision for each bit detection is declared on the basis of a comparison of the IBER with a uniformly distributed random number in the interval $[0, 1)$. An error state for each of the bits is then known, and the de-interleaving and decoding of the FEC scheme is done explicitly. Failure to successfully decode the Golay block code is declared on the basis of the "dependent-error" method for block decoding, described in ([10], p.616). It should be noted that the technique of simulation assumes that as the modulation is narrowband, hence in the radio transmission there is no intersymbol interference.

A criterion for an ARQ retransmission is based on the unsuccessful decoding of the Golay codewords, and the state of the parity check. Simulations are made of the time delay encountered for correct transmission of a frame of data, where greater delays are associated with more ARQ retransmission cycles. Simulations also show the frame reception probability after only one ARQ acknowledge or retransmission cycle, enabling the comparison of relative performance over different parameters or channel models. Included in the simulations are the effects of reverse channel errors. The ARQ is a selective-repeat scheme, where individual codewords are acknowledged or retransmitted.

The benefit of direct simulation of ARQ operation is the ability to capture delay information. Semi-analytical techniques which take possible errors in the reverse channel into account [11] give throughput results but do not provide time-delay related results.

5 Results and Discussion

Simulation results are presented and described in this section. First, simulation parameters are given.

A frame of data consists of a number of codewords that have the form determined by the FEC-interleave-repetition scheme indicated in the transmitter block diagram in Figure 5. In the simulation results, frames consist of 15 codewords for frequency-hopping operation and 33 codewords in fixed-frequency operation.

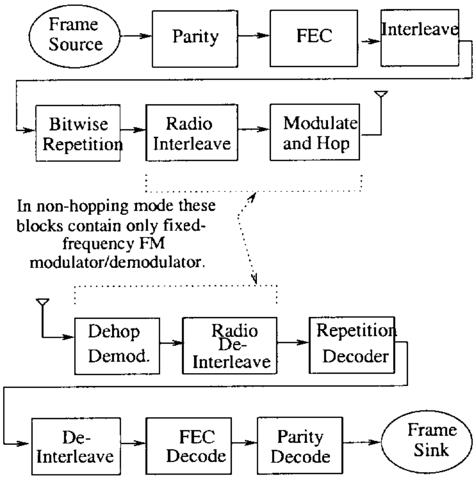


Figure 5: Radio block diagram.

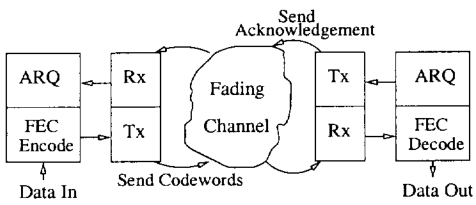


Figure 6: Transmission and acknowledgement of messages.

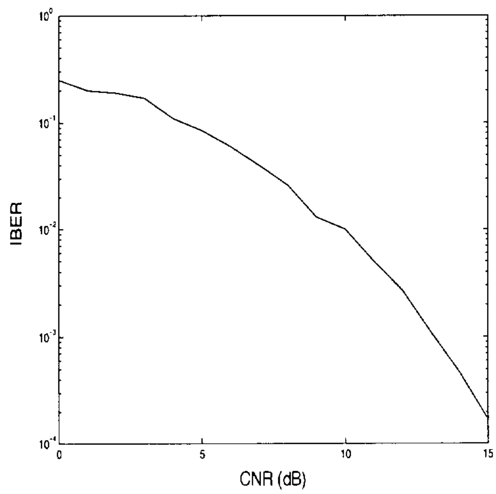


Figure 7: Simulated characteristic curve for radio.

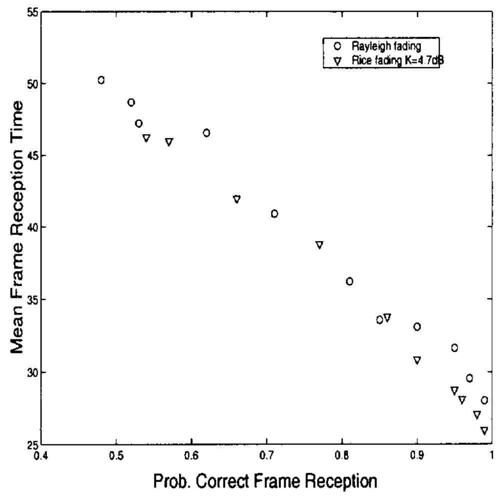


Figure 8: Delay results with ARQ, and corresponding frame reception prob. for one ARQ cycle with frequency hopping, 5 MHz bandwidth and 40 MHz centre frequency. Delay is measured in codewords, with 15 codewords per frame.

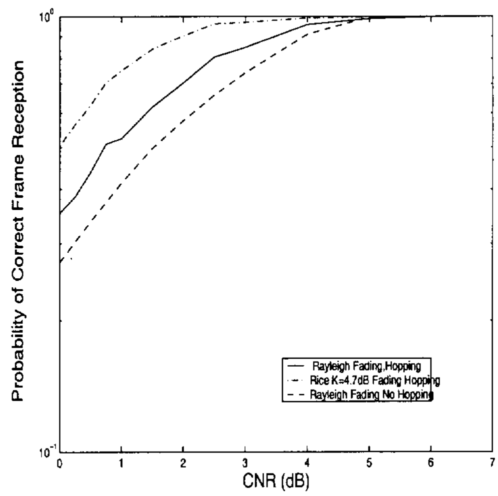


Figure 10: Frame reception results with one ARQ cycle versus average CNR for hopping and non-hopping operation with 40 MHz centre frequency and 5 MHz hop bandwidth.

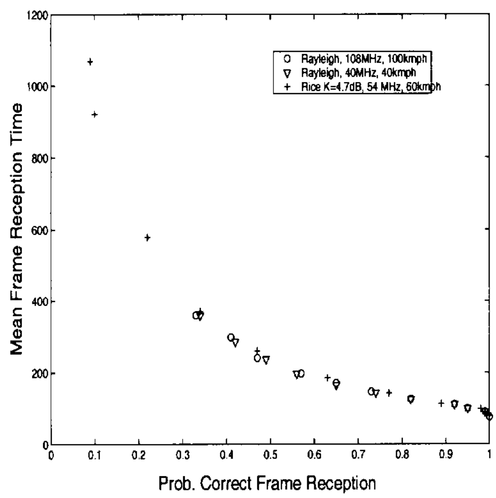


Figure 9: Delay results with ARQ, and corresponding frame reception prob. for one ARQ cycle with fixed-frequency operation. Delay is measured in codewords, with 33 codewords per frame.

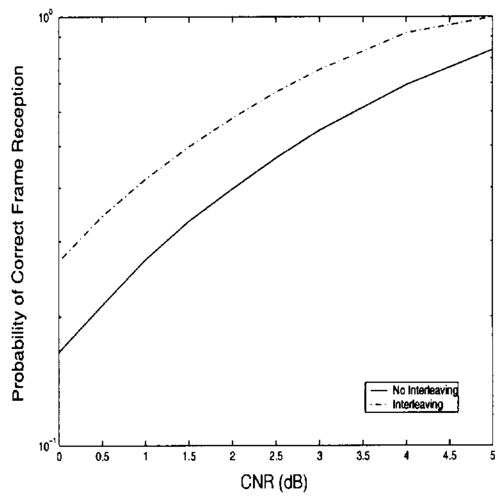


Figure 11: Frame reception results with one ARQ cycle versus average CNR for fixed frequency operation in Rayleigh fading with 40 MHz centre frequency.

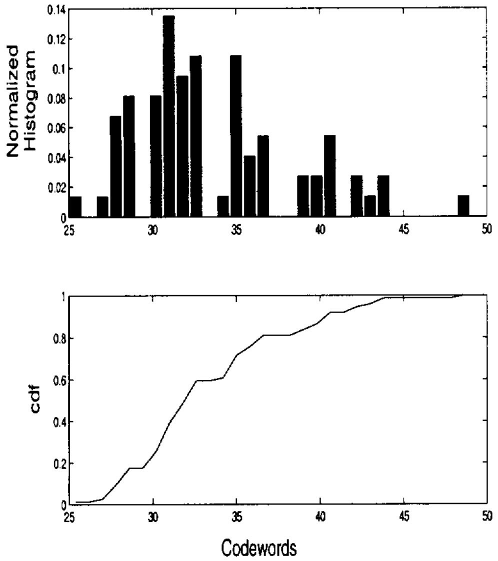


Figure 12: Histogram and cdf of transmission times, measured in codewords, with frequency hopping and Rician $K = 4.7$ dB fading, bandwidth 5 MHz, centre frequency 40 MHz, and 1.5 dB CNR.

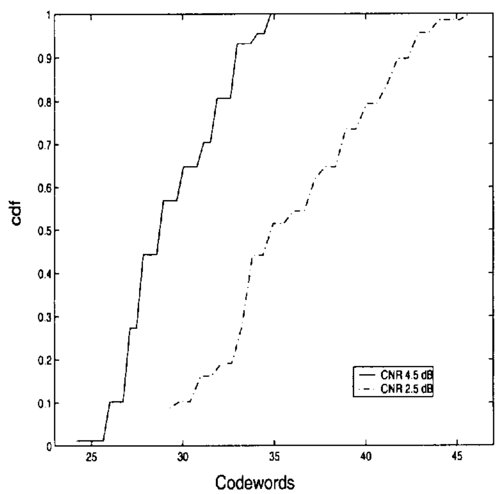


Figure 13: Cdf of frame transmission time, measured in codewords, for frequency hopping operation in Rayleigh fading for different CNR values.

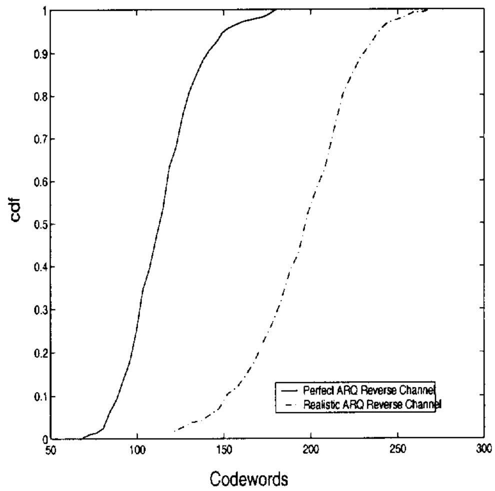


Figure 14: Cdf of frame transmission time, measured in codewords, for fixed frequency operation and Rayleigh fading showing effects of ideal ARQ reverse channel assumption, with 2 dB CNR.

These values are chosen in part from design requirements and in part on the basis of simulation time constraints required to obtain results. The number of symbols represented in a simulation run generating one frame error probability is 25 million for fixed-frequency operation and approximately 3.1 million for frequency-hopping operation. The frames on the reverse channel, used for ARQ signalling, are assumed to be shorter than those on the forward channel. In both the forward and reverse channels the delay spread is taken to be 1 μs. The reverse channel is taken to be the same type (*i.e.* Rayleigh, or Rician) as the forward channel, but with independently random behaviour.

In Figure 8 the average number of codewords required for the proper reception of a frame, a measure of delay, is shown for frames of length 15 codewords on the ordinate axis. The abscissa contains the frame reception probability after one retransmission, a measure of performance after the first ARQ cycle. This includes the effects of possible errors on the feedback path. The scale on the ordinate axis is in terms of codeword intervals, where one frame with no delays will require 15 units for reception. Simulation points are shown for operation in both Rayleigh and Rician fading channels. The higher the probability of frame reception after one ARQ cycle, corresponding to a large value on the abscissa, the lower the transmission time. This indicates that the greater the ability of the interleaving and FEC schemes to combat transmission errors, the resulting reduction in delay, and hence the reduction in number of retransmissions. Similarily, results for fixed-frequency operation with length 33 frames in Figure 9 indicates the same relationship. The ratio of delays (*i.e.* of delay to frame length, in terms of numbers of codewords) in Figures 8 and 9 for probability of correct frame reception 0.5 is approximately three for hopping operation and six for non-hopping operation. The results for Rayleigh and Rician channels shown in these figures are comparable, demonstrating how the use of ARQ reduces the sensitivity of the system to these sorts of channel conditions, when comparing frame probabilities and delay results.

Frame reception probability after one ARQ cycle for different channel models versus carrier-to-noise ratio, taking into account possible errors in the acknowledgement message on the reverse channel, is shown in Figure 10 for hopping operation. The results for the Rician channel are clearly superior than those for the Rayleigh channel. Also shown in the figure are results for fixed-frequency operation with the same frame lengths of 15 codewords as the hopped systems, giving poorer results than the comparable frequency-hopping systems. This illustrates the beneficial effects of the frequency hopping acting effectively as an interleaver to reduce the burstiness of errors due to fades.

In Figure 11 the frame reception probability after one ARQ cycle for a fixed-frequency system is shown along with that for the same system with no interleaving. This demonstrates to what extent the interleaver is successful in combatting the burstiness of the errors in the fading channel.

The probabilistic behaviour of frame arrival time, or delay, is demonstrated in Figure 12. The histogram shows the relative frequency of arrival times for a frame of 15 codewords in Rician fading, where the minimum arrival time would then be 15. Notice that the range of arrival times is from 25 codewords to just below 50. The corresponding cdf is also shown in Figure 12. This transmission delay is dependent on carrier-to-noise ratio, in Figure 12 the CNR is 1.5 dB. This dependency is seen in Figure 13 for Rayleigh fading where the delay cdf is shown for two different CNR values, demonstrating a difference in delay of 17% for a 2 dB difference in CNR at the 0.5 probability level. The delays are caused by errors in decoding the forward path messages and by the need for retransmissions. The CNR also affects decoding of the reverse-path message, and errors in these messages also contribute to delays. Increased CNR will reduce the occurrence of decoding failures in both directions.

The optimistic results arising from an ideal reverse-channel assumption are observed in Figure 14. The delay cdf is obtained for the particular ARQ scheme used in the simulation. At the 0.5 probability level the delay is seen to be roughly a factor of two lower when the simplifying assumption of no reverse channel errors is observed, an assumption often seen in the literature.

6 Conclusions

This paper has presented the simulation of data transmission with fixed-frequency and frequency-hopped VHF digital radios for communications in a terrestrial mobile environment. The techniques for the channel simulation in fixed and hopped modes have been developed. The communications system and the technique used for simulation have been described. Simulation results show appropriate density and spectral properties, as well as level crossing rates and average fade durations.

More reliable frame transmission after one ARQ cycle corresponds to lower transmission delays, a property followed for all channel conditions considered. Relative delays are lower for frequency-hopped than for fixed-frequency operation. Correct frame reception is more likely in fixed-frequency systems in Rician rather than Rayleigh channels for given carrier-to-noise ratios, and the use of frequency hopping is shown to improve upon this performance because of the improved distribution of errors. The use of interleaving improves frame reception probability. The benefit of a strong CNR on reducing delay was shown. It was also shown that errors in the reverse ARQ channel increase the simulated delay, with the consequence that system design using perfect reverse channel assumptions are optimistic in terms of delay.

Further work would include investigating the use of a more robust FEC coding scheme on the reverse channel, and the minimization of the message lengths on the reverse channel.

References

[1] W. C. Y. Lee, *Mobile communications engineering, 2nd ed.* Toronto: McGraw-Hill, 1998.

[2] H. Hashemi, "Simulation of the urban radio propagation channel," *IEEE Trans. Vehic. Technol.*, vol. VT-28, no. 3, *p.* 213-225, Aug. 1979.

[3] M. Fattouche *et al.*, "Modified simulation of the urban radio propagation radio channel (SURP)," *Electronics Letters,* vol. 29, no. 21, *p.* 1825-1826, 14 Oct. 1993.

[4] G. L. Stüber, *Principles of mobile communication.* Boston: Kluwer, 1996.

[5] W. C. Jakes, Jr., Ed., *Microwave mobile communications.* New York: Wiley, 1974.

[6] N. Chan *et al.*,"Performance of digital FM frequency hopping radios with cosite interference," *in Proc. Wireless95, p.* 357-367, Calgary, 1995.

[7] N. Chan and C. Jansen, "Performance of digital FM radios in a very high frequency fading environment," *in Proc. Wireless98, vol. 1, p.* 294-306, Calgary, 1998.

[8] S. Kularatna, "CNR(P) baseline simulations with AWGN channel", *Internal Memo*, Communications Research Centre, Ottawa, Canada, 1996.

[9] K. Pahlavan and A. H. Levesque, *Wireless information networks.* Toronto: Wiley, 1995.

[10] M. C. Jeruchim *et al.*, *Simulation of communication systems.* New York: Plenum, 1994.

[11] R. J. Benice and A. H. Frey, "An analysis of retransmission systems," *IEEE Trans. Commun. Technol., p.* 135-145, Dec. 1964.

Testbed for the Evaluation of Battlefield Information Management Techniques Applied to a Low Bandwidth Tactical Wireless Communications Environment

(June 1999)

Jean-Claude St-Jacques and Dr. Allan W. Gibb
Defense Scientists
Defense Research Establishment Valcartier
2459 boul. Pie XI nord
Val-Belair, QC, G3J 1X5, Canada

SUMMARY

Mobile communication is an important military requirement. Voice communications still occupy a pre-eminent place in Army operations. Present-generation digital data communications at the tactical level (below Brigade) are often accomplished using radio systems designed primarily with voice in mind. Data throughput tends to be very limited (300-600 bits/second is not uncommon) and highly variable. If one regards the wireless communication network as a data pipeline, there are essentially three possible ways of improving the situation: (1) increase the size of the pipeline (new/improved radios or communications hardware – desirable, but often unaffordable); (2) optimize transmission through the pipeline (network management techniques); or (3) be as smart and efficient as possible about what is put into the pipeline. The potential of the third approach is often overlooked. This paper describes a testbed being developed to study the impact of information management techniques, applied at the level of the application database in each participating node of a simulated tactical radio network, on the quality and timeliness of information distribution across nodes.

INTRODUCTION

Modern armies are undergoing a revolution in the way information is managed on the battlefield. Voice-based command, control, and communication systems are being complemented by, and in some cases replaced by (in whole or in part), digital command, control and communication systems. Digital systems offer the promise of increased battlefield awareness through a more systematic and automated distribution of relevant data than is possible with a voice-based communication system. To deliver on this promise, the communication backbone must be capable of distributing digital data among participating command and control nodes with high fidelity and a timeliness appropriate to the operational scenario. To be useful, critical information must be passed quickly enough to permit the friendly commander to stay within, and act within, the decision cycle of the enemy commander.

At the tactical level (below Brigade), units are highly mobile and have no choice but to use a broadcast medium (combat net radio) as the primary means of communication. Armies of limited means do not, in general, have the luxury of separate duplicate radio nets for passing voice and data. One radio system serves both roles. Typically, individual radios have the capability to be configured to pass voice only, data only, or voice and data. In the latter case, because voice communications still occupy a pre-eminent place in army operations, the radio system usually permits a voice transmission to pre-empt a data transmission already in progress. The data transmission is interrupted and resumes once the voice transmission has ended, after a predetermined waiting period. Voice transmissions can suffer partial loss of information and still be intelligible. Data transmissions are much more demanding in this regard. Loss of even a single bit of information may render an entire message unintelligible. Moreover, the effective data throughput will be a strong function of the varying background traffic level on the communications network. In the extreme case where radio silence is imposed, the data bandwidth can go to zero for a period of time.

The reality is that tactical command and control systems based on digital technology for which combat net radio is the primary communication means will have difficulty in many situations in passing enough digital data among participating nodes in a timely way to respond to operational requirements. Unfortunately, they will have most difficulty precisely in those cases where the comprehensive up-to-date situation picture that they can provide is most useful, i. e. when the battlefield situation is evolving rapidly and there is a high level of communications traffic.

In this context, to take maximum advantage of the digital command and control system it will be necessary to manage carefully the information in digital form that is sent over the communications network. Some management techniques, such as adaptive routing or just-in-time packet construction[1] can be implemented in

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

the network layers. Others are best implemented in the application layer. This paper deals with evaluation of techniques that can be implemented in the application layer, i.e. in the command and control nodes themselves.

CURRENT SITUATION

Field Operations Requirements and Network Topology

The tactical battlefield imposes strong physical and procedural constraints which largely determine how information will be shared. Most entities, especially at the lower echelons, are highly mobile; those entities are dependent upon an unreliable broadcast medium (radio) for sharing digital data. The radios can send and receive on only one assigned frequency at a time. The entities participate as nodes on a command and control network. A group of nodes communicating on a common assigned frequency is referred to as a sub-network. The nodes have the following characteristics:

(a) nodes can come together to form temporary sub-networks;
(b) a node can participate on only one sub-network at a time;
(c) nodes can join or leave existing sub-networks;
(d) nodes can be reassigned roles, causing them to change sub-networks;
(e) nodes can stop receiving data for extended periods due to physical separation, terrain interference, equipment breakdown, or imposed radio silence.

An analysis of the operational requirements in a tactical radio communications environment[2] leads one to several conclusions: (1) a node needs to be as autonomous as possible, i.e. it should have a local data set over which it has full ownership to permit it to function when it is temporarily disconnected from a sub-network; (2) nodes on a sub-network need to propagate changes on a peer-to-peer basis so that each node on a sub-network, at any given moment, has the same database content (i. e. consistent data sets in replicated databases); this is the only model which will permit nodes to be reassigned roles without requiring substantial one-time data transfers between nodes; and (3) data recovery, required when a node has been disconnected from a sub-network for an extended period of time and wants to recover missing data, needs to be carefully managed since it implies substantial one-time data transfers between nodes that can seriously interfere with the continual update process required to maintain database consistency across nodes;

Characteristics of the Communications System

The Canadian Land Forces will field in the next years a new communications system, the TCCCS/IRIS system. One segment of TCCCS/IRIS, the Combat Net Radio Primary (CNR(P)) is of particular interest in this paper because it is the means that each field element at the tactical level will use to communicate voice and data messages with other elements. By definition, a CNR environment has limited bandwidth when compared to wired local area networks (1 Kbit/second versus 1-10 Mbit/second). To better understand the importance of managing the information flow, it is worthwhile to review the characteristics of the CNR(P) segment that have an impact on data throughput.

In the Canadian army, the CNR(P) segment will be the main communications 'pipe' that links together all tactical field elements , i.e. battalion and below. The radios involved in this segment are either deployed as manpacks or vehicle-mounted units. They are VHF FM-radios with built-in voice and data encryption and they can be operated in fixed frequency or frequency-hopping mode. Transmission mode is half-duplex (send and receive but not simultaneously). Transmission type can be voice or data but not both simultaneously, i.e., voice is not encoded and multiplexed with digital data. Voice transmission has priority over a data transmission in progress, i.e. the data transmission is interrupted and resumed later when voice transmission is terminated.

The native bandwidth of CNR(P) is about 16Kbits/second but the technique for error correction consumes considerable bandwidth. The radio uses a 'majority vote' technique for error correction that requires each packet of digital data to be sent n times, where n is an odd number (typically between 5 and 11). Received packets are compared on a bit-by-bit basis. The transmission of the packet is successful only if a majority of received packets have identical content. If not, the transmission is rejected and must be repeated. The useable bandwidth in this case is equal to the native bandwidth of the communication system divided by n. This technique reduces the useable bandwidth to the range of only 3Kbits/second when n=5. When all factors described previously are included, the effective data throughput of a combat net radio can easily slip well below 1Kbit/second. Another factor that has significant impact on the available bandwidth is the key-up time (time to establish a stable carrier) required by a radio before transmitting. Key-up time for CNR(P) is about one-half second.

All of these radio characteristics which limit data throughput impose on the command and control application a requirement to better manage what is put in the 'pipeline'.

Situation Awareness System (SAS)

The Canadian Army will also field along with the TCCS/IRIS system a command & control application, the situation awareness system (SAS). This system automatically determines position using an on-board GPS and periodically broadcasts the positional information to friendly elements over CNR(P). It will reduce the number of voice transmissions necessary to pass friendly positional information but will create a continual stream of data transmission traffic.

Some modeling of voice and data transmission traffic for SAS operating over TCCCS/IRIS has been carried out recently. The objective of the modeling was to determine the average time between broadcast of each SAS message in order to keep the net utilization below 70%, which is the threshold for optimal performance; going over 70 % you get too many collisions. The general result of interest for this paper is that, when voice traffic plus the need to distribute additional digital tactical information (target reports, enemy contact reports, fire orders, etc) are factored in, the pause between friendly position reports necessary to keep net utilization below 70% can become very long, possibly to the point of being operationally unacceptable.

Message Exchange

Traditional military communication through use of structured military messages in a prescribed format (such as the NATO ADatP-3 format) imposes an essential discipline on the information transfer process but also imposes considerable communications overhead. The message is passed as a self-contained entity consisting of a set of data fields with separators plus metadata about the message itself (message number, date/time group etc). At the receiving end, the message may be parsed into individual data fields, and the data fields stored as individual elements within the database, tagged with an identifier of the originating message that contained the data field.

Under this message-based approach, databases serve as repositories for the data fields of structured messages, and the database organization reflects that fact. The approach is not ideally suited to a very low bandwidth environment because it involves communications overhead associated with the message itself and routinely leads to re-transmission of the same information in consecutive messages. Both factors waste bandwidth.

POSSIBLE IMPROVEMENTS

Among the COTS data distribution solutions currently available, none have been shown to work within the constraints of tactical communications, i.e. in the CNR environment. It is possible to solve the distribution problems by adding new high capacity digital radios specifically dedicated to data networks but this approach is costly and not all countries can afford it. Improvements can be made in the communications layer and in the application layer but attacking the problem in each layer in isolation is not sufficient. Solutions in both layers must be able to work together. The application layer must be aware of the communications network status in order to adjust its distribution strategy.

Communications Layer

Significant improvement can be obtained if a broadcast protocol that has minimum data overhead, that is based on negative acknowledgment and that is able to adapt the data packet size in relation to the bit error rate is used where applicable. The protocol overhead for addressing and error checking should be kept to a minimum. For example, a TCP/IP packet carries a 60 byte header overhead, which can represent a significant portion of overall message length for small messages. If positive acknowledgement were used then the duration time of each acknowledgement transmission, including the key-up time of the radio and the protocol overhead, would be a significant additional expense in terms of bandwidth.

Theoretically, for reasons of efficiency, each transmission would consist of a single packet containing the message being transmitted. However, in an environment where error rates are likely to be high, long messages would benefit from fragmentation into multiple packets. In the event of bit errors, only corrupted packets need to be retransmitted.

A method called "just-in-time packet construction" that could be applied in the communications layer, as recommended by Chamberlain[1], would be the implementation of priority queues in which are stored single sub-packet size data updates ready to be sent over the radio network. When the radio is ready to transmit, the communications protocol loads into one optimal packet those updates having the highest priority.

Finally, it is important that the communications layer provide the application layer with average distribution times in order to adjust the distribution strategy according to the network state, which can vary from 0% to 100% availability.

Application Layer

Model-Based Approach: The model-based approach to communications on the tactical battlefield, as described by Chamberlain[1], offers the potential for a substantial improvement in the utilization of available bandwidth and strong support for nodal autonomy. Under the model-based approach, the database schema are designed to support a situational model of the battlefield. Each local node maintains its own situation model in its database. The primary role of tactical data communications is to update each other's database rather than to exchange messages. Structured messages may still be exchanged when complex sets of data such as fire orders or situation reports need to be transmitted. However, simpler messages such as location reports and contact reports would be handled wherever possible as database updates without duplication of information. The model-based approach is better suited than the message-based approach to the implementation of adaptive information management techniques that are sensitive to a changing situational picture.

Information Management Techniques: The most promising improvement to information distribution will be the use of information management techniques in the application layer These techniques can take two basic forms - those which prepare information in the most efficient possible form for transmission, and those which limit what is transmitted and/or when it is transmitted. To be useful, the techniques must satisfy three requirements:

(a) they should be based upon accepted operational procedure, and implemented in a form which permits modification when operational procedures change;
(b) they should be capable of being applied automatically, with only very occasional user intervention;
(c) they should be sensitive to, and capable of adapting to, both the operational context and the conditions on the communications network.

Techniques which prepare information in the most efficient possible form for transmission (sometimes referred to as streamlining techniques) include the following:

(a) classical data compression techniques for both alphanumeric and video data;
(b) use of common lookup tables in both sending and receiving databases;
(c) transmitting database updates in place of structured messages wherever possible.

Techniques which limit what is transmitted and/or when it is transmitted include the following:

(d) algorithms which prioritize information for transmission according to user-defined rules;
(e) algorithms which scan outgoing data fields, detect if they have been previously transmitted, and arrange to send only those data fields not previously sent;
(f) intelligent message queuing, in which outgoing message content is constantly examined so that messages of higher priority are advanced in the queue and messages with stale information are removed from the queue.

A detailed discussion of the above techniques is beyond the scope of this paper. Classical data compression algorithms such as those used to transmit data over modems or files over the internet, or to store images on videodiscs, compress the data just prior to sending or storage, then decompress the data upon receipt. Some information may be lost in the compression/decompression process. Use of lookup tables permits complex battlefield entities (e.g. 25th Motor Rifle Regiment) to be assigned simple alphanumeric codes (e.g. M25); when the entity is referred to in an outgoing data field, it is always replaced by its code.

These techniques can be helpful in conserving precious bandwidth, but tend to offer fixed predetermined compression gains. Techniques (d) through (f) in the above lists, properly implemented, make use of data about the battlefield situation and communications network status contained in the local database to manage information flow. They offer the potential for a truly adaptive response in the application layer to fluctuating network conditions.

IMPLEMENTATION ISSUES

Database and Data Model

To apply the adaptive information management techniques, four things are required – (1) an up-to-date local situational picture, (2) a means of rapidly extracting data about pertinent aspects of the local situational picture, (3) a means of triggering an information management strategy based on this information, and (4) a means of implementing that strategy.

In the model-based approach, the local situation picture is kept up to date (requirement 1) through a data replication mechanism for the propagation of database updates across nodes on a sub-network. Extraction of pertinent aspects of the local situational picture (requirement 2) is facilitated by the fact that the data model contains all important entities found on the tactical battlefield and accurately models their interrelationships. The problem reduces in most cases to one of extracting attribute and state information for database entities that correspond exactly to battlefield entities.

Active database techniques[3,4,5,6] provide an ideal tool for triggering implementation of information management strategies based on changes in state of specific database elements (requirements 3 and 4). In an active database, incoming data are compared with a set of pre-defined queries called "triggers". When a trigger "fires" (i.e. when the conditions contained in the queries are satisfied), an associated action is executed. Any action may be invoked by the trigger, including actions to control the flow of information. The utility of active database techniques in adaptive replication of data across tactical wireless sub-networks has been demonstrated in the work of Chamberlain[1] and MacDonald Dettweiler[2]. Most major commercial vendors of database management systems support the use of triggers. The actions are implemented through the use of stored procedures, often implemented as SQL scripts.

Conflict Resolution and Data Ownership

With replicated data comes the possibility that local operators will independently and simultaneously update the same entity in the databases at two or more different nodes.. A conflict arises when the replication mechanism tries to propagate these conflicting updates among the nodes.

When a conflict is detected and resolved by an operator or by the system, the resolution process generates system updates that must be distributed to the other participating nodes to ensure data integrity, thus creating additional communications traffic. A simple method to reduce conflict resolutions is to avoid them by assigning single ownership to elements in the data model. Under this scheme, at any time a given entity is 'owned' by only one node. Only that node can modify the value of the entity. The modification is then replicated to other nodes. Data ownership may be transferred, but not shared. This strict approach to data ownership is necessitated by the fact that, in a distributed system of autonomous databases using peer-to-peer replication, complex interactions with the data model make it difficult to implement any simplistic or generic conflict detection or resolution schemes.

Relational vs Object-Oriented Databases

Three broad types of commercial database management systems (DBMS) are presently available – relational, object-oriented, and object-relational. Relational databases organize information exclusively in the form of tables. Relational database technology is relatively mature (the product of twenty years of evolution) and well entrenched in many enterprise systems. Object-oriented database technology, by comparison, is relatively immature. It lacks the widely-accepted rigorous approach to data structure, data integrity, and data manipulation found in relational DBMS. Object-oriented databases command only a modest market share. However, growing interest in distributed object technology (as exemplified by middleware such as CORBA and DCOM) as a basis for distributed systems is fuelling renewed interest in object technology. Object-relational databases, which combine the relational and object approaches in an attempt to capitalize on the strengths of each, is a third option presently offered by major database vendors.

From the point of view of implementing information management techniques, any of the above database types can be used, provided the database management system supports the concept of "triggers". One is not restricted, for example, to using an object-oriented database because the distributed command and control system is based on object technology. It is entirely feasible to implement relational databases in the nodes of a distributed CCIS built on distributed object technology (middleware). When relational tables are selected for broadcast, they are converted to objects, transmitted as objects using the middleware layer, and then re-transformed into relational tables at the receiving end. The choice of database type should be based on its merits in providing support for adaptive information management using the model-based approach, rather than whether or not it conforms to the object paradigm.

TESTBED DESCRIPTION

Need for a Testbed

It is possible to analyze each information management technique individually and to project its impact on the battlefield for certain limiting cases. However, because the information management techniques discussed in this paper are adaptive in nature and intended to be applied together, the combined effect in operationally-realistic conditions is extremely difficult to assess unless those conditions are simulated with some degree of fidelity. What is required is a tool which permits a) realistic tactical scenarios to be played out, b) the information management techniques to be applied to information and data passed between command and control nodes involved in the scenario, and c) data to be recorded during the scenario which will allow the impact of the techniques to be assessed both during and after the running of the scenario. After examining options ranging from storyboarding through testbed to full-scale simulation, the authors concluded that the most cost-effective approach would be the construction of a testbed specifically designed to assess the techniques in question.

Concept

The testbed will consist of a set of workstations linked together through a communications systems workstation. The latter workstation will contain a software module that simulates relevant characteristics (transmission delay, bit error rate) of the wireless communication system. Scripted tactical scenarios, spanning a spectrum of tactical Army operations, will be played out which require information to be passed between workstations through the communications system simulator. The information being passed will take the form of either structured messages or database updates. The workstations will have just the functionality necessary to mimic relevant functionality of true tactical command and control workstations (e.g. a geo-referenced map display with moveable icons). The (fixed) scenarios will be capable of running without men in the loop through use of automated scripts, or manually with men in the loop stationed at the C2IS node consoles.

Characteristics of the Communications System Simulator

The simulator will be able to:

1) impose a standard delay on the information being passed (corresponding to an assumed level of background communications traffic);
2) impose a substantial additional delay on the information being passed (simulating a data transmission interrupted by a voice transmission);
3) impose a small additional delay on the information being passed (simulating delay associated with retransmission of packets forced by bit errors)
4) block the information being passed (simulating a failed transmission)

The testbed architecture will also permit the alternative of having scenarios run with inter-node communication accomplished through the use of actual combat net radios inserted in place of the communications simulator. The use of radios provides a more realistic representation of field conditions, but permits less precise control over individual parameters than is possible with a simulator.

Implementation of Information Management Techniques

The information management techniques within individual workstations will take the form of 'triggers", i.e. rules tied to specific database elements within the workstation database which are triggered by a change of state of the database element and cause certain actions to be taken. The testbed will incorporate software "switches" permitting the rules affecting individual database elements to be turned 'on' and 'off' selectively before the running of a scenario. This will permit the effectiveness of different combinations of information management techniques in managing information flow over the limited communications 'pipeline' to be evaluated. The testbed will also permit a given information management technique to be turned 'on' or 'off' during the running of a scenario, without interrupting the running of the scenario. The scenarios will be capable of being run in real time or at an accelerated speed, depending upon the purpose of the run.

Measures of Performance, Measures of Effectiveness

Algorithms will be implemented to permit the performance and effectiveness of the information management techniques to be visualized in real time and/or evaluated through subsequent analysis. The quantities to be measured include, but will not necessarily be limited to, the following:

(a) time delay in passing a database update or structured message from sending node to receiving nodes (recorded by the communications simulation module);
(b) time since last update for certain critical information (measured at each command and control node; what is considered critical will be a function of the role assigned to that node);
(c) inconsistency in database content across nodes (measured by comparing on a continual basis, using a parallel high-speed connection between nodes, the database content in neighboring nodes). Consistency is the property that the same fact in two different databases contains the same value.

The testbed will incorporate software tools which can provide graphic displays of the functioning of the communications module, display alerts when timeliness criteria for database updates are exceeded, and display alerts when consistency criteria for database content are exceeded.

Proposed Architecture

The physical testbed architecture is shown in Figure 1. Simulated command and control nodes communicate through a combat net radio (CNR) simulator which is a software module located on a distinct node.

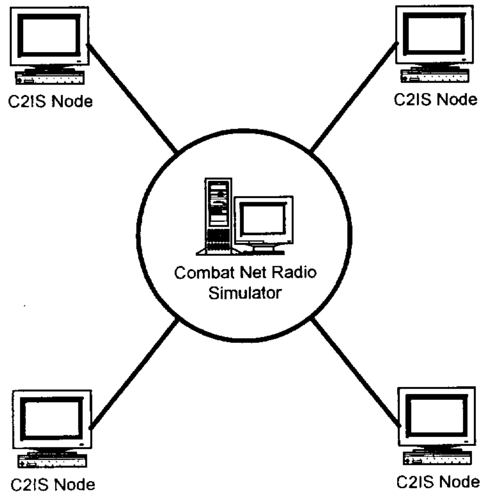


Figure 1. Physical Testbed Architecture

The logical testbed architecture is illustrated in Figure 2. Fixed scenarios can be played out automatically through automated scripts linked to an event generator or manually through people stationed at consoles at each C2IS node. In each case, a local action triggers one or more updates to the local tactical database. The replication mechanism applies the information management (distribution) rules for the changed database elements and makes a decision as to whether the update information should be broadcast immediately to other nodes, held for later transmission, or not broadcast at all. Regardless of the decision taken, the local database updates are copied directly via a high speed wired TCP/IP link to a 'truth' database residing in the central communications node. This database represents the common picture that would be shared if all local database updates were shared automatically, instantaneously, and without error – in other words the ideal that the system is trying to achieve. If the replication mechanism commits to broadcasting the database update, the data are passed through the CNR Simulator Application Program Interface (API) to the CNR Simulator. The Simulator broadcasts the data to the other nodes using a broadcast protocol (User Datagram Protocol), after applying a delay to simulate the conditions of a digital radio transmission.

The C2IS node will also receive database updates 'broadcast' from other nodes. This information is received via the UDP link between the CNR Simulator and the Simulator API. After confirmation of its integrity, the database update is applied to the local tactical database, triggering further local actions such as display updates and (possibly) the sending of database updates in response. Database monitoring and analysis tools will permit information about the time since last update for certain critical information to be monitored, and alerts to be displayed when timeliness criteria are exceeded. A high-speed wired link between tactical databases, not shown on the diagram, will also permit consistency in database content across nodes to be monitored, with alerts being displayed when consistency criteria are exceeded.

The CNR Simulator Node will also contain management tools for pre-programming the effects of digital radio transmission to be imposed by the simulator, as well as monitoring tools to measure the impact of these communications effects on the pattern of traffic through the simulator.

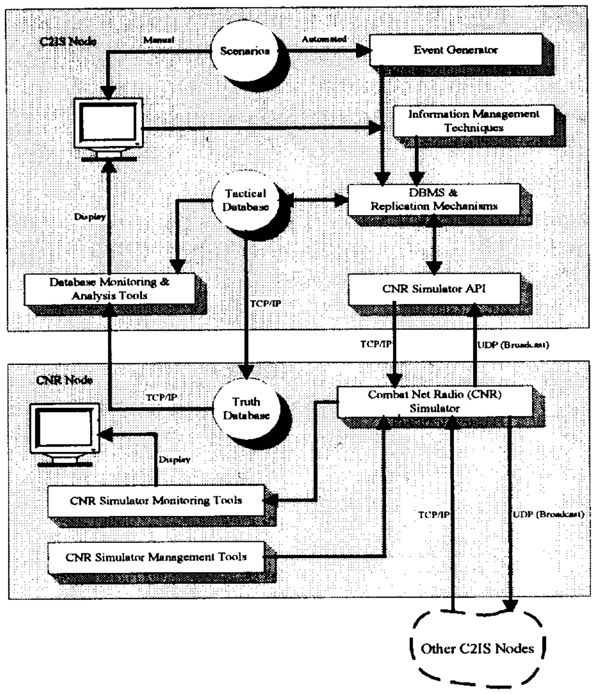


Figure 2. Logical Testbed Architecture

CONCLUSION

This paper has described a concept for a testbed to study the impact of information management techniques, applied at the level of the application database in each participating node of a simulated tactical radio network, on the quality and timeliness of information distribution across nodes. Such a testbed will permit conclusions to be drawn about the operational architecture and operational procedures required to wring maximum benefit from a distributed digital tactical command and control system forced to distribute data over unreliable low bandwidth communication links. One value of such an approach is that the conclusions drawn about the information management techniques have lasting value, since they are not dependent upon the details of the wireless communications system or network. In particular, most of the conclusions remain valid even if the bandwidth of the radio system is increased in the future.

The testbed project is in the advanced planning stages at DREV and will last three years.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Gabriel Otis and Dr. Guy Vézina, past and present Section Heads of the Information Systems Technology Section at DREV, and Mr. Denis Gouin, Leader of the User Functions Group at DREV for their support and encouragement of the work reported in this paper. We also acknowledge very helpful discussions with Dr. Sam Chamberlain of U. S. Army Research Laboratory on resilient data replication mechanisms and the model-based approach to battle command. We are grateful for the contributions of Eric Lebel, Jacquelin Bureau, and Alain Gagné of CGI in evolving the testbed concept and architecture, and for the insights of Dr. Robert Webb of Human Systems Incorporated on the subjects of scenario generation, implementation of measures of performance and effectiveness, and presentation of results. We are also grateful to Mr. Eric Dorion of DREV for discussions concerning the use of middleware in a low bandwidth context and to Mr. Paul Labbé of DREV for sharing his knowledge of information management techniques in a low bandwidth tactical naval context. Finally, a special thanks is given to the ARDS/ADM project team for their pioneering efforts in Canada in exploring solutions to the difficult problem of maintaining consistency in tactical databases over unreliable low bandwidth communication links.

REFERENCES

1. S. Chamberlain, "Resilient Data Replication Mechanisms for Battle Command Systems", in Proceedings of the Second International Symposium on Command and Control Research and Technology, Monterey, CA, pp 1-11, 1996.
2. MacDonald Dettweiler, Artillery Regimental Data System Advanced Development Model – Data Distribution Technical Note 140, Government of Canada Contract Number W8477-L-EO14A/01-SV, 1995.
3. D. McCarthy and U. Dawal, "An Architecture of an Active Database Management System", in Proceedings of the ACM SIGMOD, Portland, OR, pp 215-224, 1989.
4. D. Cohen, "Compiling Complex Database Transition Triggers", in Proceedings of the ACM SIGMOD, Portland, OR, pp 225-234, 1989.
5. E.N. Hansen and J. Widom, "An Overview of Production Rules in Database Systems", IBM Research Report RJ9023, IBM Research Division, 1992.
6. U. Dayal, E. Hanson, and J. Widom, "Active Database Management Systems", in Modern Database System: The Object Model, Interoperability, and Beyond, ACM Press and Addison Wesley, Reading, MA, 1995.

iSTAR Radio Network for Tactical Use

F. Eken and Ş. Uzun
Communications Division
ASELSAN Military Elec. Ind. Inc.,
P.O. Box 101, 06172, Ankara, Turkey

ABSTRACT:

The tactical battlefield is now becoming a ground for extensive digital data exchange where many systems, sensors, weapons, command centers need to exchange high speed data in order to perform effectively. More so, these units need to carry out their data exchange while on the move because the new military doctrines heavily emphasize mobility and flexibility.

Many tactical command control functions such as maneuver control, fire support, SHORADS (Short Range Air Defence System), electronic warfare, intelligence and logistics rely on tactical mobile radio communications. No army can afford to have dedicated and separate radio systems to support different applications. What is needed is a single tactical radio system that will be able to support most if not all the needs of the tactical area.

Considering the requirements of the future battlefield a new concept for the future digital tactical radio has been introduced. This new concept is called iSTAR (Integrated Services Tactical Radio), VRC-5100. The iSTAR concept is based on radio networking and packet communications. On the tactical field, iSTAR radios automatically form a radio network where all the network management functions are carried out in a distributed fashion.

INTRODUCTION:

New requirements for tactical command control, and rapid technological advances generate increasingly sophisticated weapons and systems which completely revolutionize the battlefield. All these changes impact most heavily the area of tactical communications which is to carry the information required to execute the desired command control functions.

The basic requirements which should be addressed by a future tactical radio communication system are summarized below:

Mobility: Technological advances reduced the duration of combat from months and years to days and hours. In this new environment the need for an uninterruptible mobile communication system becomes very crucial. Therefore, the tactical mobile radio system should be easy and flexible to deploy, set-up and operate.

Electronic Protection Measures: Today's technology allows gently killing the enemy i.e. electronically disabling the communications and tracking capabilities. As a consequence, a modern tactical communication system must provide a reasonable level of protection against electronic warfare threats. The protection should cover both jamming and direction finding threats.

Integration of Services and Interoprability: Both voice and data communication services are required in the tactical area and the users cannot afford to carry separate radios for these services. All the data and voice communication services for the tactical users should be integrated and provided by one single radio system.

Mobile users also demand communication services integrated with other available communication infrastructures such as Tactical Area Communication Systems, Strategic Communication Systems and the PTT. Interoperability enhances survivability and enables wide area networking.

Radio Networking: Computer systems will become an essential part of the digitized battlefield. Any future tactical radio communications system should explicitly provide support for computer communications. That is, the tactical radio system

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

should offer networking facilities to the users. Radio networking also facilitates automatic relaying and routing of data.

Automatic Relaying and Routing of Data and Voice over Radio*:* Availability of radio connectivity is always a concern for mobile communications which can be impaired by terrain effects or jamming. When radio connectivity degrades, setting up relays become cumbersome and limit the mobility and reduce the survivability of communications. For future tactical radio systems, having the ability to automatically relay both voice and data traffic without affecting the quality of service to its own users will extend the range of communications and enhance the survivability of the system.

Near Real Time Data Communications*:* The sophisticated sensor and weapon systems which are becoming increasingly important for tactical command and control require high speed near real time data communications. This kind of sensor to weapon data communication is intolerable to delays. Thus, tight delay constraints are imposed to the mobile radio system which has to deliver the information to the destination in time.

Packet Switching*:* Efficient use of available bandwidth in tactical communications is of great importance. Packet switching provides the means for efficient utilization of communication resources. Using the virtual circuit concept, packet data communications also facilitates point-to-multipoint data communications which is an important requirement of tactical command control applications.

Position Location and Distribution: Conveying geographical position information to other friendly units is very crucial in the tactical field. Gathering up-to-date geographical positions of the subordinate units is vital to the tactical commanders. A tactical radio communication system that automatically distributes the position information of each individual unit over the radio network will greatly enhance the tactical decision making process and also reduce the voice communication traffic.

Distributed Management*:* Distributed network management is the key for radio network survivability. Centralized control approach will create an Achilles heel in the system. The network management functionalities of the radio units should be uniform.

iSTAR RADIO NETWORK :

iSTAR is a new generation radio family which combines both SCRA (Single Channel Radio Access) and Packet Radio concepts. The iSTAR (Integrated Services Tactical Radio) concept is based on radio networking and packet communications. On the tactical field, iSTAR radios use Time Division Multiple Access technique and automatically form a radio network where all the network management functions are carried out in a distributed fashion.

The iSTAR radio network can provide many simultaneous voice and data connections to the mobile users. iSTAR radios have built-in encryption, EPM and LPI/LPD features which enable secure communications in the battlefield. The iSTAR radios are also equipped with GPS (Global Positioning System) receivers the radio system automatically distributes the GPS information of each mobile unit over the entire network.

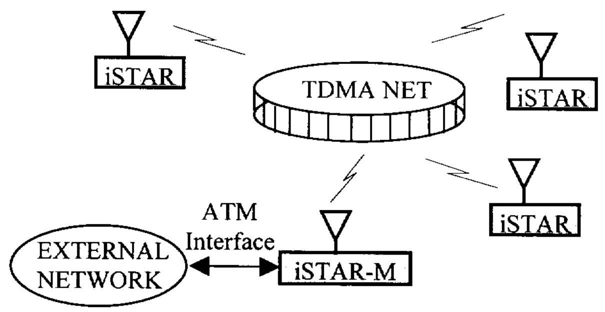


Figure 1: iSTAR Radio Network

Upto 60 iSTAR radios can participate in a radio network. The iSTAR-M radio is the gateway for the iSTAR radio network for accessing external networks. The iSTAR-M contains a 2 Mbit ATM (AAL1, AAL5, Q.2931 signalling) interface to tactical, strategic and PTT systems.

The iSTAR-M radio directs the mobile traffic into the external network and vice versa enabling the integration of iSTAR radio network with different types of military networks.

iSTAR CONFIGURATIONS:

There are three basic iSTAR radio configurations:

VRC-5110 : A vehicular radio set for mobile subscriber

PRC-5114 : A handheld radio for personal mobile use

VRC-5115 : A vehicular radio configuration for the gateway iSTAR-M

MAJOR TECHNICAL FEATURES :

Major technical features of the iSTAR system are summarized below:

- 225- 400 MHz Operating frequency band
- TDMA (Time Division Multiple Access) technique for channel access
- Use of Spread Spectrum techniques for EPM (Electronic Protection Measures) and LPI/LPD (Low Probability of Intercept / Detection)
- High speed burst transmission
- Rake Receiver and Forward Error Correction for improved reliability under multipath
- Distributed and Dynamic Routing
- Automatic relaying of voice and data
- 4.8 kbits/s CELP (Code Excited Linear Prediction) coded speech
- Near real time packet data communications with X.25 user data interface
- Position location (GPS) and network-wide distribution
- 2 Mbit ATM(AAL1, AAL5, Q.2931 signalling) interface to tactical, strategic and PTT systems.

iSTAR COMMUNICATION SERVICES:

iSTAR radios provide their users the following communication services:

- 4.8 Kbits/s CELP Coded Voice
- Asynchronous Data
- Synchronous Data (including 64 Kbits/s video)
- X.25 Packet Data
- Near Real-Time X.25 Packet Data

Voice and data services can be provided simultaneously. X.25 Packet Data services features ARQ (Automatic Repeat Request). Using the X.25 packet data service a single iSTAR radio can establish 32 different virtual circuits with other radio units simultaneously.

The iSTAR system provides near real time packet data communications and supports the X.25 user data interface. Near Real-Time X.25 Packet Data service is a low latency data service for sensor to weapon communications.

ISTAR SUPPLEMENTARY SERVICES:

The following supplementary services are supported by the iSTAR radio network:

- Calling Line Identification Presentation /Restriction
- Connected Line Identification presentation / Restriction
- Sub Addressing
- Call Transfer
- Call Forward
- Call Waiting
- Call Hold
- Camp on Busy
- Conference (Secure and Non-secure)
- Broadcast
- Closed User Group
- Priority & Pre-emption
- Non-secure Warning

iSTAR NETWORKING FEATURES:

iSTAR radio network, regardless of radio connectivity topology each iSTAR radio:

- Continuously updates a routing table indicating the connectivity paths to all the radios in the network

- Acts as a relay to any connection in the network while providing uninterrupted voice and data services to its own user.
- Establishes voice or near real time data communications with any other radio in the network using other units in the network as automatic relays.
- Provides simultaneous voice and data services to its user.
- Provides to its user the geographical position information of all the radios in the network.

ISTAR PROJECT STATUS:

The iSTAR development program which has been underway since 1996, produced the first engineering prototypes recently. The field trials of these prototypes are scheduled to start by the end of 1999.

Figure 2: First Engineering Prototype of the iSTAR radio

CONCLUSION:

The modern battlefield will contain many sources of high speed data which need to communicate on the move. The tactical mobile radio communication systems of the future should respond to the near real time data communication needs of these sophisticated equipment as well as providing the required voice services to mobile users.

The existing tactical radio systems and concepts are insufficient to fulfill the future needs of the battlefield. Development of a new tactical radio system, iSTAR (Integrated Services Tactical Radio), has been underway to meet the future needs of the digitized battlefield. The iSTAR solution provides a single unique radio system to support the present and future needs of the tactical mobile users.

A Wireless Ad hoc Multihop Broadband Network with Quality of Service Support

Matthias Lott, Bernhard Walke
Communication Networks, Aachen University of Technology, D–52074 Aachen, Germany
E-Mail: {lott,walke}@comnets.rwth-aachen.de

Abstract— **In this paper a wireless ad hoc multihop network is described. The decentrally organized network is able to guarantee the bandwidth contracted to a connection in a hidden station environment by means of contention-free data transmission on real channel connections (RCCs) that are established and used for the duration when data have to be transmitted, and that are released otherwise. Protocols for the air interface of the proposed network that have been developed to support real-time oriented services with quality of service and that support the prioritized, quick re-establishment of real channel connections are described.**

The proposed algorithms efficiently exploits the available frequency spectrum and protects established links in a hidden and exposed station environment. Performance results for the ad hoc network with different connectivities are presented and indicate low delay and high utilization even for multihop operation.

Keywords—**ad hoc, multihop, self-organizing network, hidden station, exposed station, dynamic channel allocation, decentral control, quality of service, prioritized access**

I. Introduction

AD HOC wireless communication has been of interest for a long time in military tactical communication [1]. Besides, mobile ad hoc computing is becoming more attention in the commercial communication, too [2], [3]. This has been started with the standardisation of 802.11 [4], and HIPERLAN[1]/1 [5], [6], that rely on the MAC[2] protocols carrier sense multiple access (CSMA) with collision avoidance (CA) and EY-NPMA[3], respectively, to assign transmission capacity to competing stations within a radio cell. Improvements with respect to collision probability have been proposed in [7].

Current proposals for MAC protocols for wireless ATM[4] systems at 5.2 GHz with data rates up to 54 Mbit/s rely on dynamic slot assignment (DSA) TDMA/TDD[5] MAC schemes and the concept of nodes that centrally control the transmission of stations [8], [9], [10], [11]. Though, at the moment the standardisation of HIPERLAN/2 focuses on systems with central control represented by an access point (AP) to a fixed network and one-hop communication of wireless stations to the central node, the need for ad hoc systems, that operate in environments without an infrastructure, will become more important. One typical future application is the connection of fixed, movable or mobile stations to a fixed network access point across a so-called ad hoc network, in which stations themselves identify their current radio connectivity from their current location and the other active stations [12]. They calculate the route across intermediate (relay) stations over multiple one-hop connections (multihop connection) to the final destination station [13]. This application includes the replacement of an indoor local area network (LAN) by a wireless network. Other applications are in-house networks that connect portable electronic devices with plug-and-play features [14], mobile computing in conference sized ad hoc mobile environments, and radio LAN outdoors, e.g. for rescue missions.

Different from the protocols mentioned above the proposed network relies on self-organized stations with decentral control able to route forward (relay) packets received according to virtual channel connections (VCC) established beforehand, and on dynamic channel allocation (DCA) and the concept of real channel connections (RCCs) [15]. This approach combines the advantages of decentrally controlled channel access with reservation based collision-free transmission of packets in containers of a framed TDMA system in a hidden station environment.

Next to the general problem of hidden and exposed stations (sec. II), protocols for the air-interface developed for ad hoc networks are presented in section III. To guarantee quality of service the best as possible in wireless networks, decentral scheduling and connection admission control schemes are proposed in section IV and VI, respectively. Results for the proposed ad hoc network that have been derived by means of stochastic event-driven simulations and that demonstrate the ability to guarantee the QoS aimed at are summarized in section VII.

II. Hidden and Exposed Stations

Ad hoc radio networks are characterized by stations that are placed randomly over the area and built a network in a spontaneous manner. Thus, no radio coverage planning can be applied and in many cases not all of the terminals in the ad hoc network have direct radio contact to each other.

A. Hidden station

In the latter case a station S_1 forwarding a packet to some other station S_2 is unable to control the usage of the respective radio medium in the receive range of the station addressed (c.f. Fig. 1).

A so called hidden station S_3 with a distance to the transmitter, that is larger than the detection range R_{det}, might

[1] High Performance Radio Local Area Networks
[2] Medium Access Control
[3] Elimination Yield-Non-Preemptive medium access
[4] Asynchronous Transfer Mode
[5] Time Division Multiple Access/Time Division Duplexing

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

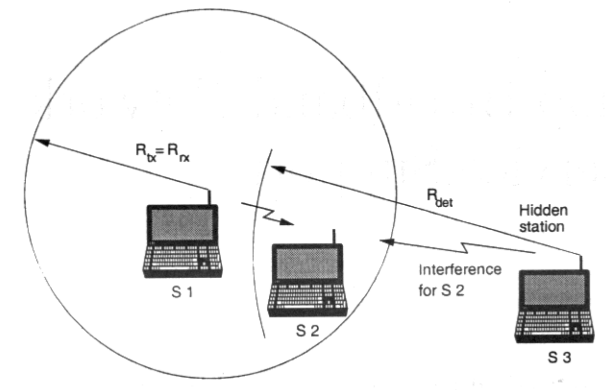


Fig. 1. Interferences for non-synchronized systems in a hidden station environment

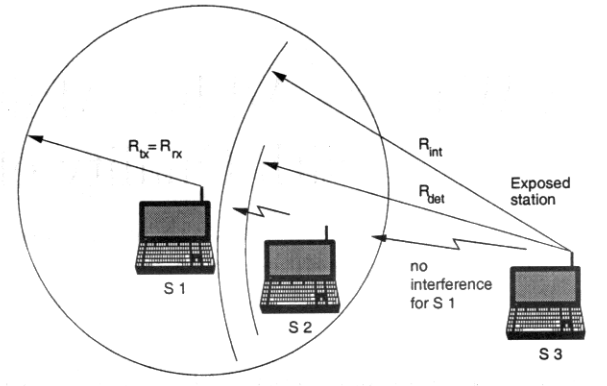


Fig. 2. Exposed station

cause interference at the receiver at the same time since neither the transmitter nor the interferer are aware of the transmission process of the other station.

Hidden stations affect the system throughput severely and tend to make a quality of service (QoS) guarantee impossible. The problem always arises when no means are available in a system to extrapolate from the presently observed spectrum occupancy to the future usage, e.g., of a time slot. Real channel connection (RCC) (see also sec. III-C) based systems are potentially better suited to guarantee an agreed transmission capacity in an ad hoc environment than systems based on dynamic slot assignment or random MAC protocols, since they operate with unpredictable random usage of the radio medium in the receive range of any node or station involved. An RCC is synchronously used in time-division duplexing (TDD) mode and the position in time of the potential interference energy is therefore known by any station within the detection range of the transmitter and receiver. Hidden stations are aware of the RCC in use and are forced to cooperate. The channel oriented communication has the advantage that an RCC measured by some station to have a too high signal strength will not be used by that station, since with a high probability it will find the channel still occupied in the near future.

B. Exposed station

Another problem in partially meshed networks that decreases the spectral efficiency are exposed stations, as shown in Fig. 2.

A station is called exposed if it resigns to transmit at some time to not interfere another communication relationship but in fact could communicate without disturbing the respective communication. E.g., a station S_3 detects a station S_2 transmitting to another station S_1 and defers from transmission to avoid a collision. But if the receiver S_1 is located outside the interference range R_{int} of the potential transmitter S_3, a simultaneous transmission of S_3 would cause no collision at station S_1. With the introduction of RCCs the stations will measure the received signal strength (RSSI[6]) of the two stations using an RCC and then are able to decide whether they are exposed stations or not.

[6]Received Signal Strength Indicator

C. Multihop operation

The aforementioned problems are typical in ad hoc networks with partial connectivity and where multihop connections have to be established to reach all stations within the network. Therefore, stations require a radio relay function implemented to extend the one-hop connection between two stations that have radio contact by another one-hop connection, and so on, to a multihop connection for an end-to-end relation between two stations without radio contact. Typically, only a small amount of hops is recommended to limit the end-to-end packet transmission delay that is a critical parameter for real-time oriented traffic.

In the following protocols for the air-interface are presented that are based on RCCs. They can cope with hidden stations and can make use of exposed stations to most efficiently use the spectrum. Because of independent operation of stations and their partial connectivity, each station has a different view of the interference situation. The proposed protocols are designed for self-organized stations with switching capabilities and use decentrally controlled medium access.

III. WANET Protocol Stack

The proposed wireless ad hoc network (WANET) is aimed to be able to support the QoS known from ATM fixed networks.

A. Frame structure

The physical channels result from dividing the given frequency band into FDM channels and from introducing TDM channels based on periodic time slots used in a TDD mode of operation, cf. Fig. 3.

A time slot carries a burst containing fields for settling time for automatic gain control (AGC), synchronization, error control (EC) and user data, and has some spare space called guard period to account for the propagation delay. The example TDMA frame comprises 16 slots each having a duration of 23.2 μs for a data rate of 20 Mbit/s and 448 bit

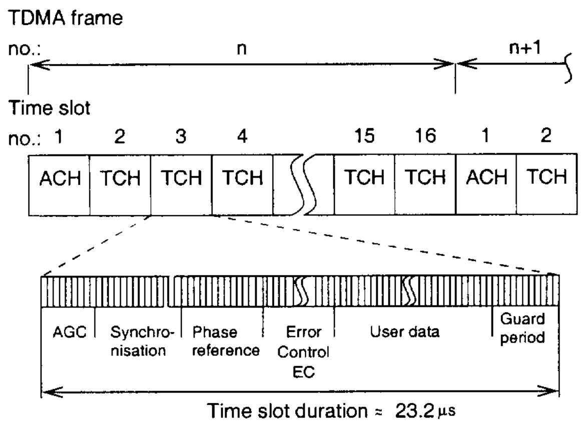


Fig. 3. Frame structure (370 μs)

carried in a slot. Each burst transmitted in a slot provides transmit capacity to the MAC protocol represented by a traffic channel (TCH). Random access with contention (using a key) is performed on the access channel (ACH). This channel is based on a slot structure that comprises a contention phase of 320 bit duration and signalling burst of 112 bit. The structure of the ACH is described in more detail in the following Section III-C that deals with the reservation of TCHs for one-hop connections.

B. Logical channels

Traffic channels (TCH), formed from a number of periodic slots, provide exclusive transmit capacity on the MAC layer for point-to-point communications as long as needed. A TCH x/y is using a number x of slots per frame, where y is the repetition period of these slots counted in frames. E.g., TCH 1/3 defines a physical channel capacity according to 1 slot/frame every third frame, whilst a TCH 3/1 uses three slots every frame, see Fig. 4.

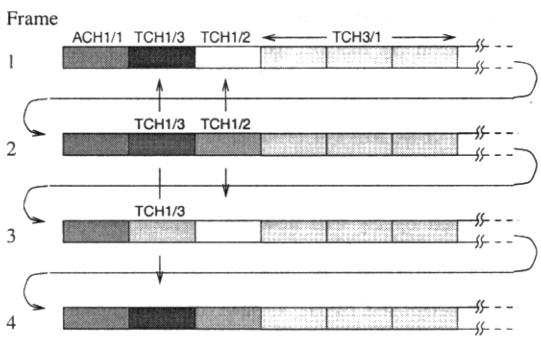


Fig. 4. Logical channel structure

By means of defining different slot-to-frame relations sub-multiplexing of the transmit capacity is possible. The smallest traffic channel capacity available is TCH 1/y where y is a design parameter that is chosen according to the traffic characteristic and QoS requirements. For the results in this paper $y = 1$ has been used. Any number of TCHs can be grouped to realize a so-called real channel connection (RCC).

An RCC is a layer 2 logical channel based on one or more TCHs. Its range of validity is for one hop only. The establishment of an RCC requires that a virtual channel connection (VCC) has been established before as a layer 3 end-to-end connection. Otherwise no QoS can be guaranteed and the system would be able to serve connectionless data transmission only, e.g. IP packets. A VCC might consist of a sequence of RCCs, each based on a different setting x/y of the related TCHs, e.g. TCH 1/1 for the first hop and TCH 2/2 for the second hop according to the available slots per hop. Different from a VCC an RCC reserves continuous physical transmit capacity during its (short) life time.

TCHs are used in a time-division duplexing mode of operation and the number of slots used forward and backward can be defined during establishment and dynamically during the data transfer phase (see also Section III-D.2). Access channels (ACHs) are provided separately in the TDMA-frame to allow stations to acquire TCHs under the control of a multiple access protocol.

All periodic time slots are supervised by the station's management system by periodically measuring their signal strengths and updating a local list of usable channels. An Organisation Channel (OCH) is used to exchange system connectivity information, parameters of the stations and network management system and radio resource related information. A TCH x/y is used to realize an OCH throughout the network. Depending on the number of slots per frame the parameters x and y are chosen to provide the required capacity for the updates.

C. Reservation of transmit capacity for one-hop connections

To establish an RCC between two stations an access (ACC) protocol data unit (PDU) is transmitted via the ACH.

The PDU is protected by a frame check sequence (FCS) to detect errors and contains an ID to identify the type of PDU, e.g. access control or network management, cf. Fig. 5, and the addresses of the transmitting and receiving stations of the one-hop connection. An abbreviated unique VCC identifier refers to a VCC that has been previously established (if the network does support QoS guarantee), and a local channel list contains the proposed channels that have been measured to be silent and can be used as TCHs by the responding station to acknowledge the requested RCC. Further, the PDU comprises signalling information of higher layer protocols, e.g. the QoS that informs the MAC layer about the throughput and delay requirements for the requested RCC.

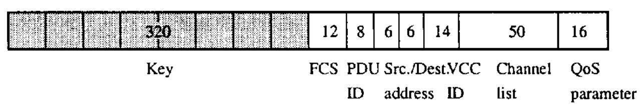


Fig. 5. Access PDU

The ACC PDU is preceded by a key, a code representing priority numbers. In Fig. 5 the key has 40 bit per code symbol and 8 binary symbols per key. A 'one' symbol is generated by transmitting an energy burst during the equivalence of 40 bit duration whereby the 'zero' symbol carries no energy and defines an equivalent listening duration. With the example key 2^8 sets of energy-burst-and-pause combinations, called keys can be distinguished. The lowest priority is represented by all zeros' and the highest priority of 255 is 'all ones' symbols. A station intending to transmit an ACC PDU uses its own key and listens during its symbol transmit pauses. If another station is heard transmitting the station defers from transmitting its ACC PDU in the current slot. This access protocol, CSMA/CA[7] is known from the ETSI/HIPERLAN/1 access protocol EY-NPMA [5]. The key length is a design parameter. CSMA/CA serves to reduce the collision probability of an ACC PDU (see also sec. VII).

A station, say S_1, that was not forced by keys of other stations to defer sends its ACC PDU to the destination (Dest) S_2, cf. Fig. 6, and all other stations in the transmit range with radius R_{tx} of station S_1 mark the proposed TCHs contained in that message as reserved for a time duration T_{res}.

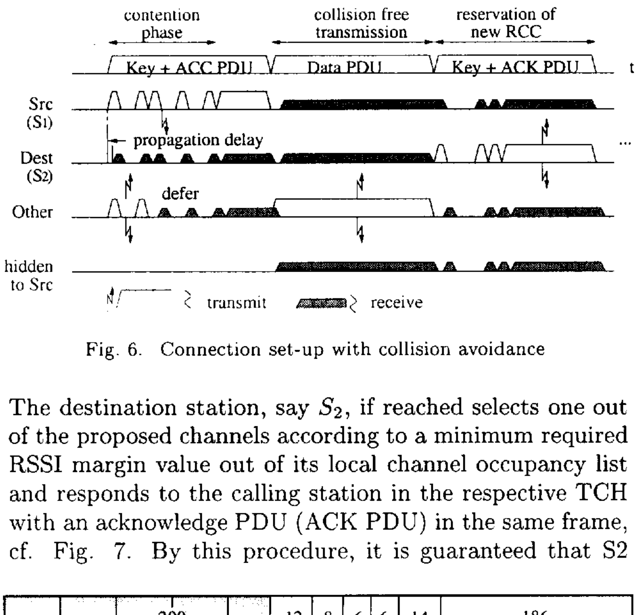


Fig. 6. Connection set-up with collision avoidance

The destination station, say S_2, if reached selects one out of the proposed channels according to a minimum required RSSI margin value out of its local channel occupancy list and responds to the calling station in the respective TCH with an acknowledge PDU (ACK PDU) in the same frame, cf. Fig. 7. By this procedure, it is guaranteed that S2

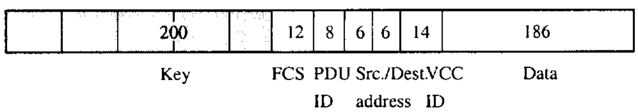


Fig. 7. Acknowledge PDU

will reach S_1 safely and vice versa with a high probability. (This procedure is similar to the RTS/CTS scheme used with IEEE 802.11 where a station transmits first a request-to-send (RTS) packet to the receiving station and waits for its acknowledge by a clear-to-send (CTS) packet.) Both PDUs are recognized by the stations within the detection range of S_1 and S_2 and all stations in the receive range of S_1 and S_2 are aware of the TCH being in use as soon as station S_1 is transmitting on the respective channel.

Besides management related information an updated channel list from station S_2 is transmitted within the data field of the ACK PDU. This provides station S_1 with information to identify other free channels for communication with station S_2.

The key contained in the ACK PDU of station S_2 serves to establish the TDD back-channel of station S_1 free of collisions. The key is able to eliminate the rare event that another station S_4 that is close to S_2 also has decided to transmit an ACK PDU in answer to a channel request of a station S_3, cf. Fig. 8.

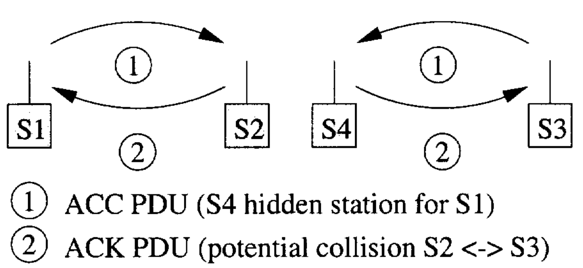


Fig. 8. Collision avoidance during RCC establishment

This might happen when a hidden station S_3 has successfully transmitted its access PDU to station S_4 at the same time as S_1. If S_4 is also a hidden station for S_1 or the ACC PDU of S_1 has been captured by the ACC PDU of S_3, S_4 will respond to the ACC PDU of S_3.

If station S_1 collides during its initial attempt or it will not receive an answer of S_2, it will repeat the ACC PDU later. The collision probability on the ACH can be guaranteed to be very low by dynamically defining the number of ACHs used. Nevertheless, a decentrally controlled number splitting algorithm [16] might be applied to prevent from potential instability and reduce RCC establishment delays. If station S_2 has no free channel available matching the channel set proposed by S_1, it will answer on the ACH and propose another set of channels to S_1. A positive, negative, or no answer of station S_2 will be recognized by the stations in the transmit range of S_2. All channels having been reserved temporarily for a duration T_{res}, but the one selected by S_2 will now be marked free internally in these stations. It might be that a station does not receive the answer of S_2 to S_1 and consequently will mark all the channels proposed by S_1 as free after T_{res} has expired. This does not severely harm the functioning of the network since these station soon will measure the allocated TCH in use in a TDD mode of operation. Thereby, all stations in the receive range of S_1 will detect its usage and release all the other channels reserved for the duration T_{res} and mark the used channel as occupied. The next hop, say from S_2 to S_3, is established in the same way. During the establishment of the next RCC on a route along a VCC, the capacity of a previous hop might be increased on demand

[7] Carrier-Sense Multiple Access / Collision Avoidance

by signalling the respective demand inband via the existing RCC.

For the procedure that has been described before it is assumed that on request of a new RCC at least one TCH can be found. To support QoS even for a highly loaded network where all TCHs are in use, established RCCs with low priority can be interrupted in favour of connections with higher priorities as described in Section IV-B.

D. *Data transfer*

After an RCC has been established, the data transfer is started. The packets that have in this design the size of the payload of an ATM cell, are transmitted transparently as payload of the data PDU, cf. Fig. 9, and are acknowledged on the backward channel.

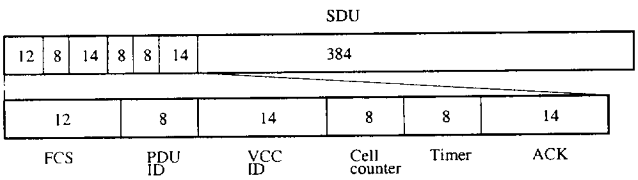


Fig. 9. Data PDU

Next to the cell counter of 8 bit and acknowledgements (ACK) of 14 bit for the ARQ protocol, the VCC-ID is transmitted within the PDU to allow statistical multiplexing of different VCC on the same RCC. Furthermore, the PDU comprises timing information (Timer) that determines the residual life time of the PDU and supports prioritized scheduling of urgent cells in intermediate stations of a route.

All other stations in the receive range of the respective stations will recognize and respect the occupancy of the channel through measuring the signal strength and a collision free transmission with guaranteed QoS becomes possible. Time division duplexing (TDD) on the same frequency is advantageous compared to frequency division duplexing (FDD), since all stations in the receive range of a station and node are able to detect the occupied channels without scanning the system frequency bands.

The number of time slots allocated to an RCC can be dynamically changed according to the actual needs of the communicating stations by transmitting a signalling PDU instead of a data PDU (see Fig. 10).

12	8	6	6	14	50	16	336
FCS	PDU ID	Src./Dest. address		VCC ID	Channel list	QoS parameter	Channel occupancy list

Fig. 10. Inband RCC signalling PDU

This PDU contains the same information as an ACC PDU and therefore can be used to increase the capacity of an active RCC or set-up a new RCC to a different destination station. Instead of the access key the current view of the station regarding the occupancy of the TCHs (e.g. reserved TCHs and respective RSSI values) are transmitted. This information is used by all receiving stations to update their routing tables and to temporarily reserve the proposed channels in the channel list. By this procedure, changes of the capacity of an one-hop connection can be performed in a very short time e.g., to allow to multiplex traffic of multiple VCCs to an RCC and to make quality of service re-negotiation very easy to introduce. This especially becomes important for real-time services with bursty traffic characteristic, e.g. video applications. One possible mode of operation is to keep continuously at least a TCH 1/y for an RCC and dynamically increase/reduce the capacity by opening/releasing parallel TCHs.

D.1 Error control

To guarantee a very low residual packet error probability the logical link control (LLC) layer applies an ARQ protocol on a per hop basis. The data PDU contains 14 bit for piggybacked acknowledgements and another byte for a sequence counter. If no data has to be transmitted on the reverse channel that can carry the ARQ information a signalling packet is transmitted instead that carries additional ARQ information and the channel occupancy list as described before to update the routing tables (cf. Fig. 11).

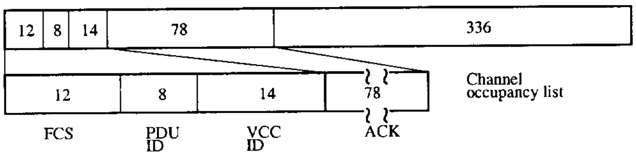


Fig. 11. ARQ signalling PDU

To improve efficiency of ACK reception, the multiple ACK protocol described in [17] can be used also.

D.2 Asymmetric traffic load

Under symmetric traffic load each slot is used alternating to carry forward and reverse traffic in a TDD mode of operation. To support asymmetric traffic flows, any relation of forward-to-reverse usage of a slot may be agreed per hop by the communicating stations, e.g. 8:1 would assign eight times the capacity forward, compared to backwards. To increase/decrease the relation of forward-to-reverse slots the signalling information will be transmitted over the RCC piggybacked with a data PDU. The receiving station will acknowledge the next higher/lower relation inband by the next packet. The resulting relations are given by

$$2^n : 1 \underset{\text{decrease}}{\overset{\text{increase}}{\rightleftarrows}} 2^{n+1} : 1, \qquad n \in [0;3]. \tag{1}$$

If a request for an increase of the forward-to-reverse relation would reduce the capacity of the backward channel below what is acceptable to the respective station a new TCH is reserved instead.

A threshold for the number of packets in a queue is used to decide whether a new TCH is needed instead a change of the forward-to-reverse relation. This guarantees that first the TDD mode is optimized to the asymmetric traffic and thus the reserved capacity is fully exploited.

E. Connection release of RCCs

Bursty traffic sources tend to use a channel as can be described by the packet train model [15], [18]. After an inactive phase of the source, a train of data PDUs is generated with the inter-car gap smaller than a specified maximum, referred to as the maximum allowed inter-car gap (MAIG). The inter-car gap is defined as the time between the end of a data PDU and the time of arrival of the following one.

Consecutive trains are separated by inter-train gaps. We propose to define a train so as to contain the MAIG as a parameter, controlled by the management system dependent on the current load of the network. The inter-train gap is a second parameter to control the life-time of an RCC and to define the MAIG. It is measured by the network management system. The MAIG might be defined service class specific.

An RCC is released either explicitly through a connect release message from one of the stations involved in a hop, or by a decentralized decision from the other stations observing that the MAIG has been exceeded so that the respective slots are marked as free in their local channel occupancy list.

To adapt the reservation of capacity to the varying load the number of ACHs is dynamically adjusted depending on the train length observed by the management system and is communicated via the OCH. With increasing train length the access intensity decreases and thus the number of ACHs can be decreased and the capacity for TCHs increased. A station is allowed to multiplex traffic of different applications and related VCCs to a given RCC to extend the train duration and reduce the number of accesses to the ACH.

F. Establishment of an end-to-end VCC

To establish an end-to-end virtual channel connection, an RCC will be requested with a new VCC-ID for an one-hop connection. The destination station can be retrieved from a station's local routing table that contains the one-hop next station for an end-to-end connection. If the RCC has been established, a connect request signalling PDU is transmitted to the one-hop (next) station that comprises the address of the end-to-end destination station. In case the addresses of the one-hop station and the end-to-end station are different, the station stores the new VCC-ID and the quality of service parameters and relays the connect request to the next station on the multihop connection. The destination station will respond with a connect confirm PDU that will be relayed by the intermediate station(s) to the source station and the new VCC is established then.

To protect the existing VCCs and to avoid the situation that the network will be overloaded, each station involved in the set-up procedure performs connection admission control, as explained in section VI, and rejects the connection request if sufficient capacity is not available.

IV. Decentral capacity assignment control for QoS guarantee

A. Prioritized access

To support both continuous and real-time variable bit rate (CBR, rt-VBR) services, which have stringent real-time requirements, opposed to other services (available bit rate, ABR, unspecified bit rate, UBR), which are much less sensitive to a time variance of throughput and delay, sets of keys for access PDUs representing different priorities are provided. The key contained in an access PDU will avoid collisions across service classes. For real-time oriented services a set with a high priority is assigned allowing the quick re-establishment of an RCC when the next packet train arrives. ABR and UBR service classes are assigned low priorities. If a station has been forced to defer during an access trial the key priority might be increased for the next attempt to improve the probability of success.

B. Requested release

It might be that a station S_1 having established a VCC but having released an RCC due to a pause in communication needs to re-establish an RCC to a station S_2 but is unable to name silent TCHs in its access PDU as candidate channels to be used. Station S_1 will then check all RCCs it is currently operating for low priority service classes and will select one and propose its TCHs for a new RCC to station S_2. To establish the new RCC to S_2 the respective TCH(s) between S_1 and the current neighbour station is(are) interrupted by S_1. In case no interruptable TCH is found by station S_1 the same procedure will be initiated by station S_2 when having received an access PDU from S_1 on the ACH. Station S_2 then will check all the RCCs it is currently operating with respect to the service classes they are supporting. In a responding PDU from S_2 transmitted on the interrupted TCH, station S_1 is informed about the new RCC to be used between S_1 and S_2. Trains with a higher (service class related) priority thus will be able to interrupt lower priority services on a per-hop basis. An interrupted RCC will be allowed after a service class specific delay T_s to try a re-establishment.

C. Forced release

To allow high priority services to find a free TCH whenever needed (re-establishment or increase of the number of TCHs for an active RCC), at least one TCH per hop is reserved for the service class with the highest priority.

When this channel will be accessed by a station with the respective priority, a station running a connection with a lower priority forced to release a TCH to always have one free TCH available. A similar approach is described in [19]. This requires to decentrally organize the forced releases of TCHs. As long as all stations involved have stored the service classes for each reserved TCH, the TCH with the lowest service class and with the longest life-time can be released. This approach is equivalent to a FIFO (first-in first-out) strategy for the lowest-priority service and tries to serve all stations in a fair manner by avoiding that one station continuously transmits its low-priority data while other stations can not find free TCHs or have to release the TCH after a short time in favour of requests of higher service classes.

But as long as a station is not aware of the services classes of the TCHs reserved before, because the respective stations are outside the decoding range, this station might be the only station that supports the lowest service class and therefore has to release its TCH when the last available TCH is used by a station with a higher priority. In this case, all stations that have no knowledge of the services classes of the other TCHs and support low priority services classes have to release one TCH. Still, the released TCH may not be usable for another station since its view of the interference situation is different.

From this discussion it becomes clear, that a combination of the forced-release and requested-release approach should be considered to support time delay sensitive service classes.

V. Power saving

Power saving is important in wireless networks with mobile stations due to their limited battery capacity. The protocol is able to support some percentage of mobile stations unable to continuously follow what is going on and possibly not willing to relay traffic of other stations. The latter has no impact on the functioning of the protocol but only on the meshing, since the connectivity of stations is then reduced. Stations that are not willing to transmit may switch to a sleeping mode. To enable mobile stations to be addressed and reached even when sleeping sometimes, they must select a nearby (fixed) station and inform it when during their next planning horizon they will power on and be ready to receive.

VI. Connection Admission Control

To be able to support real-time services with an appropriate QoS, these services are assumed to open a virtual channels connection (VCC) before transmitting their user data. Before opening a new connection the required capacity is therefore compared with the available capacity at the station. All VCCs operated at stations within the detection radius and the new VCC are taken into account and make up an equivalent capacity C_{new}. If this capacity is below the total capacity C_{total} that can be provided by the station involved, the new connection is accepted [20]. A new virtual end-to-end connection will be accepted if this procedure is successful on all the hops involved.

Non real-time services need not to apply for admission but will be interrupted whenever real-time services need the radio capacity occupied by them. Of cause it appears practical also to apply some admission rules for non real-time services to be able to guarantee them at least some minimum average throughput.

Since routing in the network is much more efficient due to short addresses when VCCs are used, we apply VCCs for all types of services in the WANET.

VII. Performance Results

The protocols described above have been analyzed by means of event-driven stochastic simulation. Different radio connectivities for 20 stations have been considered and the impact on the performance of end-to-end connections have been investigated. The connectivity of a radio network is defined by the mean number of stations reached in one radio hop, normalized by the number of all the other stations N,

$$c = \frac{1}{N \cdot (N-1)} \cdot \sum_{i=1}^{N} n_i, \tag{2}$$

where n_i is the number of neighbours to station i. For a fully meshed network the connectivity is 1.

Stations once for ever randomly select their final destination station and service class and establish the respective VCC. No signal fading is taken into account; the packet error ratio (for all PDUs) is set to 1 % instead. The stations are assumed to have identical source traffic behaviour and to generate symmetric traffic. Two traffic sources are modeled by a Poisson stream of packets with constant length either of 53 byte (source 1) or 1590 byte (source 2).

For the simulation study, the maximum allowed inter-cell gap is assumed to be two frames.

Figure 12 shows the payload throughput over the traffic load for networks with connectivity 1.0, 0.77, 0.57 and 0.5 for source 2.

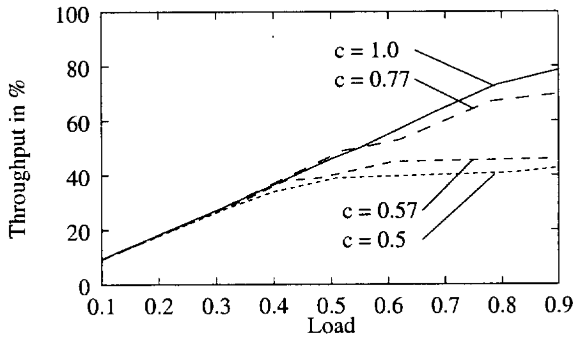


Fig. 12. Throughput vs. offered traffic for source 2

The offered traffic load results from the payload (ATM cells) only and is related to the available transmission capacity of 20 Mbit/s. Owing to the protocol overhead at the air interface the payload throughput is more limited the lower the connectivity is. With a connectivity of 0.5 the network is saturated for an offered traffic of 50 % of the transmission rate. A mean number of 1.9 hops per end-to-end connection for this connectivity has been observed so that the load is approx. twice the offered traffic.

Because of the long packets (source 2) and the large sojourn times, cell trains typically consist of a continuous stream of data PDUs and one-hop connections are released after each packet. Throughput linearly increases with load until saturation is reached.

The end-to-end mean cell delay of source 2 packets shown in Figure 13 is nearly constant until the network approaches saturation. The number of TCHs for each RCC have been increased dynamically with load in this simulation experiment to keep delays small. With loads approaching net-

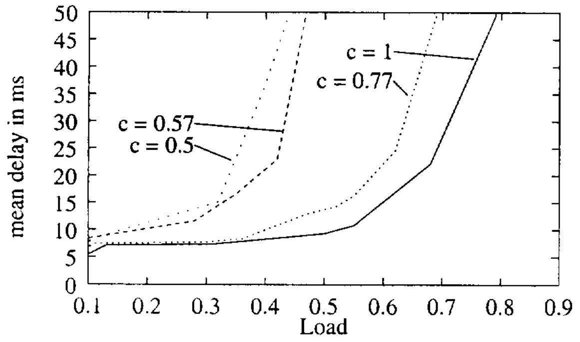


Fig. 13. End-to-end cell delay vs. offered traffic for source 2

work saturation the cell delay increases substantially as queueing in buffers is then dominating.

The delay can be reduced under small to medium load by allocating traffic channels based on more slots, e.g. TCH x/1, cf. Fig. 14.

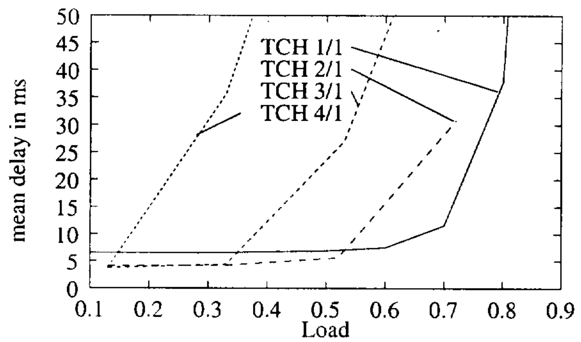


Fig. 14. End-to-end cell delay vs. offered traffic for source 2, c = 1.0

Using a TCH 4/1 for an RCC only 4 active RCCs can be supported simultaneously for a frame size of 16 slots. The mean delay under small load is reduced then, since arriving packets are served by a high transmit capacity in a short time.

The delay can be further decreased by increasing the number of ACHs enabling a quick reservation of TCHs especially for short train durations [15]. With increasing load the trains become longer and the number of re-establishments of RCCs decrease so that the number of ACHs can be reduced and the protocol will continue to operate in a stable condition.

VIII. Conclusion

A self-organized wireless ad hoc network supporting multihop operation with guaranteed QoS has been presented. Using channel connections based on a TDMA frame, contention free data transmission can be realized by so called real channel connections (RCCs). RCCs in combination with the proposed access control protocols provide high throughput and a small delay even when hidden and exposed stations exist.

Performance results indicate that the end-to-end mean delay of cells can be kept nearly constant until the network saturation is reached, when the capacity of TCHs for each RCC is increased dynamically with its load.

References

[1] Z.H. Haas and et al., "Milcom'97 Panel on "'Ad-Hoc Networks"'," in *IEEE MILCOM*, Monterey, California, Nov. 1997.

[2] P. Agrawal, "Special issue on Mobile Computing," *Wireless Personal Communications*, vol. 4, no. 2, Mar. 1997.

[3] I. Chlamtac, "Mobile Computing: When Mobility Meets Compuation," *IEEE Transactions on Computers*, vol. 46, no. 3, pp. 257–259, Mar. 1997.

[4] LAN/MAN Standard Committee of the IEEE Computer Society, " IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications ," Standard, IEEE, June 1997.

[5] ETSI, "Radio Equipment and Systems (RES); High PErformance Radio Local Networks (HIPERLANs) Type 1 Functional specification," Draft prETS 300 652, ETSI, Sophia Antipolis, France, 1996.

[6] Stephen K. Barton, "Special issue on the HIgh PErformance Radio Local Area Network (HIPERLAN)," *Wireless Personal Communications*, vol. 3, no. 4, 1996.

[7] J. Deng and Z. Haas, "Dual Busy Tone Multiple Access (DBTMA): A New Medium Access Control for Packet Radio Networks," in *IEEE International Conference on Universal Personal Communication (ICUPC)*, Florence, Italy, Oct. 1998, pp. 973–977.

[8] F. Bauchot and et al, "MASCARA, a MAC Protocol for Wireless ATM," in *ACTS Mobile Communications Summit*, Granada, Spain, Nov. 1996, pp. 647 - 651.

[9] D. Petras, A. Krämling, and A. Hettich, "MAC protocol for Wireless ATM: contention free versus contention based transmission of reservation requests," in *PIMRC'96*, Taipei, Taiwan, October 1996, pp. 903–907.

[10] Y. Du et al., "System architecture of a home wireless ATM network," in *IEEE ICUPC*, Sep 1996, pp. 477–481.

[11] M. J. Karol, Z. Liu, and K.Y. Eng, "An efficient demand-assignment multiple access protocol for wirelss packet (ATM) networks," *Wireless Networks*, pp. 267–279, 1995, J.C. Baltzer AG, Science Publishers.

[12] C-K Toh, *Wireless ATM and ad-hoc networks*, Kluwer Academic Publishers, 1997.

[13] B. Walke and R. Briechle, "A Local Cellular Radio Network (LCRN) for digital voice and data transmission at 60 GHz," in *Cellular & Mobile Communications*, Nov 1985, Online, pp. 215–225.

[14] J. Haartsen, "BLUETOOTH - The universal radio interface for ad hoc, wireless connectivity," Ericsson Review No. 3, 1998.

[15] B. Walke, S. Böhmer, and M. Lott, "Protocols for a Wireless-ATM Multihop Network," in *International Zurich Seminar*, February 1998, pp. 75–82.

[16] D. Bertsekas and R. Gallager, *Data Networks*, Prentice-Hall, 1992.

[17] V. Brass and B. Walke, "Inforamtion Distribution for Network Management in a Multihop Packet Radio Network," in *39th IEEE Vehicular Technology Conference (VTC'89)*, San Francisco, USA, Apr. 1989, pp. 793–798.

[18] R. Jain and S.A. Routhier, "Packet Trains- Measurments and a New Model for Computer Network Traffic," *IEEE Journal on Selected Areas in Communications*, vol. SAC-4, no. 6, pp. 986–995, September 1986.

[19] B. Xu and B. Walke, "Protocols and Algorithms Supporting QoS in an Ad-hoc Wireless ATM Multihop Network," in *Proc. of EPMCC'99*, Paris, France, Mar. 1999, pp. 79–84.

[20] M. Lott and B. Walke, "Performance analysis of a Wireless Ad hoc Network with QoS Support," in *Symposium on Performance Evaluation and Telecommunication Systems (SPECTS'99)*, Chicago, Illinois, July 1999, Accepted for publication.

A Connection Level Priority/Pre-emption Service for ATM Communication Networks

T.D. Johnson[1], Dr. G. Nourry[2], Eldon J. Mellaney[2], Dr. H. Rahman[3]

[1]TACOMS Post - 2000
International Project Office
18, rue du Docteur Zamenhof
92131 Issy-Les-Moulineux,
France

[2]Communications Research Centre Canada
P.O. Box 11490, Station "H"
Ottawa, Ontario, Canada
K2H 8S2

[3]Department of Electrical and Computer Engineering
Royal Military College of Canada
Kingston, Ontario, Canada
K7K 5L0

Abstract - The development of new services for ATM networks continues, yet one service still required is end-to-end Connection Level Priority/Pre-emption (CLPP). This paper details the CLPP requirement, presents factors affecting its implementation, and provides a solution for the implementation of this service in the control plane of the ATM protocol reference model. Also presented is an analysis of a subset of the results obtained from over three thousand CLPP OPNET simulations, which were continually refined over a three-year period.

1.0 INTRODUCTION

Current ATM standards do not support a Connection Level Priority/Pre-emption (CLPP) service. During periods of congestion, connection requests may be denied access to an ATM network. Depending on the importance of a call/connection request relative to other calls in the network, it may be desirable to admit the connection to the network at the expense of removing one or more other existing connections of lower importance. This statement summarises the concept of CLPP.

Although the exact implementation of the CLPP service and the reasons underlying its requirement differ, both the civilian and military sectors require ATM CLPP. For military purposes, CLPP is considered essential for combat effectiveness. Therefore, the addition of the service to future ATM-based defence networks is considered mandatory [1,2]. For civilian purposes, CLPP is required for time-critical applications (e.g. tele-medicine) and by ATM service providers who wish to implement pricing frameworks based on favoured network access.

2.0 DESIRED ATTRIBUTES OF CLPP

Irrespective of whether the user is military or civilian, the CLPP service must meet his different requirements while ensuring the service:

- pre-empts the minimum number of lowest priority connections required to establish a higher priority request, *only when the higher priority request has established an end-to-end ATM connection*;
- is effective in all network topologies;
- minimises the effect on standardisation to allow straightforward interconnection to public and private ATM networks that do not support the service;
- provides adequate network controls to minimise effects such as cascading and malicious use, while not severely restricting the service provided; and
- minimises connection set-up delay.

3.0 EXISTING WORK

To date, a number of contributions have been made to the ATM Forum's technical committee working groups on the matters of call re-routing, call priority, and connection-level priority and pre-emption.

It is important to note that re-routing and CLPP are two different services designed to meet different requirements. In the event of link or node failure, the re-routing service attempts to find an alternate path for an *existing* call that is affected by the failure. It cannot be used to allow a time-critical connection request guaranteed access to the ATM network during periods of congestion. Only a CLPP service can provide this functionality.

In terms of call priority, two main approaches have been proposed to the ATM Forum's technical committee. The first recommends the adoption of ITU-T Q.2959 [3] with a modification to support three priorities per connection: one for each of connection establishment, connection holding, and connection re-routing. Due to its complexity, the triple priority scheme would result in an increase in the connection set-up time. It is unclear whether this increase in complexity, and consequently connection set-up time, can be justified. In contrast, the second approach recommends the adoption of the call priority information element specified in ITU-T Q.2959 that indicates a single priority level for the connection – the priority of the connection with regard to the connection's establishment in the network. This is the same approach that is used in the proposed CLPP service presented in this paper.

In addition to the contributions made to the ATM Forum's technical committee on call re-routing and call priority, a few connection-level priority and pre-emption algorithms (CLPPA) have been proposed in the open literature [4,5]. A polynomial and exponential CLPPA are presented in [4]. However, neither is suitable, as the priority of the connections to pre-empt is considered only after one or two other optimisation criteria in the polynomial and

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

exponential algorithms, respectively. Similarly, the algorithm proposed in [5] is unsuitable due to its iterative nature which makes pre-empting a large number of connections too computationally expensive. As such, it is not a viable solution due to the delay it would impose on the connection set-up time and computational resources.

4.0 FACTORS AFFECTING IMPLEMENTATION OF CLPP

There are many factors that will determine if, how, and when the CLPP service is used. These factors are described in the following sections.

4.1 Priority Level Assignment

There are two basic considerations with respect to Priority Level Assignment: how many priority levels are required and how to assign them. The number of priority levels should be sufficient to meet the objectives behind the decision to implement the CLPP service. Nevertheless, it must be understood that as the number of levels increases, calculating which connections to pre-empt and controlling negative effects such as cascading become much more involved. Furthermore, as the number of levels increases, so does the probability of pre-emption for low priority connections.

In terms of assigning priority levels, there are two primary options: assignment to end terminals and assignment to users. In assigning priority levels to end terminals, each Virtual Path Connection (VPC) crossing a User-Network Interface (UNI) can be configured to restrict the amount of resources used and/or the number of connections at each priority level. Assuming network VPCs are permanent, configuration may be completed at network subscription. The semi-permanent VPC framework with the CLPP configuration is advantageous as it improves the ability to dimension the network to guarantee service levels to high priority connections. However, the disadvantage is that the service provided to the user is inflexible in that users are tied to the capabilities of the terminals. The alternative is to assign each user a maximum priority level and a maximum bandwidth resource. This approach offers the user improved service flexibility, but is much more complex to implement. Consider the possibility that one user generates several high priority connections on different terminals throughout the network (i.e. a sabotage attempt). Centralised network control is required to ensure that the service is not abused, requiring nontrivial network resources and an increased connection set-up delay. In addition, network dimensioning to ensure a guaranteed network performance is almost impossible.

4.2 Crankback

The routing mechanism called "crankback" is used to determine an alternate routing path after an attempt to make a connection through a previously selected route has failed [6]. The use of crankback will decrease, but not eliminate, call blocking in a network. The use of crankback can significantly increase connection set-up delay. Therefore, crankback be used successfully in conjunction with CLPP only where connection set-up time is not a critical factor. The combination of these two services allows the network to route around congestion points when possible (thus avoiding pre-emptions) and to use pre-emption to traverse the congested area when no alternate routes are available.

4.3 Network Topology and User Requirement

Network characteristics such as traffic load, traffic patterns, network size, and switch size, combined with user requirements, will determine how often the CLPP service is invoked. For instance, the user requirement may stipulate that crankback is not to be used, or, if it is, when it is to be abandoned in favour of CLPP. Additionally, the user requirement will determine how rapidly the network pre-empts connections. High priority connections may be regarded as a time critical. As such, when pre-emption occurs, the switch may pre-empt using a RELEASE COMPLETE (R.C.) message instead of a RELEASE (REL) message, allowing immediate pre-emption and recovery of network resources. To effectively optimise the CLPP service, characteristics about the expected network topology and user requirement must be incorporated into the ATM connection control procedures.

4.4 Routing and Call Admission Control

The routing and Call Admission Control (CAC) algorithms in conjunction determine, based on one or more metrics (e.g. additive delay) what path a connection request follows in the network to the requested destination. Clearly, the decisions these algorithms make will, in part, determine when CLPP is invoked. As such, it is important to ensure that the routing decisions are as accurate as possible. For example, if routing tables are not updated frequently, a connection request may be forwarded on a link with fewer resources than the tables indicate, which may result in unnecessary pre-emptions at downstream nodes. It is equally important to ensure the CAC algorithm is precise. In general, most CAC algorithms allocate more or less bandwidth than is required [7]. As such, in a network with CLPP, the network may pre-empt connections unnecessarily or accept the new connection when pre-emption should have been attempted.

4.5 *Cascading*

One of the main challenges with implementing a CLPP service is limiting cascading. Cascading refers to the situation in which an attempt to re-establish a connection that itself has been previously pre-empted results in pre-emption of one or more connections. If not controlled, a single pre-emption can cascade into a multitude of subsequent pre-emptions, severely disrupting communications in the network. In ATM networks, cascading may be controlled via proactive and reactive measures as outlined below. A fuller explanation of both approaches appears in [8].

4.6 *Proactive Cascade Control*

The algorithm used to determine which connections to pre-empt is known simply as the CLPP Algorithm (CLPPA). The CLPPA must know the priority level of all Virtual Channel Connections (VCCs) currently active in a switch. Assume that the ATM switch Management Information Base (MIB) is extended to include the priority level of each VCC for this purpose. If the CLPPA is designed to calculate the lowest priority connections to pre-empt, subject to certain criteria (e.g. sustainable cell rate), the probability of these connections pre-empting other connections during a re-establishment attempt is minimised. This is a proactive cascade control mechanism as it suppresses cascading before it begins. To implement this mechanism, the CLPPA must consider all possible pre-emption combinations on all potential destination routes. When the number of combinations is small, this approach is reasonable. However, in large switches, the number of search iterations will cause the connection set-up delay to rapidly exceed limitations. Clearly, some compromise must be reached between the complexity of the CLPPA and the effort to minimise cascading.

4.7 *Reactive Cascade Control*

If the CLPPA cannot control cascading proactively due to time constraints, the network (not the user) must control cascading in a reactive manner. The point of origin of the cascading effect occurs where a user tries to re-establish a pre-empted connection: at the UNI. Clearly, the most advantageous place to stop the cascading effect is where it originates. To achieve this, the local ATM switch must be able to recognise a call request as a re-establishment attempt. When a REL or R.C. message is received at a switch, it may analyse the Cause Information Element (IE). If the Cause IE indicates release due to pre-emption (a new Cause value is required) and the message destination is across a UNI (i.e. a local user), the switch may add the call to a list of locally pre-empted connections. To recognise a specific call, the list would contain the physical port number, the source and destination address, and the time the call was pre-empted. Assume that the switch deletes list entries after a predetermined time has elapsed from the pre-emption time. In this manner, when the switch receives a connection request from a local user, the request may be compared against the list. If there is a match, the switch *may* choose to reject the request as a re-establishment attempt.

4.8 *Malicious Use of CLPP*

With respect to malicious use of the CLPP service, there are two areas that need to be addressed. First, the network must ensure that users who are pre-empted do not try to raise the priority level of their requirement to try to re-establish their connection. The reactive cascade control mechanism ensures that users cannot simply raise their priority level to force re-establishment. Second, the network must protect itself from attempts to flood the network with high priority connections. As seen earlier, this is accomplished by restricting the maximum priority level and resources used by a user. Other security features for user authentication/affiliation, etc. are beyond the scope of this paper.

4.9 *Connection Requests without Priority Indication*

When a connection request is received from a network that does not support the CLPP service, the SETUP message will not indicate a requested priority. The network supporting CLPP must assign a priority to the connection request. The simplest solution is to assume that all calls originating outside the network are not as important as those originating inside and, therefore, are assigned the lowest priority. A more sophisticated approach is to provide a gateway function that compares the source address or user authentication code, contained in a higher layer protocol data unit, to a predefined list to determine what priority level should be assigned to the connection request.

5.0 CLPP SERVICE ADDITION TO ATM CONTROL PLANE

This section presents an overview of the proposed CLPP service implementation that is simply referred to as the ATM Control Plane Implementation. This implementation is an extension to the standard ATM protocol. Specifically, the extension occurs in the control plane of the ATM Protocol Reference Model (PRM). It is assumed that the ATM-based network uses a dynamic link state routing method (e.g. Open Shortest-Path First).

5.1 *CLPPA – Proactive VPC Search Mechanism*

The concept of the CLPPA was introduced in the section on proactive cascade control. It is the heart of the CLPP system. In short, the CLPPA must, for each VPC on a route towards the destination, determine if there are enough lower priority VCCs on the VPC to allow the higher priority request access to the network. The algorithm must also determine the minimum number of lowest priority VCCs to pre-empt. This must be accomplished without overly increasing the call set-up delay. For the proposed implementation, a mechanism called the Proactive VPC Search was developed.

In addition to extending the MIB to include the priority level of each VCC, the information regarding each VPC is extended to include the number of VCCs and the sum of their Peak Cell Rates (PCRs) and Sustainable Cell Rates (SCRs) at each priority level (i.e. all information required by the CAC algorithm). These MIB fields must be updated every time a VCC is accepted or released.

When routing and CAC refuse a connection request, the CLPPA is invoked. It searches all the suitable VPCs in the order they are ranked in the routing table. For each VPC, the average PCR and SCR of its VCCs are calculated for each priority level. An *average* VCC is therefore defined as a VCC on a given VPC with the average PCR for the VPC. Starting at the lowest priority, the CLPPA determines how many *average* VCCs would have to be pre-empted at each priority level in order to accept the connection request. If the priority level of the request is less than or equal to the highest priority (highest priority level is numerically smallest) that would have to be pre-empted, the VPC is excluded. This process is completed for each VPC. When the search is complete, the VPC that requires the pre-emption of the smallest number of *average*, lowest priority connections is selected. Once the VPC is chosen, the SETUP message is immediately forwarded to the next switch. At this point the CLPPA determines and *marks* for pre-emption the VCCs to be pre-empted on the selected VPC, such that the number is minimised. It is important to note that no pre-emption has yet occurred.

Clearly, all VCCs on the VPC with a priority level lower than the minimum priority level to be pre-empted are marked for pre-emption. For the VCCs with a priority level equal to the minimum priority level to be pre-empted there are two possibilities. If the number of average VCCs to pre-empt equals the total number active, then all VCCs are to be marked and no further determination is required. If this is not the case, the CLPPA performs a recursive combination search algorithm that tries all combinations of VCCs up to a specified depth. At the end of the search, the combination that provides the least amount of wasted resources is chosen and marked for pre-emption. To limit the search time, the depth is limited to the number of average VCCs required to be pre-empted. While this may not always yield the best combination, it can significantly limit the number of search iterations required. The number of possible combinations of the restricted search is given by Equation 1.

$$N_c = \sum_{i=1}^{A_{vcc}} \binom{N_{vcc}}{i} \qquad \text{Equation 1}$$

where A_{vcc} is the average number of VCCs required to pre-empt at the highest priority level to be pre-empted, and N_{vcc} is the number of VCCs at that highest priority level on the VPC.

When $A_{vcc} \ll N_{vcc}$, the restricted search is many orders of magnitude less than the unrestricted one. Moreover, when the traffic on the VPC is fairly homogeneous, with high probability the amount of resources between the optimal selection and the actual selection made is minimal. Finally, limiting the search depth to the number of average VCCs will always produce a combination that provides the required resources to complete the pre-emption. Therefore, given the limited ill effect of minimising the search depth, this approach was taken.

By using the Proactive VPC Search method, each switch only has to determine the VPC on which the pre-emption will occur before forwarding the SETUP message. This significantly reduces the end-to-end call set-up delay and still determines a VCC pre-emption combination for the VPC that will minimise the cascading effect.

5.2 *System Overview*

When an ATM switch receives a SETUP message, all the information required to process the message, including the requested priority level, is contained within. As per current standards, the switch will perform routing and CAC for routes with the appropriate Quality of Service (QoS) to the desired destination. If the connection is accepted, the SETUP message is forwarded to the next switch along the chosen route to the destination. However, if routing and CAC determine that there are not enough resources for the connection, the switch initiates the CLPP system.

First, the system ensures that the request is not a re-establishment attempt. If it is not, the proactive VPC search is used to determine if pre-emption is possible. If pre-emption is possible, the SETUP message is forwarded to the next switch and the CLPPA then and marks which VCCs are to be pre-empted on the chosen VPC. At this point the switch waits for four possible situations with respect to the connection request:

- a CONNECT message is received;
- a RELEASE message is received;
- a new SETUP message is received; or
- second Expiry of timer T303 or first expiry of T310.

A logic diagram of the operation of the CLPP service for these various events is given in Fig. 1.

5.2.1 *CONNECT Message Received*

There may be a relatively long period of time between the forwarding of the SETUP and the arrival of the CONNECT message. During this time resources may have been released on the VPC chosen for the pre-emption (i.e. a VCC not marked for pre-emption was released). If resources have been freed, the CLPPA determines if pre-emption is still required, and, if so, whether enough resources have been released such that fewer VCCs have to be pre-empted. If pre-emption is no longer required, all marked VCCs are cleared and no pre-emption occurs. If sufficient resources have been released so that not all of the marked VCCs have to be pre-empted, the CLPPA ascertains which of the marked VCCs will be pre-empted; otherwise, the VCCs marked originally are pre-empted.

If pre-emption must occur and time is not a critical factor (determined from the priority level of the request), the switch will send a RELEASE message towards the source and destination of each VCC marked for pre-emption (Fig. 2a). The RELEASE message Cause IE indicates that the connection is being released due to pre-emption (This requires a new value for the Cause IE for the RELEASE and RELEASE COMPLETE messages). This information is passed to the upper layers and provides a mechanism to warn the user that he is being pre-empted. On receiving a RELEASE COMPLETE from all the pre-empted connections, or the expiration of the network timer(s) (in which case a RELEASE COMPLETE is sent to pre-empted connections which have not responded), the CONNECT message is forwarded to the source.

If pre-emption must occur and time is critical, the switch sends a RELEASE COMPLETE to the source(s) of the VCCs marked for pre-emption (Fig. 2b). This permits the switch to immediately recover the resources from the pre-empted connections and instantly forward the CONNECT message to the source. The pre-empted users are given no warning of the pre-emption, but the reason the connection was released may be determined from the Cause IE.

5.2.2 *RELEASE or RELEASE COMPLETE Message Received*

If a RELEASE message is received from the upstream switch, the destination, or the source, it is forwarded either toward the source (Fig. 3a) or destination (Fig. 3b), as appropriate, and the VCCs marked for pre-emption are cleared.

5.2.3 *New SETUP Message Received*

When a new SETUP message is received (indicating a new connection request), a parallel call control process is initiated. If the new request is a re-establishment attempt, it is rejected. If it is not and pre-emption is required, the CLPPA is initiated. As before, if pre-emption is possible, the SETUP message is forwarded on the VPC chosen for the pre-emption and the CLPPA determines which VCCs are to be marked for pre-emption.

Should the new request mark for pre-emption the request currently awaiting a response from the network (i.e. the original request), a RELEASE message is immediately sent toward the source and destination of the original request and the VCCs marked by the original request are cleared. Only then will the CLPPA determine which VCCs are to be marked for the new request. The release of the original request and the subsequent re-determination is completed for two reasons. First, there is a possibility that there were more available resources on the VPC when the original request was received. When the original request is released, these resources are returned to the VPC and, as a result, are now available for the second, higher priority request. Consequently, the CLPPA may now only need to mark a smaller number of low priority VCCs for pre-emption. Second, the VCCs marked by the original request will always have a lower priority than the minimum priority required to be pre-empted by the new request. Therefore, these connections must be the first considered for pre-emption (i.e. marked first) if the integrity of the CLPP system is to be maintained.

5.2.4 *Timer Expiry*

If no response is received from the upstream switch within a certain timeframe (i.e. timer T310 or T303 timers expire), the switch clears the current request by sending a RELEASE message towards the source and destination and all marked VCCs are cleared.

5.3 *Effect on Standardisation*

The addition of a new service to a standardised protocol must endeavour to minimise its effect on the standard. The addition of the proposed CLPP mechanism is simplified given that steps have been taken for its introduction (e.g. ITU-T Q.2959) although no mechanism to implement the service has been chosen to date. Specifically, a new non-mandatory IE has been added to the SETUP message [3]. The IE will contain, among other things, the requested priority level of the connection request. With bit 4 of the IE Action Indicator set, the CLPP information can be transparently passed through networks that do not support the service. Irrespective of this planning, other changes are still required. To implement the control plane CLPP service, the following changes are required to the ATM standards:

- The addition of a new cause value to indicate network release due to pre-emption will be added to the RELEASE and RELEASE COMPLETE messages. When a standard network finds a cause value in these messages that it does not recognise, it assumes a cause value of 31: "normal, unspecified." Therefore, the addition of the new cause IE will not affect the operation of standard ATM networks, but will allow networks with the service to identify when a connection has been released due to pre-emption.

- The CLPPA algorithm must be added to private ATM switches. This will include augmenting many of the states in ITU-T Q.2931 [9] and the control plane software accordingly. This will require co-operation from the ATM switch vendors, but does not require any hardware modifications.
- The MIB must be modified to include the changes outlined in section 5.1. Again, this will require co-operation with the switch manufacturer. The additions to the MIB will not prevent standard User Management Entities (UMEs) from obtaining required information.

5.4 *Key Benefits of ATM Control Plane CLPP*

The key benefits of implementing the CLPP service in the control plane of the ATM PRM are:

a) Any CLPP service added to the ATM PRM must make use of the current call control procedures. As such, the extension of these procedures to include a CLPP service is a natural evolution;

b) The implementation meets the stated requirements of section 2.0 in that it:

- pre-empts the minimum number of lower priority connection required to establish a higher priority request, *only after the higher priority request has established an end-to-end ATM connection*;
- eliminates network effects such as cascading, while still allowing pre-empted connections to be re-established if sufficient resources are available so that re-establishment can be completed without pre-emption; and,
- minimises connection set-up delay through the use of the Proactive VPC search mechanism

c) The control plane CLPP service is transparent to ATM public or private networks that do not support the service;

d) The control plane can readily communicate with higher layers during call/connection control. Therefore, information from higher layers required to implement the service may be easily accessed (i.e. passing connection priority information). Similarly, information regarding pre-emption may also be delivered to the control processes of higher layers (e.g. the ISO OSI has 16 priority levels in TP 0-4); and,

e) The control plane CLPP service is implemented in software. Therefore, any change/upgrades required by the network with regard to CLPP can be achieved without purchasing new hardware.

6.0 CHARACTERIZATION OF ATM CONTROL PLANE CLPP THROUGH MODELLING AND SIMULATION

To test, refine, and characterise the proposed CLPP service, it was modelled and simulated with OPNET (OPtimized Network Engineering Tool). The CLPP system was added to the standard ATM Model Suite. This included the modification of the ATM MIB structure (to allow the use of the Proactive VPC search algorithm) and the network and user control states of ITU-T Q.2931 to add the CLPPA.

Over 3000 simulations have been completed since the models were first produced. The functionality of the CLPP was verified. The CLPP model functionality has also been independently tested and verified. Due to space limitations, only a small subset of the simulation results is presented here.

6.1 *Modelling and Simulation Methodology*

The aim of the simulations was to determine the effect of the CLPP system on end-to-end connection set-up delay and to determine the call blocking probability for each priority level. To characterise the proposed CLPP service, it has been simulated in three network topologies for 4, 8, and 16 priority levels, using various call priority distributions, with and without crankback. Each network was simulated without a CLPP system, with and without crankback, to provide a basis for comparison.

In each simulation, any traffic source generating a call to the network selects its destination uniformly. The average call duration and the average call inter-arrival time are set to 7 s and 1 s, respectively. In all simulations, the CAC algorithm uses the Peak Cell Rate (PCR) as it provides the worst case scenario and pre-emption is completed via a RELEASE COMPLETE message. To effect network congestion, and thereby evaluate the effectiveness of the proposed CLPP service, a traffic scaling factor is employed: in any given simulation, the minimum and maximum possible PCR of a connection request are determined by the value of the load factor. In each simulation completed, the maximum PCR is 12 000 cell/s when load factor is 1.0. To ensure statistically valid results, each parameterised simulation is run with nine distinct simulation seeds. In all cases, a statistical result presented represents the mean value of that statistic and is always presented with the standard deviation computed for that mean value. The simulation data presented below captured the behaviour of the network over a 1200 s period, which is more than sufficiently long to obtain statistically valid results.

7.0 SIMULATION RESULTS

The simulation results for the network topology shown in Fig. 4 and Fig. 5 will be presented and analysed. In each simulation, eight priority levels are used. All links in the network have a capacity of 9000 cell/s or 3.816 Mb/s. The average time required to execute the CLPPA at a switch, as computed in the OPNET simulation running on a general-purpose workstation, is approximately 50 μs. Thus, the additional time imposed on the connection establishment by the CLPP service is negligible.

7.1 *Uniform Call Priority Distribution Function*

In this scenario, the effectiveness of the proposed CLPP service is evaluated for the network when the priority level assigned to a call is selected from a uniform distribution. Two cases are considered: i) crankback is not employed in the network; and, ii) crankback is employed in the network, with a maximum of two re-route attempts possible at a given switch.

7.1.1 *No Crankback*

A plot of the call blocking probabilities, by priority level, versus load factor clearly shows that the proposed CLPP service ensures that high priority calls are given preferential treatment over low priority calls (Fig. 6). For load factors less than 0.3 there is very little congestion in the network. Accordingly, the blocking probability of all priorities is near 0. As the load factor increases, so does network congestion, causing the CLPP system to be invoked. The result of the addition of the CLPP system is clearly reflected in the plot of the call blocking probabilities. As expected, the blocking probability for priority level 0 (highest) calls increases slower than those for lower priority calls. For example, at load factor 0.5, the absolute difference in the blocking probabilities for priority level 0 and 7 calls is 24.31% and 2.28% for priority 0 and 1 calls. When load factor is 1.0 (network is very congested), the absolute difference between the blocking probabilities for priority 0 calls and those of priority 7 and priority 1 calls is 18.88% and 8.3%, respectively. Overall, the effectiveness of the CLPP system is clearly demonstrated by the graph.

7.1.2 *With Crankback*

The use of the CLPP service on its own reduces the blocking probabilities for high priority calls during network congestion (Fig. 6); however, when crankback is employed in conjunction with the proposed CLPP service, additional reductions in the blocking probabilities for high priority calls are realised (Fig. 7).

At load factor 0.5, the blocking probability for priority level 0 calls is 12.03% without crankback and 9.02% with crankback. Similarly, for the lowest priority, the blocking probability at load factor 0.5 is 38.18% without crankback and 33.33% with crankback. Therefore, the use of crankback resulted in a 25.0% and a 12.7% reduction in the call blocking probabilities for priority level 0 and 7 calls, respectively. When load factor is 1.0, i.e. much higher levels of network congestion, the use of crankback decreases the blocking probability for priority level 0 and 7 calls by 13.3% and 1.9%, respectively.

As expected the use of crankback with the CLPP service reduces the blocking probabilities for all calls in the network. The most significant gains occur for higher priority calls when network congestion is relatively low, i.e. load factor less than 0.4. These gains result from routing around congestion in the network, when possible, and pre-empting connections only as required. The trade-off for the lower blocking probabilities is an increased call set-up delay.

7.2 *Incline Call Priority Distribution Function*

In this simulation scenario, the priority of a call is determined from an "incline" probability distribution function (PDF). The incline PDF results in relatively few high priority connection requests made to the network compared to the number of lower priority ones. This call priority PDF more accurately reflects the actual distribution of priorities in a network that would employ a CLPP service. In this simulation, crankback is used in conjunction with the proposed CLPP service.

Fig. 8 shows the blocking probabilities for the calls in the network. Clearly, compared to the previous two scenarios, the absolute (and percentage) differences in the blocking probabilities of high and low priority calls are greater. This improvement is attributed to the fact that there is a significant increase in the ratio of low to high priority connections in the network. As a result, the probability increases that a high priority request can be accommodated in the network by either re-routing around congestion points or by pre-empting lower priority calls.

8.0 CONCLUSION

A Connection Level Priority and Pre-emption (CLPP) service for ATM-based networks has yet to be defined. Many network, user, and technical factors must be considered if the CLPP service is to meet the user requirements. The proposed ATM Control Plane CLPP implementation meets these requirements through unique techniques like the Proactive VPC search mechanism. The proposed service has been extensively tested and characterised through detailed modelling and simulation in OPNET. The simulation results conclusively demonstrate the effectiveness of the ATM Control Plane CLPP system. Moreover, the recent addition of a Call Priority IE to the

SETUP message by the ATM Forum means that this system may be added to the ATM standards with minimal disruption.

REFERENCES

[1] Hardich, W., Reinhold, W., and Schade, H. "Analysis of a Requirement-Oriented, Integrated Services Tactical Telecommunications Network of BW/NATO POST 2000". ALCATEL. March 1994.

[2] TTCP STP - 8 Workshop, "Technical Challenges arising from the Interconnection of ATM Networks via Radio Links". October 14, 1993. p 5.

[3] ITU-T Recommendation Q.2959. "Digital subscriber signalling system No. 2 – Call priority". July 1996.

[4] Peyravian, M. and Kshemkalyani, A. D. "Connection pre-emption: issues, algorithms, and a simulation study". Proceedings of 1997 IEEE International conference on computer communications (INFOCOM '97). Kobe, Japan. April 7-12, 1998. Vol 1. pp 143-51.

[5] Poretsky, S. and Gannon, T. "An algorithm for connection precedence and pre-emption in Asynchronous Transfer Mode networks". 1998 International Conference on Communications (ICC '98). Atlanta, Georgia. June 7-11, 1998. pp 299-303.

[6] ATM Forum. ATM User-Network Interface Specification. Prentice Hall. Toronto. 1995.

[7] Onvural, R.O. Asynchronous Transfer Mode Networks: Performance Issues. Artech House. Boston. 1995. pp 329-367.

[8] Johnson, T.D. Connection Level Priority/Pre-emption service for Asynchronous Transfer Mode Communication Networks. MEng. Thesis. Royal Military College of Canada. Kingston, Ontario, Canada. March 1997.

[9] ITU-T Recommendation Q.2931. "Broadband Integrated Services Digital Network (B-ISDN) – Digital subscriber signalling system No. 2 (DSS2) – User-network interface (UNI) Layer 3 specification for basic call/connection control". February 1995.

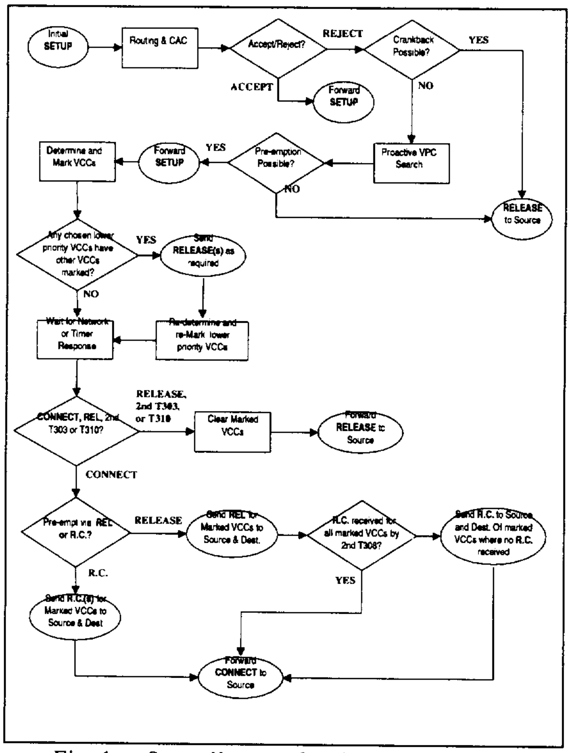


Fig. 1 State diagram for the CLPP service

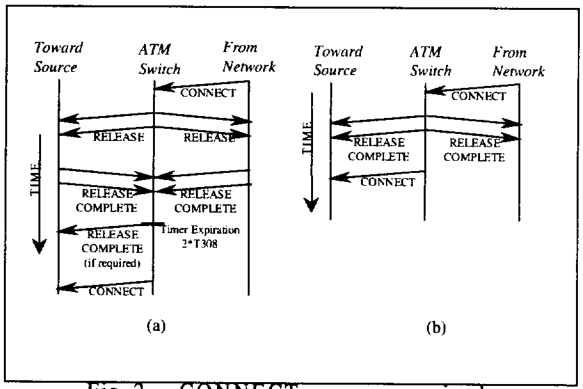


Fig. 2 CONNECT message received.

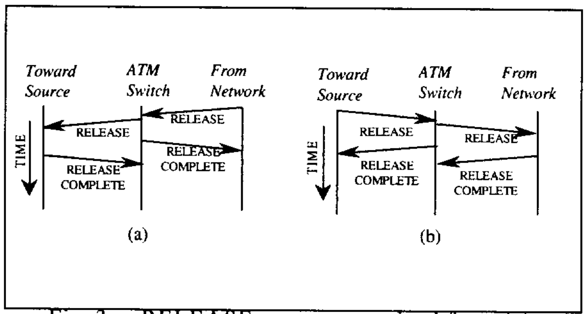


Fig. 3 RELEASE message received from (a) destination and (b) source

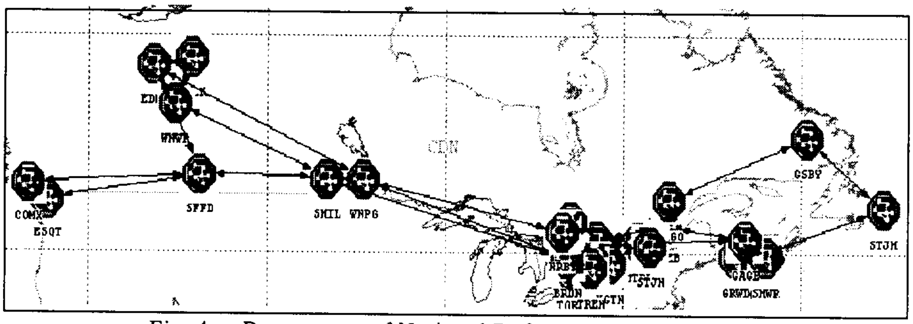


Fig. 4 Department of National Defence strategic network

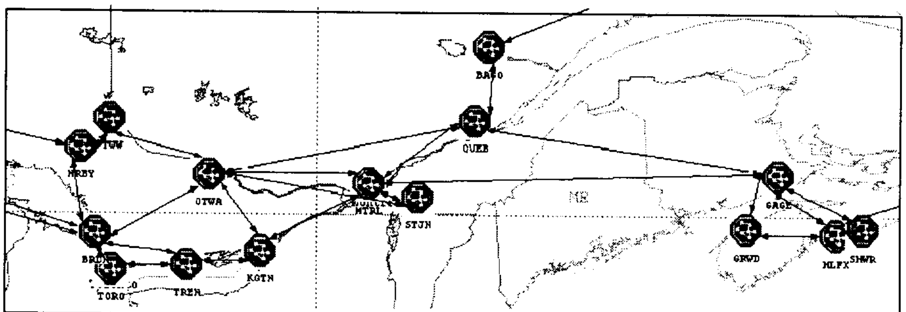


Fig. 5 Department of National Defence strategic network (Central and Eastern Canada)

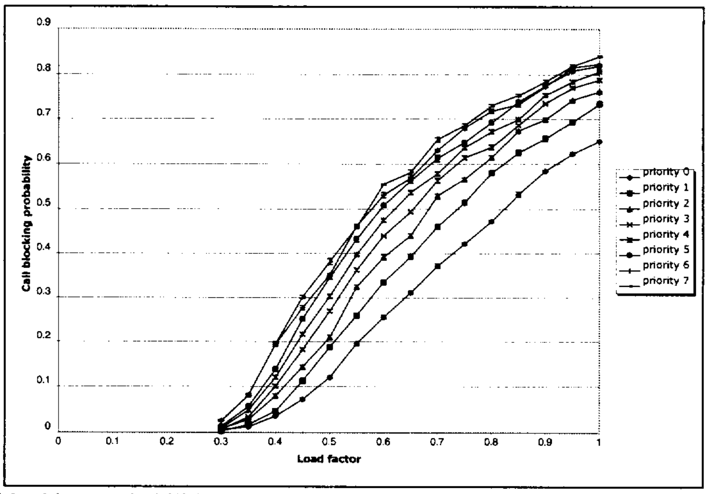


Fig. 6 Call blocking probabilities vs. load factor – uniform priority distribution, no crankback.

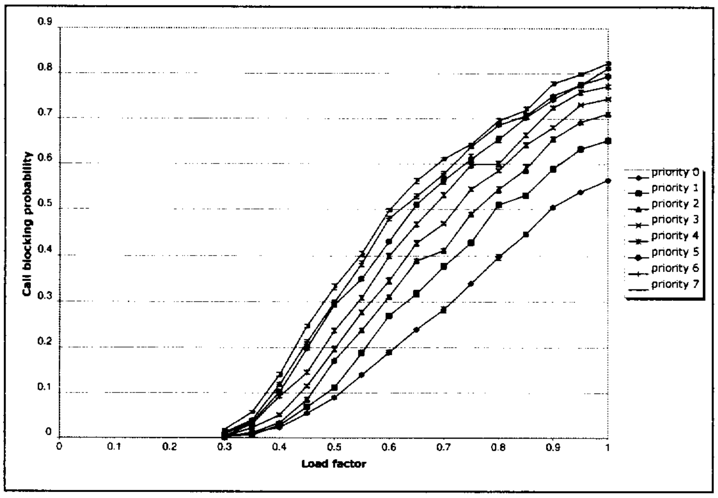


Fig. 7 Call blocking probabilities vs. load factor – uniform priority distribution, max. two crankback attempts per switch.

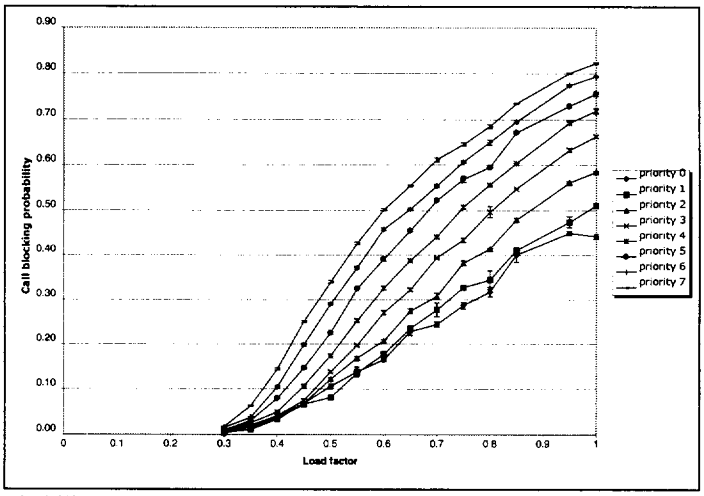


Fig. 8 Call blocking probabilities vs. load factor – "Incline" priority distribution, max. two crankback attempts per switch.

Mobility Management for Tactical ATM Networks

(June 1999)

P. Walker
Marconi Research Centre
Elstree Way, Borehamwood,
Hertfordshire WD6 1RX, England.

This paper evaluates the current and proposed ATM standards for COTS equipment against tactical network requirements. It presents a set of requirements and describes some of the schemes under consideration in the ATM Forum to implement mobility. It concludes that the emerging ATM standards that enable terminal mobility are worthy of consideration for tactical networks, as are those for switch mobility in a non-operational state. However, some features, such as operational switch mobility and connection pre-emption, require further standards development.

1. Introduction

COTS technology

Commercial-Off-The-Shelf (COTS) equipment is, by definition, ready to buy and use. Ranging from small equipment to complete systems, COTS products are developed for a broad (commercial) market, and often benefit from massive R & D investment. The economies of scale result in a high performance-to-cost ratio, though users have to accept that such a product is not tailored to their own individual requirements.

Hence a challenge for the military is to be able to exploit COTS. However, defence organisations are neither a typical nor dominant COTS customer – there are special requirements and COTS equipment may not be able to satisfy these directly. Therefore it is important to investigate whether COTS equipment can be modified or extended in order to address any shortcomings while taking advantage of the high performance-cost ratio.

ATM

ATM has received considerable attention worldwide as a technology that offers a highly flexible networking service. It is scaleable and the Quality of Service (QoS) options make it appropriate for voice, video, data and other applications such as control.

ATM's QoS options are facilitated because ATM is *connection-oriented*, where users intercommunicate over dedicated connections. In ATM these are virtual rather than physical. They are typically set up on demand by signalling software. The Virtual Connections (VCs) are multiplexed for transmission between nodes by switch logic.

Bringing ATM to a market of varied applications requires much cross-vendor standardisation of interfaces, protocols and frameworks. This role is currently performed by the ATM Forum and the ITU. Many standards are in place to support static ATM networks but the standardisation process continues as the industry seeks to satisfy ever more complex network requirements.

Attention is now moving towards the use of radio as a bearer. The benefit of radio in LAN applications is mobility and wiring freedom (thereby decreasing installation costs) and the main motivator in WANs is the phenomenal success of technologies such as GSM for mobile telephony. There is an increasing need to offer more flexible services than GSM can currently support, e.g. high quality video services to ambulances. However, the high error rates inherent in radio systems are the antithesis of the original ATM drivers, so there is now much activity investigating solutions, mainly concentrating on error correction techniques to resolve the radio problems below the ATM protocol layer. The ATM Forum are developing standards in support of ATM over radio links under groups focussed on signalling, routing and security enhancements.

Motivation

Given this background, it is no surprise that there is defence establishment interest in using ATM in tactical networks. Its flexibility makes it ideal for a range of scenarios. For such systems COTS technology offers the prospect of cost savings, good interoperability with other tactical and civilian systems, and well-tested standards. Moreover, since ATM is being deployed in public networks, the technology should be long-lived.

This paper summarises the results of a study on the suitability of ATM for tactical networks. The study was recently conducted as a joint activity by DERA and Marconi Research Centre. The tactical requirements were determined, against which a COTS ATM solution could be assessed, and any shortfalls characterised. In order to facilitate assessment it was necessary to develop an example deployment, referred to as an "example corps" in this paper, incorporating projections of UK needs around 5 years into the future. The subject under

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

examination was principally the current and emerging standards of the ATM Forum.

The following sections address:

- tactical requirements and assessment of ATM technology against each of them
- the existence of established ATM technology, and evolving future ATM technologies
- the different issues associated with terminal mobility and switch mobility.

The assistance of DERA in producing this paper is gratefully acknowledged.

2. Requirements of tactical networks

In a future corps, it is anticipated that communications between assets will be provided by a network typically composed of the following parts:

- A land-based network consisting of switches and wireless paths, providing a trunk system
- Air or space-based switches, allowing the trunk system to work over greater distances
- Switches within vehicles or groups of vehicles forming HQs. These access the trunk network via access switches.
- A system to allow users to access the trunk system whilst mobile. Access to the trunk network is via Radio Access Points (RAPs)
- Gateways to other communications systems.

Typically, a land-based tactical network will need to support communication between fixed HQs, mobile HQs and various other users, some of which are mobile (e.g. man packs) while others are stationary (e.g. unattended sensors). The network will cover a forward and rear battle area, which may be as large as 80 km width by 200 km depth.

The trunk system, based upon point-to-point radio links, will serve to interconnect fixed HQs and RAPs, which provide multidirectional coverage of an area for access by mobile HQs, stationary sensors, etc.

The main requirements of the tactical network for the future corps described above are as follows:

- Transparency of mobility, offering seamless communication, between stationary or mobile users over an infrastructure of which each part may be stationary or mobile. Types of mobility include offline mobility, in which a unit goes offline prior to a move and goes online again after the move, and personal mobility, in which the user changes terminal or changes role (as happens during a change in command) without the need for terminal or network movement. Offline/online transitions must be executed with minimal disruption to the network.
- Tolerance of network fragmentation and de-fragmentation.
- Graceful degradation, allowing the graduated and controlled loss of service arising under conditions of excessive demand.
- Prioritisation and pre-emption where higher priority connections should be preferentially allocated capacity, pre-emptively when applicable. Where possible, connections being displaced by those of higher priority should be offered reduced QoS rather than clearing the connection.
- Scalability, guaranteeing that mobility shall be applicable over a range of network sizes.
- Radio silence, in which unidirectional connections are allowed to equipment operating in this mode.
- External gateways, which provide an interface to other networks.
- Switch installation capability within an operational network.
- Survival of the unplanned loss of a switch or link. Affected connections shall have the option of being re-routed, with some traffic loss, rather than clearing.
- Support for diversion of traffic from a switch prior to a planned outage, e.g. for a non-operational move.
- Handover capability of a terminal from one RAP to another with minimal data disruption over active connections.
- Multicast and multisource support
- Routing for efficient use of network resources such as the signalling bandwidth required.
- Link error rate parameters should be considered as part of QoS for each connection.
- Addressing by name, role or terminal identifier.
- Location management to provide the mapping between a called user and their current virtual address (location), and to provide resilience in the event of loss of individual location servers, or fragmenting of the network.
- Authentication of user to network, network to user, and switch to switch, with resilience in the event of loss of individual authentication servers or network fragmentation.
- Management control of switches, RAPs, terminals and connections. Logging, statistics-gathering and diagnostics shall be included.

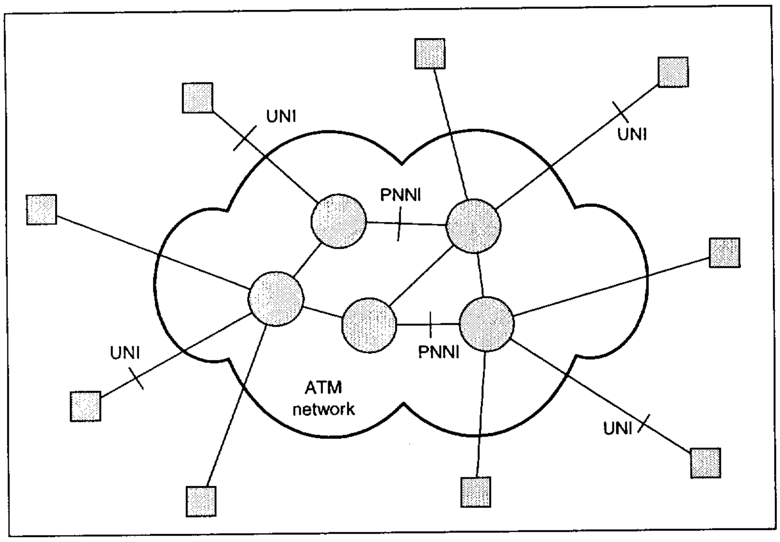


Figure 1 - ATM network

3. ATM background

An ATM network consists of switches, freely interconnected according to user requirements, that is, without a topology prescribed by the technology. A mesh topology is typical. Connected to the switches are users' terminals (Figure 1).

Users intercommunicate by placing calls to one another, much like traditional telephony. The connections may be set up on demand by user signalling. Multicast and multisource connections are also possible. As mentioned, connections are virtual (hence VCs), in that switching is performed by logic, which directs user data, packaged into 53-octet "cells", onto the appropriate output port for multiplexing with cells of other VCs. Resources such as buffer space are allocated to VCs at each switch according to their negotiated QoS.

The terminal interface to the network is specified by the ATM Forum's User-Network Interface (UNI) [2]. The network infrastructure is given cohesion by the ATM Forum's Private Network-Node Interface (PNNI) [1]. PNNI is an inter-switch protocol with two components: a signalling protocol and a routing protocol. The signalling protocol is used to direct the setup of connections through the network. The routing protocol is used to propagate topology information (which switches are connected to which, and with what spare capacity) and reachability information (the terminal addresses accessible through each switch) around the switches of the network. Each switch maintains a database of topology and reachability information, and database changes are flooded to other switches using the routing protocol. Each switch is responsible for originating the flooding of its own information, e.g. when a new link is added to the switch.

The topology of a PNNI network is influenced not only by the distribution of physical links, nodes and addresses, but also by PNNI's ability to summarise this information in a hierarchical way by the following means:

- Aggregation of addresses, where only one entry is required to cover a set of closely related and collocated addresses
- Grouping nodes for flooding of full topological information, and flooding only an abstraction of the topological information outside and vice versa
- Aggregation of physical links.

It is this hierarchical summarisation which allows PNNI to scale to large networks. Without this hierarchy there is a risk that PNNI flooding will saturate the capacity of large networks. The hierarchy is built from groups of switches termed "peer groups", and in standard PNNI version 1.0 peer group membership is configured by management.

Peer groups are nested, each having a switch that acts as a peer group leader, taking part in inter-peer group routing protocol exchanges on behalf of the peer group. The peer group is treated as a logical node as viewed from further up the hierarchy. Peer group leaders are chosen as part of an ongoing PNNI election protocol, and a new leader is appointed if the old leader fails.

Connections are routed by an ingress switch on the basis of routing information in its database. The route is expressed in detail within the peer group of the ingress switch, and summarised in terms of peer groups thereafter. Switches in other peer groups subsequently amend the connection setup message to include details of the routing within their peer groups. Hence PNNI is source routed, albeit not fully source routed in the multiple peer group case.

When a connection request arrives at a switch it performs Connection Admission Control (CAC) to decide whether there are sufficient resources to accept the connection. The CAC algorithm is not standardised and depends upon switch technology, wireless state in an adaptive system, and vendor-specific software design. To overcome this problem, PNNI specifies a generic CAC algorithm to be executed by switches when making routing decisions. This is, effectively, a best guess about the outcome of a particular CAC decision at each switch along the connection path.

4. User mobility issues

Users associate themselves with terminals, which are attached to some point in the network and may be either fixed or mobile assets. The system must identify location, provide authentication, and enable mobile terminal handover.

Location management

Location management is the means by which a network locates terminals. While PNNI routing is capable of providing sufficient information for a call to be set up to a static terminal, it is not suited to doing so for a mobile terminal. Routing protocols can accommodate a slowly changing network but are not generally fast enough to keep pace with the whereabouts of a mobile terminal.

Some military networks implement location management with flooding techniques that are very robust but do not scale, and have therefore not found favour within COTS standards. Hence we consider the only known alternative: the use of a database which maintains the current point of attachment for each mobile terminal. This database is updated each time a terminal relocates to a new switch and is queried each time a connection is to be established to a mobile terminal. The location database works alongside the routing database; both are needed to place a call to a mobile terminal.

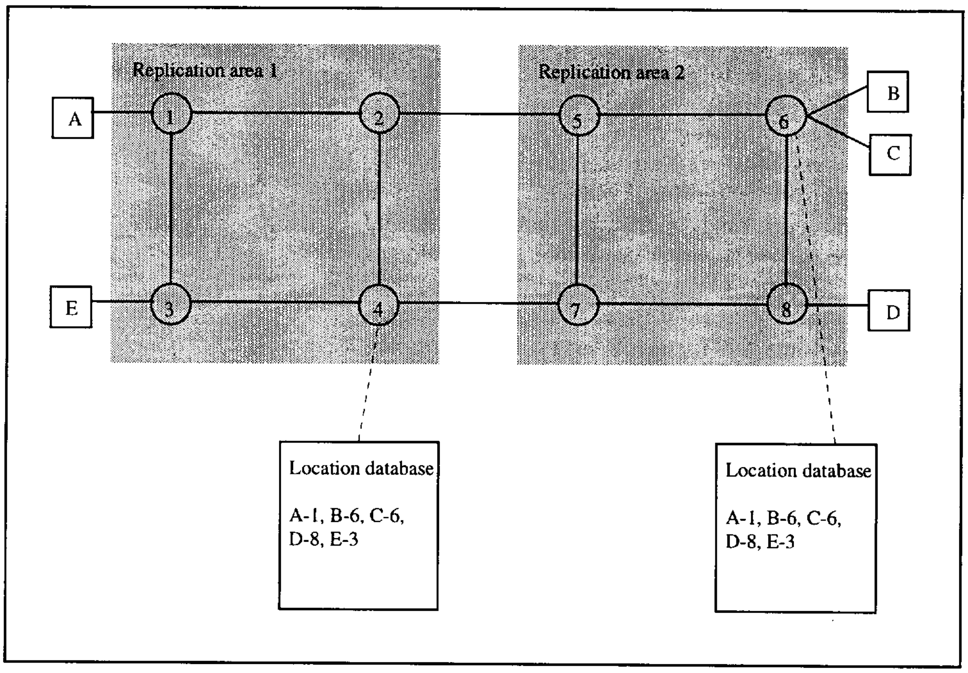
The location database concept is simple, but there are many implementation options for its structure. The database should be replicated (i.e. elements copied) and the replicates distributed about the network to avoid a single point of failure. Replication also has the benefit that the average distance between a client requiring access and the nearest replicate is reduced, thereby improving the performance of queries. However, the performance of the location service depends on rapid synchronisation of location database replicates, and updating multiple replicates is a complicated process, leading to a performance trade-off. The degree of replication must be such that the risk of loss of location service is kept to an acceptable level. Network fragmentation is also a threat, survival of which requires a replicate in each potential fragment. Replication of location databases is hence a detailed network design issue, which must take due account of areas which run a high risk of becoming fragmented (e.g. islands). It is clear that a tactical network will require at least 2 replicates in a well-connected backbone plus 2 in each area with a significant risk of fragmentation. The location database distribution could be as illustrated in Figure 2.

One aid to location management is *paging*, which permits a lower resolution in location tracking. If paging is used then location management tracks the mobile terminal by location area (a group of RAPs), rather than a single RAP. Upon connection setup all RAPs in the location area are commanded to attempt to contact (i.e. page) the terminal, and the one which obtains a response is the one which carries the connection.

Paging offers a trade off: location update traffic is reduced but connection establishment has a greater overhead. Paging is therefore more appropriate in GSM networks, where there is often considerable terminal movement between calls, than it is in the tactical network where there are relatively frequent connection set-ups, and which typically will be to the same cell as the previous connection.

Caching is also a popular computing technique for reducing remote access by storing the results of previous queries, and is an obvious candidate for reducing location database accesses. However, without detailed studies of movement patterns, its effectiveness is doubtful.

Figure 2 – Example of location database distribution. The location database has two replicates.



Authentication

Two-way authentication between the terminal and the network is implemented by a two- or three-way handshake between newly associated entities where the transfers must be encrypted and include a pseudo-random element to protect against replay attacks.

The network requires a database with information about the privileges of all authorised terminals so that the rights of each newly arriving terminal can be established. Distribution of the authentication database follows the same considerations as distribution of the location database – replication is recommended but partitioning is not.

Handover

Handover is concerned with re-routing existing connections to a new RAP when the terminal relocates. This is achieved by building a new connection segment between the new RAP and some point on the path of the existing connection. With appropriate handshaking and buffering the handover operation could avoid loss of data in transit.

Handover implies a change in location, so the location database must also be updated in the usual way when there are connections.

Personal mobility

Personal mobility is achieved as a user registers with a static or mobile terminal, and subsequently re-registers with a new terminal, typically following user movement. Alternatively, a user may re-register with the same terminal under a different role.

Personal mobility requires a database to map user name (or role) to terminal. This database, which is called a directory, is updated each time a user changes terminal or role, and is queried whenever a connection is to be established. As with location management, the database must be replicated to increase availability of the service.

In principle, the name and location servers could be combined into a single server, with a single query returning both the routable and the permanent terminal address. However, this has the disadvantage that any partitioning, if required, would be difficult because of the different characteristics of the name and location spaces, e.g. when a name is mapped to a different location, the new location may reside in a different partition. Moreover, although each service operates in a similar way, the location service will be required to handle a fast-changing data set that is automatically updated by the network, whilst the role/name-to-number mapping of directory services will alter less frequently and may be manually updated.

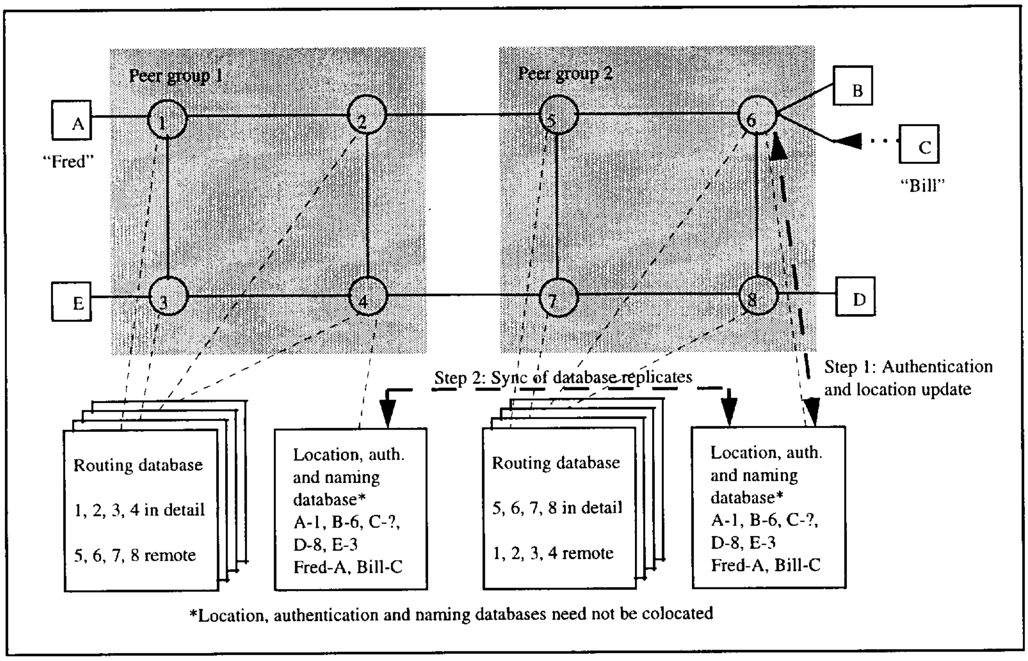
ATM and user mobility

The WATM standard will specify location management, handover and extensions to signalling (both UNI and PNNI). WATM is the name of the group set up by the ATM Forum to establish standards for "Wireless ATM". The standard will also specify the Access Point Control Protocol (APCP), the protocol for controlling an Access Point which is a unit for adapting traditional wired ATM interfaces to wireless. WATM is not specifying the radio layers; this is left to groups such as ETSI/BRAN.

The WATM standard specifies how calls are set up to a mobile terminal. When a mobile subscriber attaches to a switch, it is first authenticated and then there is a location update as illustrated in Figure 3.

When a subscriber wishes to establish a connection to another, the connection is routed towards a switch

Figure 3 - Illustration of authentication and location update for C. The "?" in the location database indicates that the location of C was unknown before the update. After the update the entry is C-6. For simplicity the UNI signalling is not shown.



hosting a location database replicate, based on the permanent address of the called mobile terminal. Any switch on the path, if so equipped, can query the location database to obtain the current address of the mobile terminal. The connection request is then forwarded towards the mobile terminal. In Figure 3, for example, terminal A may wish to call terminal C. The call might be routed through switches 1 and 2 to 4. Switch 4 is able to perform a location query to obtain the address C-6, which is used to route through 7 on to 6, and finally to C.

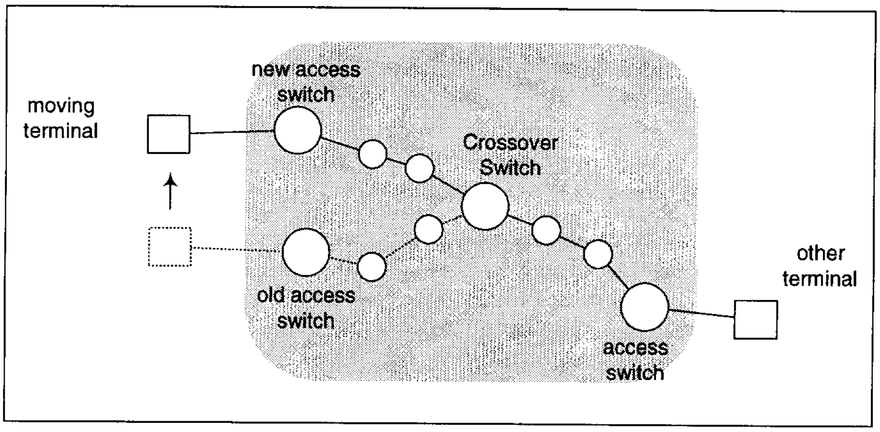


Figure 4 - Terminal handover example

The WATM standard [4] also specifies terminal handover, the process by which a mobile terminal changes its RAP, and possibly the network access node, in a co-ordinated manner while maintaining its active connections. The ATM Forum has identified several possible variations in the process and has pursued some of these.

In the case of an inter-switch handover each VC to the mobile terminal is re-routed away from the old access switch to reach the mobile terminal via the new access switch. At some point on the path of each VC is a switch, termed the Change-Over Switch (COS), which is the limit of the re-routed segment of the VC. This is illustrated in Figure 4.

The process of handover as recommended by the ATM Forum entails the release of the old data connection before the new data connection is established and does not introduce cell misordering or duplication. However, there is the possibility of cell loss, due to asynchronous changeover of connection segments at the terminal and the COS, or cells being discarded along the old path if connection release propagates faster than user cells. It was estimated that data loss over 0.1 seconds could occur, corresponding to just over 1 cell on average for a 4800 b/s CBR connection and around 40 cells for a 128 kb/s CBR connection. Data loss is sensitive to the lengths of the old and new paths of the re-routed segment, due to the buffering effect of the network, and also to the length of time it takes for the mobile terminal to disassociate and re-associate with the new access node. The possibility of cell loss may pose problems for encrypted data that requires synchronisation.

5. Network mobility

Network mobility implies the mobility of switches and the links that interconnect these switches. There are broadly two cases of switch movement:

1. movement while not operating
2. movement while operating (i.e. carrying traffic)

If a switch moves in a non-operational state then, in general, its ports will be reconnected to different switches following the move. The tasks involved in non-operational movement are as follows (in order of execution):

- traffic blocking
- diverting existing connections
- clearing any remaining connections
- link closure
- switch (re)installation

Traffic blocking is where new calls are prevented from attempting to use the switch. By restricting access under such circumstances, the subsequent task of clearing existing calls is reduced. Moreover, the potential waste of network resources in setting up connections that will soon be dropped is minimised. Clearly, prediction of the removal of a resource is sometimes impractical (e.g. equipment failure, hostile action, etc.), but in many instances infrastructure mobility will be a planned event.

Many connections may be in transit across a trunk switch, and wherever possible such calls should be diverted away from the switch prior to power-down. Otherwise, at best, all failure-protected connections will attempt to re-route simultaneously, whilst unprotected connections will clear, with the likelihood of reconnection attempts being initiated by users, possibly from both ends, resulting in an increase in the transient network signalling load.

When diverting connections away from a trunk switch, it may not be possible to transfer *all* the traffic from the switch without exhausting resources in other parts of the network. Selection according to priority and/or service type may be necessary.

Existing transit connections may be diverted using Edge-Based Re-routing (EBR), shown in Figure 5. EBR is a means for re-routing existing connections within a PNNI network. A new route is selected between the ingress and egress switches on the basis of the PNNI database in the ingress switch at the time of the reroute operation. Hence the new route may have segments in common with the old route.

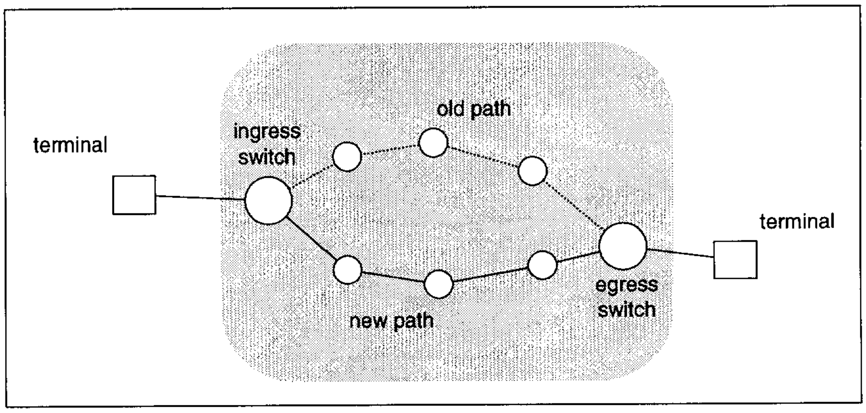


Figure 5 - EBR example

The ATM Forum is developing two forms of EBR: 'hard', which uses a 'break-before-make' approach such as may arise in the event of link or equipment failure, and 'soft', in which a new route is established alongside the existing one - a 'make-before-break' approach. The latter offers the prospect of diversion of a connection path at the instigation of a management function, and is transparent to the end-users.

Clearing existing connections should be done only if they cannot be re-routed.

Soft EBR could be used for operational switch mobility, although further standards development is needed. For example, a mechanism for advance warning before links become unavailable is needed.

An alternative approach to operational switch mobility is to treat a HQ as a terminal, attached to the trunk network via a RAP, and communicating with the nearest trunk switch over the UNI rather than PNNI. Although this is feasible from the location management standpoint, it precludes the use of the HQ access switch for transiting trunk connections.

6. Assessment against requirements

The study assessed the use of ATM in mobile tactical systems. An example corps was assumed and used for quantitative analysis. This is described, followed by a baseline of ATM Forum standards forming a solution. Finally, the extent to which the solution satisfies the tactical network requirements is indicated.

Example corps

The example corps may be viewed as representative of a large UK deployment in around 2005. The example corps assumes over 120,000 personnel, consisting of some combat and some support divisions. Each division has brigades, each brigade has battle groups (BG) and each BG has companies (coy).

Each BG has equipment acting as data sources and/or sinks. These might include:

- Sensor equipment (ESM equipment, radar, unattended tactical sensors)
- Sensor vehicles (air and ground reconnaissance vehicles, unattended airborne vehicles (UAV))
- Weapon and/or logistic control systems
- Key patrol/soldier/appointment data terminals
- Vehicle sensors (which provide information on the state of the platform)

Each of the elements (corps, division, brigade, BG and coy) have a number of headquarters (HQ). Their roles and function differ. At higher levels of the organisation, a 'handover' of the command and control of the battle between HQs is necessary at various times and rates. At lower levels, the HQs command the battle whilst on the move.

Baseline of the ATM mobile network

The ATM Forum standards and work in progress used in this study were:

- PNNI 1.0 [1]
- UNI 4.0 [2]
- ILMI 4.0 [3]
- WATM Capability Set 1 as laid down in BTD 1.09 [4]
- PNNI Edge-Based Rerouting Extension as defined in BTD 1.03 [5]
- PNNI Mobility Extension as defined in BTD 2.03 [6]
- ATM Security Specification v1.0 (draft) [7]
- PNNI Secure Routing Addendum as defined in BTD 2.00 [8]
- UNI Signaling 4.0 Security Addendum [9]
- PNNI Version 1.0 Security Signaling Addendum [10]

The study assumed the use of a directory service, either a standard such as X.500 or, possibly, a proprietary service developed with an understanding of land-based tactical naming.

The PNNI peer group layout is shown in Figure 6. It comprises 3 levels: one peer group at the top, 1 level-2 peer group in the Rear Battle Area (RBA) containing a

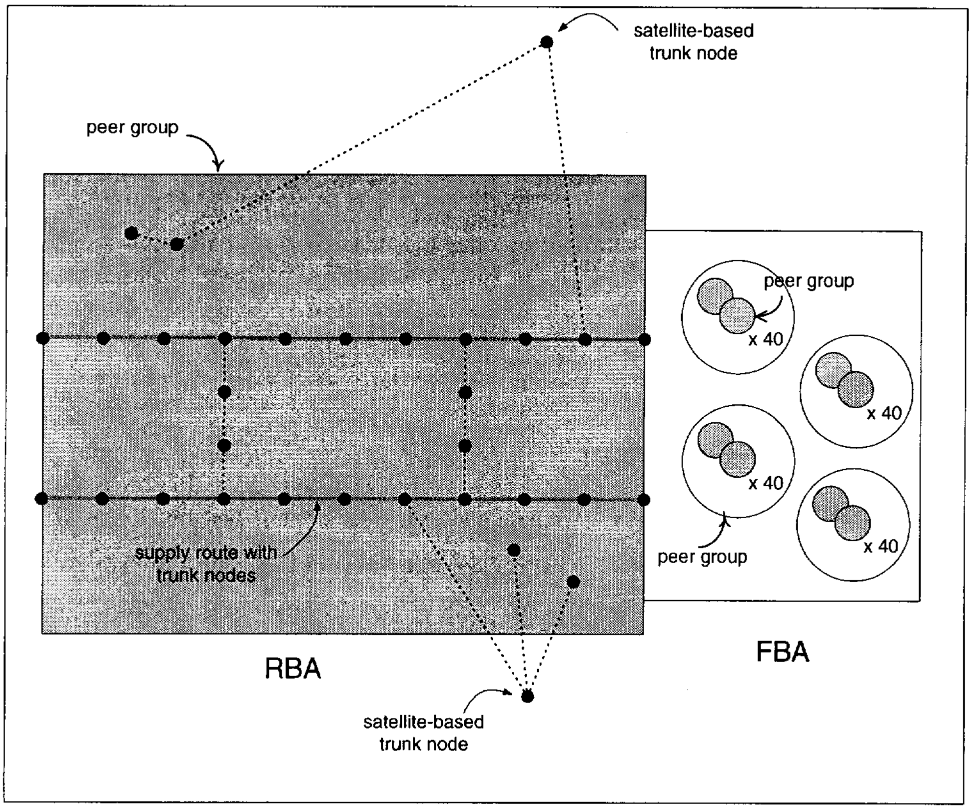


Figure 6 - PNNI peer group layout

mix of around 40 HQs and trunk switches, and 4 level-2 peer groups in the Forward Battle Area (FBA) each containing a mix of around 40 HQs and trunk switches. Each HQ is a level-1 peer group and in terms of PNNI routing appears as a single switch to other HQs and trunk switches. Other PNNI variables, such as thresholds for advertising changes in available bandwidth across links, were assumed to be their default values.

Extent of satisfaction of tactical requirements

Transparency of mobility The requirements are satisfied, however:

- Directory database access requires support by the user and/or application software. This is not considered onerous because it is done, for example, by Internet Web browsers which use DNS to convert URLs to IP addresses.
- The PNNI source route imposes a limit of 20 nodes at each level of the PNNI hierarchy. Normally this is not a problem, but it would be a problem if the network was damaged to such an extent that routes, after allowing for the PNNI peer group structure, exceed this length.
- Legacy switches must run PNNI V1.0 (or later) for compatibility with mobile ATM.

Network tolerance of fragmentation The COTS standards are unlikely to support the necessary replication of the critical location and authentication databases. It may be that a solution can be identified to this problem, possibly using additional features in a product from a particular vendor, but it is not possible to be certain of this until the standard is stable and product information is available. The key problems to be solved are:

- Synchronisation of replicates
- Finding alternative replicates after loss of access to primary replicate
- Re-establishing authoritative information following merger of network fragments

Note also the following:

- A network fragment cannot operate unless it has at least one replicate of the directory, location and authentication databases. This can be mitigated by careful network design.
- It is important that applications are designed to avoid too frequently repeated connection attempts as this could cause a high signalling overhead.

Graceful degradation Graceful degradation is provided by the prioritisation and pre-emption mechanisms when supported (see below).

Prioritisation and pre-emption Prioritisation and pre-emption are not supported by the standards analysed but the ITU and ETSI are working on these features in ISDN [11] and GSM [12]. Although each connection can be assigned a priority by extension of the signalling protocol, pre-emption is problematic. The decision to pre-empt is taken at each switch, and must be taken by all switches in the path. If pre-emption is refused by a downstream switch due to the absence of resources in use by lower priority calls then reservations already made will have to be relinquished. Added to this, the decision on which calls to pre-empt is complex – differences in call bandwidth as well as priority lead to trade-offs between one call and another set of calls of lower bandwidth.

Scalability The requirements are satisfied for networks up to the size of the example corps. The PNNI peer group structure is needed for routing information summarisation but partitioning of directory, location and authentication databases are not required.

Radio silence Full support for radio silence by COTS products is not expected. Power saving features are common, but a full listen-only state, with no possibility of location updates whilst the network continues to transmit unidirectional traffic to the terminal, is unlikely. Radio silence could be added to an existing product as an extension, controlled by management or proxy signalling.

External gateways External gateways are supported. Communication with an external PNNI network is seamless, and in future tunnelling options may exist to communicate via non-PNNI networks.

Switch installation A switch can be installed while the network is operational. After authentication, PNNI enables routing data exchanges to make the switch known to the network and vice versa. It was estimated that this process would typically complete within 20 seconds.

Switch loss and survival of the network Users connected to a switch in a non-redundant manner will obviously be impacted by switch loss. Connections broken as a result of a lost trunk switch can be re-requested, either by hard EBR or by the application software. Unfortunately this can cause a high peak in signalling traffic when a large number of connections attempt to reconnect simultaneously, and this is made even worse by crankback (a PNNI mechanism for trying alternative routes in the presence of congestion) due to the possibility of several connection requests selecting the same resources. It may be that the problem can be reduced by careful sequencing in EBR and/or random delays in the application software.

Switch relocation Non-operational mobility will be possible when means are specified in the standards to trigger soft EBR. These could be either by management or signalling, although the latter is preferable because signalling normally has a smaller network overhead than equivalent management operations.

More significant changes to the standards are necessary to support operational mobility but, in principle, soft EBR could be used provided:

- There is advance warning before links become unavailable
- There is spare capacity in order that the connections can be re-routed

Link failures A fast failing/recovering link could cause a high management overhead so requires closing down by management. An additional management application would be required to do this.

Handover The network overhead due to handover is calculated to be acceptably low – less than 10% of the capacity of each 2Mb/s trunk link.

Further work is in progress within WATM to ensure that QoS is preserved in the new connection segment following handover.

Multicast and multisource Multicast and multisource are not supported by the current WATM work, and neither is multicast over the air supported by ETSI BRAN. However, these features are becoming increasingly important for other COTS standards such as conferencing and MPLS (MultiProtocol Label Switching). Future support is anticipated.

Note that multicast and multisource can be replaced by a set of unicast connections when resources permit.

Routing The PNNI routing overhead was estimated to not exceed 10% of the bandwidth of 2Mb/s trunk links with the configuration variables set to their default values. These include thresholds for advertising changes in available bandwidth across links.

Link Error Rate Parameters The ATM standards do not treat link error rates as a signalled parameter. Instead the approach is to treat link error rates as a network-defined property, and remove links from service if they cannot offer the intended reliability. Extensions to the signalling standards would be necessary to satisfy this requirement.

Addressing Addressing by name, role or terminal number is achievable with the assistance of directory services, currently under development at the ATM Forum.

Location management Location management is provided and the overhead is a small proportion of the link bandwidth. Call setup times are estimated to be increased around 50% due to location management. Location database replication introduces the problem of synchronisation. Lack of synchronisation will result in failed calls.

It is considered unlikely that it will be possible to inhibit location updates in COTS equipment to support the radio silence requirement.

Authentication management The ATM Forum are addressing authentication in their Security Specification [7] and in secure PNNI routing [8] and signalling [9,10]. In addition, WATM are developing a specification for authentication of a mobile terminal to the network and vice versa, entailing a 3-way handshake.

Management requirements Most of the management requirements are satisfied provided all equipment has the full standard MIBs, including RMON in the switches.

7. Conclusions

COTS ATM technology is well supported by many vendors. It is becoming well tested, is available at a good price-to-performance ratio compared with non-COTS solutions, and is future proof - the technology will exist for a long period and is likely to evolve in functionality and performance.

However, the full requirements of land-based tactical networks are not yet satisfied, though standardization continues. Location, directory and authentication databases, when eventually implemented, are unlikely to satisfy the full requirements, especially for replication at different sites. Even with extensions, they are unlikely to be as robust as a flooding-based solution, although they will scale better. COTS standards do not support radio silence.

Although re-routing standards are being developed, there are none as yet for triggering the re-routing, nor for sequencing the re-routing to reduce overload. Further development of COTS standards in this area may be beneficial. In addition, there are no standards yet for operational trunk switch mobility, but again it may be possible to develop COTS standards in this area. Multicast and multisource connections are not currently supported by WATM, although there are prospects for future standardisation. Standardisation of priority and pre-emption is progressing, although pre-emptive routing in multi-service networks presents many challenges. Link error rates are not treated as signalled QoS parameters by COTS standards. WATM still has much work to carry out in the area of authentication.

It should be noted that the majority of the issues are not specific to ATM and indeed apply to most technologies, even those developed specifically for military applications.

8. References

[1] Private Network-Network Interface Specification Version 1.0 (PNNI 1.0)
ATM Forum document ref: af-pnni-0055.000
March 1996

[2] ATM User-Network Interface (UNI) Signalling Specification Version 4.0
ATM Forum document ref: af-sig-0061.000
July 1996

[3] Integrated Local Management Interface (ILMI) Specification Version 4.0
ATM Forum document ref: af-ilmi-0065.000
September 1996

[4] Wireless ATM Capability Set 1 Specification – Draft
ATM Forum document ref: btd-watm-01.09
October 1998

[5] Edge-Based Re-Routing for Point-to-Point Connections
ATM Forum document ref: btd-cs-pnni-reroute-01.03

[6] PNNI addendum for Mobility Extension
ATM Forum document ref: btd-ra-pnni-mobext-02.03

[7] ATM Security Specification Version 1.0
ATM Forum document ref: af-sec-0100.000
February, 1999

[8] Secure PNNI Routing, Security Addendum to PNNI v1.1
ATM Forum document ref: btd-ra-pnni-rsec-02.00 draft
December 1998

[9] UNI Signaling 4.0 Security Addendum
ATM Forum document ref: af-cs-0117.000
March, 1999

[10] PNNI Version 1.0 Security Signaling Addendum

ATM Forum document ref: af-cs-116.000
March, 1999

[11] Multi-level Precedence and Pre-emption Service (MLPP)
Recommendation I.255.3
July 1990
ITU

[12] Enhanced Multi-level Precedence and Pre-emption Service (eMLPP)
Stage 1, GSM 02.67, version 5.0.4
ETS 300 924
ETSI, July 1997

E-mail: phil.walker@gecm.com

URL: http://www.gecm.com/

Proposed Concept for a Non-LOS Tactical Wireless LAN

Svein Haavik
FFI (Norwegian Defence Research Establishment)
P.O. Box 25, NO-2027 KJELLER
svein.haavik@ffi.no

SUMMARY

New concepts for land warfare and the increasing use of civilian computer and communications systems are leading to a demand for more bandwidth on the battlefield. The introduction of dispersed command posts with WLAN interconnection is a tough challenge for communications systems designers, especially when combined with a requirement for mobility and ease of operation. SHF-systems to be used in forest terrain will need high antenna masts or must try to overcome the large propagation loss caused by vegetation. We are investigating the possibility of operating a high-capacity SHF-based LAN-system for dispersed command posts in typical Norwegian terrain, allowing for line of sight to be obstructed by vegetation.

1 BACKGROUND

The introduction of new command & control information (C2I) systems for tactical land forces is accompanied by an increasing demand for communication capacity. In addition, the growing use of commercial hardware and software products - designed for 10/100 Mbps or more local area network (LAN) environments - in tactical systems, contributes to this increasing demand. This is assumed to affect both local area (command post) communications as well as wide area (trunk) communications in the near future.

One of the lessons learned from the Gulf War was the growing threat from precision weapons. This may well affect the command posts (CPs) of tomorrow. One alternative to meeting this threat is the introduction of dispersed command posts. With a distance of a few hundred meters between the different cells (vehicles) a single precision weapon or artillery attack is less able to set the whole CP out of operation. The desired distance between cells in a dispersed CP is about 500 m, with a minimum requirement of 300 m. Such a dispersed CP makes new demands to the communication system. The use of optical fibre increases the time required to establish and move/re-establish a CP, but gives sufficient bandwidth. Radio communications, although being able to offer less bandwidth, give a greater flexibility in deployment and increased mobility.

Taking into account the requirements for bandwidth, the potential problems of radio communications is range and the threat from electronic warfare (EW). Considering the general technological move towards higher frequencies, the EHF band is very promising for short range systems within line-of-sight (LOS), with 60 GHz offering exceptional electronic protection measures (EPM) for local area systems, due to the strong oxygen absorption. But the use of EHF band requires strict LOS conditions, at the sacrifice of mobility and easy deployment in forest terrain.

At FFI we have studied different radio LAN concepts for the past years. Our goal is to be able to provide high capacity radio communications with a range of a at least 300 m allowing for vegetation to block absolute line-of-sight.

2 PROPAGATION MODELS

A number of propagation models have been developed, and some are also suited for the prediction of the propagation loss experienced when microwaves propagate through vegetation. ITU-R has proposed a model [1] which is valid for the frequency range 200 MHz to 95 GHz at vegetation depths up to 400 m.

$$L = 0.2 f^{0.3} d^{0.6}, \qquad (1)$$

where f is frequency in MHz and d is distance in meters. This model is assumed to be inaccurate at microwave frequencies as it is based on measurements carried out mainly at UHF. Weissberger [2] has proposed a slightly modified model.

$$L = 0.187 f^{0.284} d^{0.588} \qquad (2)$$

Al-Nuaimi and Stephens [3] have presented another modification to the UTU-R model based on measurements at 11.2 and 20 GHz. This latter model, the fitted ITU-R model (FITU-R), distinguishes between two different foliation states: in-leaf and out-of-leaf.

$$L_{in\text{-}leaf} = 0.39 f^{0.39} d^{0.25} \qquad (3)$$

$$L_{out\text{-}of\text{-}leaf} = 0.37 f^{0.18} d^{0.59} \qquad (4)$$

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

All four models are plotted in Fig. 1 as function of frequency at our minimum required range of 300 m. The free-space loss is included as a reference. We see that there is a large discrepancy between the different models, with the UTU-R model predicting the highest attenuation at all frequencies. Note also that the FITU-R model is probably not suited below 8 GHz where the in-leaf and out-of-leaf curves intersect (at 300 m distance).

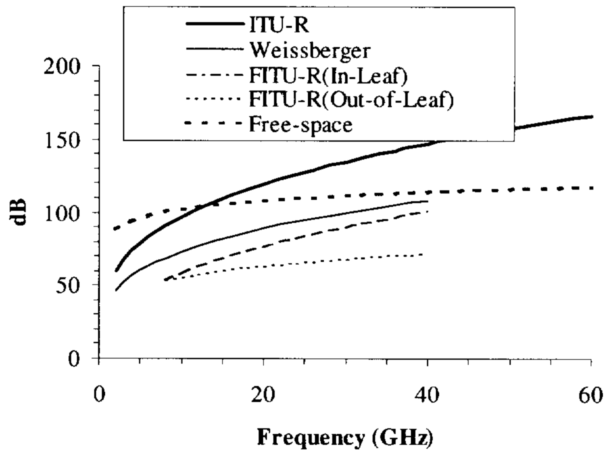


Fig. 1 Vegetation attenuation at 300m distance as a function of frequency.

For a system with antenna gains G_{tx} and G_{rx}, output power P_{tx} and a receiver sensitivity $P_{rx,min}$, the maximum allowed propagation loss is given as:

$$L_{max} = P_{tx} + G_{tx} + G_{rx} - P_{rx,min} \qquad (5)$$

An affordable and technically realisable system with omnidirectional antennas might have the following parameters: $G_{tx} = G_{rx} = 7.5$ dBi, $P_{tx} = 30$ dBm and $P_{rx,min} = -99$ dBm (for a link of 1 Mbps). This gives $L_{max} = 144$ dB, which must be used to fight both the ordinary free space loss as well as the vegetation loss, ignoring other effects such as diffraction and precipitation. Combining free space and vegetation loss in Fig. 2 we see that a range of 300 m (our minimum requirement) is unattainable at frequencies above approximately 3 GHz, even with the most optimistic vegetation loss model.

Often, even higher data rates than 1 Mbps are desired for several command post systems. And for a military communications system EW capability is usually required, adding additional bandwidth requirements. As a result, it may be difficult to find sufficient bandwidth for a military radio LAN without moving to higher frequencies.

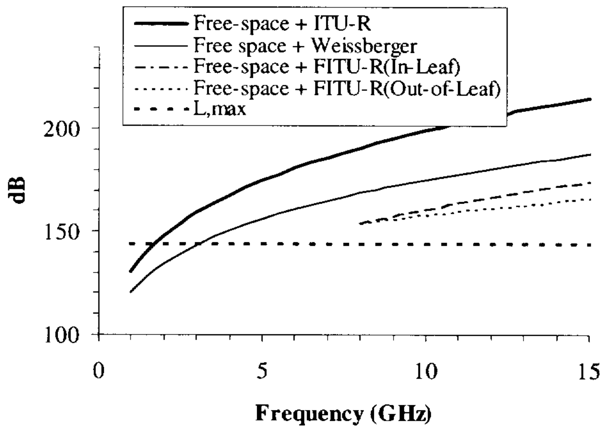


Fig. 2 Total attenuation at 300m distance, and the loss limit for a multi access system with omnidirectional antennas.

3 PREVIOUS WORK

At FFI we previously studied the possibility of developing a multi access system in the 2.4 - 2.5 GHz IMS-band. Such a system offers the necessary range for a wireless LAN (WLAN), as will be shown below, and supports mobility. The work was not carried further due to the shortfalls of this concept regarding capacity and lack of both LPI (Low Probability of Intercept) and anti-jam performance.

During the study we performed several measurements of the propagation loss in this frequency band [4], [5]. A large part of these measurements focused on foliage loss in different types of vegetation (deciduous and mixed deciduous/conifer forest). Some of these results are plotted in Fig. 3, together with model predictions. We see that our results are well below the predictions of ITU-R, but are comparable to the predictions of Weissberger, even though the measurements seem to be systematically lower than the predictions. One possible reason for this discrepancy is that the measurements are performed in natural forest where you will find large variations in the foliage density on a random section, while the predictions assume a constant high density.

Fig. 3 includes two limits. One of them is the estimated theoretical attenuation limit of the affordable and technically realisable system mentioned in the previous chapter. The other limit is that of a civilian COTS product - a DS SS WLAN according to IEEE 802.11 which will be discussed in the next chapter. We see that it should be possible to make an omnidirectional WLAN in this band which will have a range of 300 m or more in typical Norwegian forest (assuming only vegetation to obstruct line of sight). But such a system may not operate within the above mentioned IMS-band without violating the regulations regarding output power.

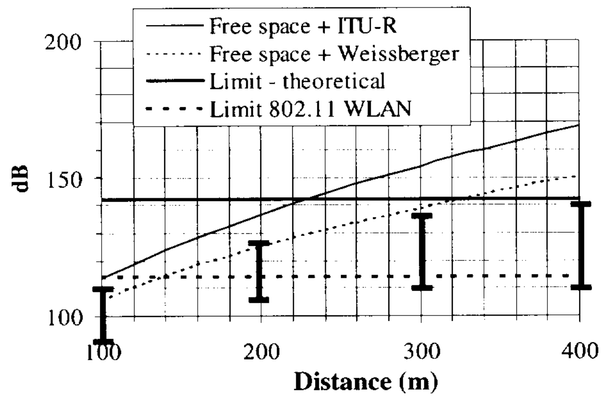


Fig. 3 Total attenuation for a 2.5 GHz system based on omnidirectional antennas.

4 CIVILIAN COTS PRODUCTS

For the civilian market there is a growing range of WLAN products, primarily intended for indoor environments. The IEEE 802.11 standard is based on spread spectrum (SS) techniques, but allows two different physical layers for radio: frequency hopping (FH) and direct sequence (DS). The direct sequence choice seems to be the best one for use in an outdoor environment with vegetation blocking direct sight between the network nodes, as it can tolerate a larger propagation loss (typically 10 dB above FH).

The IEEE 802.11 standard has no error correction included, making the systems very vulnerable to fading channels or marginal radio conditions. With increasing bit error rate the re-transmission of all erroneous packets will soon reduce the effective system capacity to an unacceptable level. The standard also has a very limited relaying capability as only the base station may relay, allowing a maximum of two radio hops.

In cooperation with Thomson-CSF Norcom we have performed some initial tests of one of these candidate WLAN products in Norwegian forest terrain [6]. The primary goal was to reveal the practical range in such outdoor environments. The measurements turned out to be in very good accordance with the predictions of Weissberger (see Fig. 3). Our measurements determined the practical range limit to be about 100 - 150 m, which makes such systems unsuited for use by dispersed CPs in forest terrain. A modification of such products to meet our requirements may turn out to be very comprehensive and expensive, thus impairing its greatest advantage - the very low cost of procurement.

5 PROPOSED CONCEPT

In order to meet the user requirements for a WLAN to support dispersed command posts, and at the same time giving data rates comparable to office environments, we propose the following concept:

Users within one cell (typically a vehicle) may for instance use a standard LAN product (e.g. Ethernet) for user access to the communication system. For inter cell communication (between vehicles) we propose a system where the cells are connected by point-to-point communication links. Such a communication system is very consistent with the architecture of ATM (Asynchronous Transfer Mode) which is a prime candidate for switching in future military communication systems. As mentioned in chapter 2, high user capacity generally requires the use of higher frequency bands which experience a greater attenuation from vegetation and precipitation. At the same time, deployment in forest is unavoidable, and even desirable due to its camouflage effect. Due to the attenuation this normally leads to a requirement for line-of-sight (LOS), usually achieved by the use of high antenna masts.

We propose links to use radio transmission in the medium to upper SHF band. In order to support flexibility in deployment and to reduce the time to deploy and move command post cells, we intend to avoid the requirement for line of sight. The greatest uncertainty of our concept is whether the necessary distance between cells may be supported, due to the large anticipated propagation loss with radio links traversing great vegetation depths. The idea is to try to overcome the attenuation by using high gain/ narrow beam antennas. For the measurement system/prototype these antennas are parabolic, but other antenna concepts may be more appropriate for a final system.

Using some data from our COTS-based measurement system, we find that a possible communications system might be realised with the following parameters: Antenna gains of ~33 dBi, output power of at least 25 dBm and a receiver sensitivity of -90 dBm for a 10 Mbps system. This gives an attenuation limit of about 180 dB. Fig. 4 gives the model predictions for a system range of 300 m. From the estimated system limit we see that frequencies above 15-20 GHz are not suited if these models are reliable.

The primary aim of our present study is to determine whether a WLAN system based on directional antenna links may reach the required range of at least 30 0m in various Norwegian forest terrain, operating in the frequency band above 10 GHz. Fig. 4 may indicate that this is possible, at least if the average foliage density of a link turns out to be lower than that assumed by the prediction models. Previous measurements at 2.5 GHz

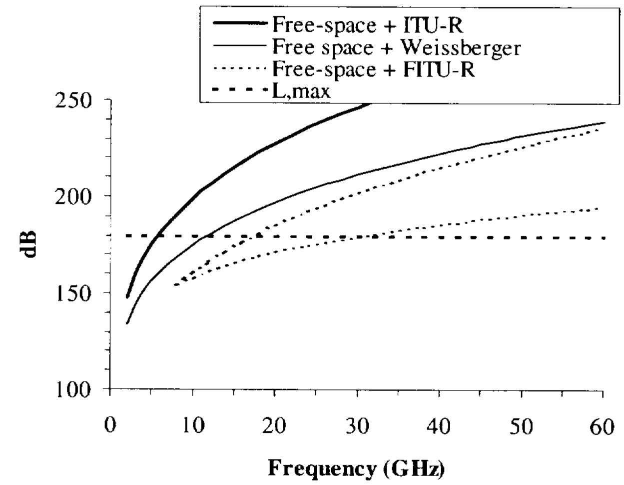


Fig. 4 Total attenuation at 300m distance, and the loss limit for a point-to-point system based on narrow beam antennas.

may indicate that such is the case. Also, none of the propagation models have been verified for vegetation depths over approximately 150 m. All this indicates a great uncertainty and our proposed concept needs further investigation into foliage attenuation before concluding on its potential applicability.

We also want to investigate the possibility of operating a link by pointing both antennas above the target. As indicated in Fig. 5 we intend to reveal whether a lower total attenuation loss under certain conditions may be achieved by aiming an angle upwards. This gives a shorter foliage depth, but introduces a large loss due to the (two) diffractions required.

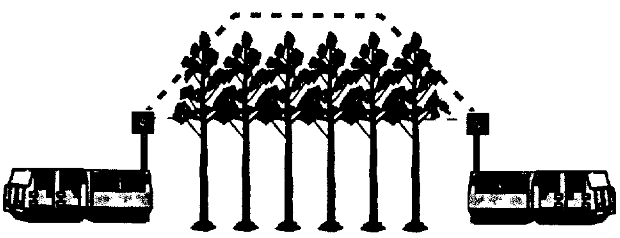


Fig. 5 Alternative concept for SHF narrow beam link establishment, resulting in diffraction at tree top level.

One drawback of a link-based system vs. a multi access system with omnidirectional antennas is that each cell (vehicle) must be equipped with at least two transceivers to constitute a network (unless allowing for a base station concept), thus increasing the cost. Another practical drawback it that the antennas must be directed towards each other with large precision as the beam width may be very small in order to achieve the required gain. One possible solution will be mentioned in chapter 7, as we have implemented a partly automatic way of doing this in our measurement system.

A system operating within forests may experience a link loss varying from low to very high values. This should be exploited by constantly adapting the link bit rate to the available radio link quality, employing a number of different modulation and error correction techniques.

6 PROPAGATION MEASUREMENTS

The three different vegetation loss models from chapter 2 are plotted in Fig. 6 for 18 GHz and Fig. 7 for 38 GHz.

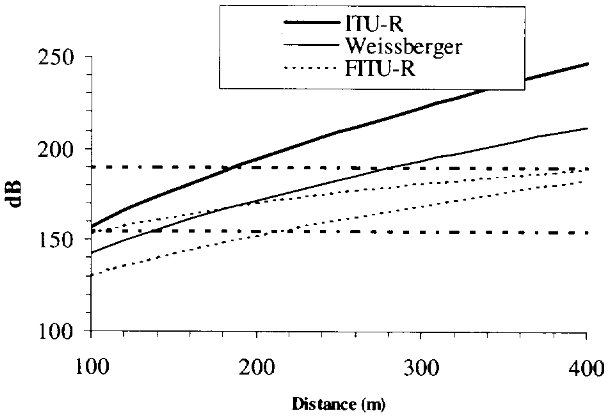


Fig. 6 Predicted total loss at 18 GHz. The two dashed lines indicate attenuation limits for our measurement system (narrow and wide band).

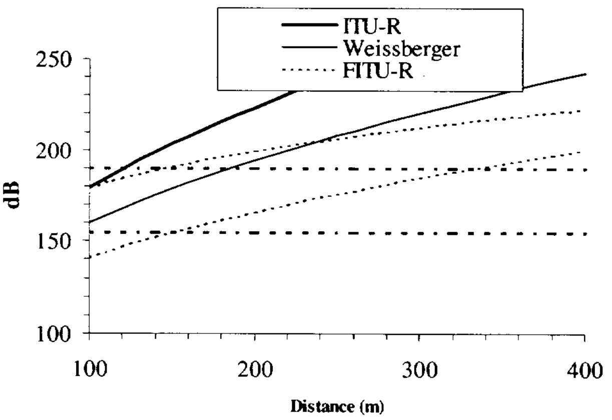


Fig. 7 Prediction models for vegetation loss at 38 GHz.

These are the two frequencies we have selected. This selection is partly justified by the availability of COTS link products which would simplify our construction (and at the same time could be reused for future purposes). The other justification is that the availability of bandwidth requires the move to frequencies above 10-15 GHz. In addition, the EW-protection generally increases with frequency. The 38 GHz choice is most likely unusable for our system, but we wanted to contribute to the research on foliage loss at higher frequencies. We chose to use narrow beam antennas in order to be able to perform measurements at greater foliage depths than previous known measurements.

With our measurement system we expect to be able to register total propagation losses up to about 210 dB at 18 GHz (190 dB at 38 GHz). From the models we find an anticipated measurement range from 200 to 400 m at 18 GHz, and from about 100 to 200-300 m (depending on foliage state) at 38 GHz. But the predictions disagree largely, and the models are not supported by previous measurements combining such high frequencies and foliage depths.

A series of field measurements are planned in Norwegian terrain with different vegetation, both conifer, deciduous and mixed forests. We intend to measure at different weather conditions and seasons, considering the effects of precipitation in the air as well as rain, snow and foliage on the trees.

In addition to measuring the propagation loss using narrow band transmission we intend to perform some wide band measurements to study not only the mean attenuation, but other statistical parameters such as the fading characteristics and the delay spread. For this we will use a channel sounder with a bandwidth adjustable up to 200 MHz.

Due to our use of narrow beam antennas we expect to find a flat and fast fading channel, at least at bandwidths below 100 MHz. Since future WLAN system probably will have capacities limited to 34 Mbps, wide band measurements will primarily be performed at a bandwidth of around 30 MHz. This gives a maximum propagation loss of 175 dB at 18 GHz (155 dB at 38 GHz). According to the models we should be able to perform wide band measurements at foliage depths of about 150-300 m at 18 GHz (Fig. 6) and 100 m at 38 GHz (Fig. 7).

7 MEASUREMENT SYSTEM

A large effort has been put into the system in order to achieve a high degree of automation. With narrow antenna beams (1-2 degrees) and line of sight obstructed by vegetation, a system for automatic antenna positioning is required. Based on compass, Global Positioning System (GPS) for precise positioning and inclination meters for compensating non-levelled vehicle/antenna, the system is able to point the receiver antenna towards the transmitter antenna (and vice versa). But to compensate for a limited accuracy the system will have to search a certain sector (both in azimuth and elevation) to find the (strongest) signal.

With a 2-axis electrical rotor to control the antennas this process takes a number of seconds. The process may be speeded up using phased array antennas. Also, the use of such antennas will reduce the need for mechanical antenna adjustment and will be able to quickly compensate for vehicle motion which otherwise could cause link dropout. A small study on phased array antennas for our application has been performed by SINTEF Telecom and Informatics [7]. The conclusion is that it is possible to realise such an antenna, with many technical and practical advantages, although the cost may be too high. Depending on extended commercial use of such technology the cost may fall in the future, making use of such antennas more feasible for our purpose.

Fig. 8 Our equipment mounted in an old military all-terrain vehicle (Volvo) - exterior view.

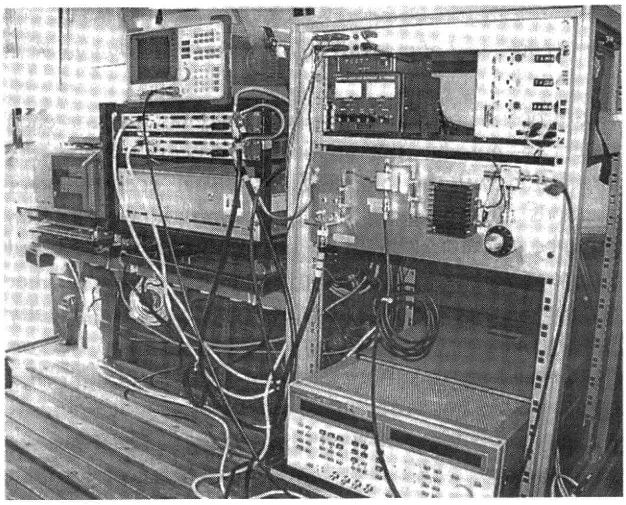


Fig. 9 An interior view from our RX measurement vehicle.

Fig. 8 shows a photo of one of our two measurement vehicles. On roof top we see a generator (220V), a magnetic compass and an electrical rotor with two microwave units with antennas

Fig. 9 shows the interior of our receive vehicle, but without the channel sounder to be mounted in the front rack for wide band measurements. Fig. 10 shows the outline of our measurement system with all major components.

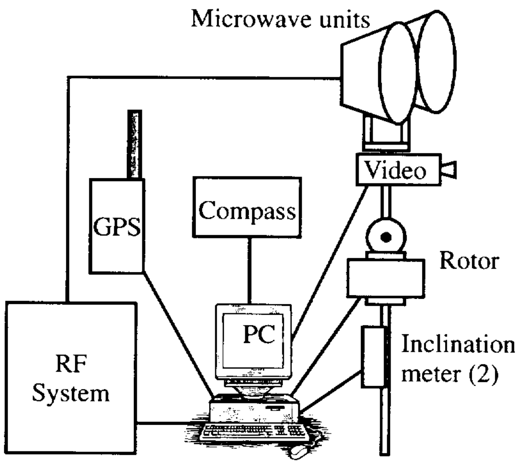


Fig. 10 Schematic view of the measurement systems.

Fig. 11 and Fig. 12 show the RF parts of our measurement system, both for the transmit and receive vehicle.

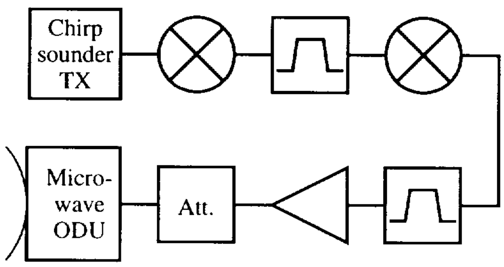


Fig. 11 RF-part of measurement system - TX vehicle.

The process of pointing the two antennas towards each other is not an uncomplicated process. Since the vegetation will be blocking for visual contact, other methods must be used. First, the approximate direction must be calculated based on information from GPS and compass and information on the position of the other vehicle. In case the vehicle is not positioned horizontally we also use information from the two inclinometers to compensate for this. With a limited accuracy of both GPS (military, but not differential) and the magnetic compass we may not achieve an overall accuracy that matches the antenna beam width. Due to this, there is a need for both vehicles in turn to scan a larger angular area to locate the strongest signal.

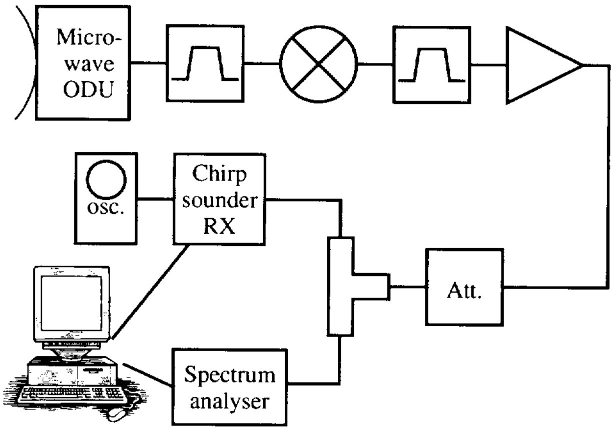


Fig. 12 RF-part of measurement system - RX vehicle.

8 RESULTS AND CONCLUSION

The discrepancy between the different propagation attenuation models is rather large. In addition, the models have not been verified for the combination of high frequency and large foliage depth that we require. Due to this, we found it necessary to perform our own propagation measurements to determine whether or not our proposed WLAN concept is worth further attention.

At the time of writing (end of May 1999) our measurement field campaign is about to start and no results are yet available. But we hope to be able to present some preliminary results at the conference.

According to both [2] and [3] we expect to find results that give a smaller attenuation than that indicated by the ITU-R model [1]. Also, if the results of [5] may be extended well above 2.5 GHz, we hope to find results indicating that even the Weissberger model is too conservative for our purpose in Norwegian terrain. In addition to measurements of the mean attenuation, other statistical parameters of the channel will be evaluated, such as the fading characteristics and the delay spread. With a wavelength the size of leaves and sprigs we expect to find a fast fading channel, even at calm days.

Based on the assumption that the measurements indicate a usable system link range of 300 - 500 m we may state a preliminary conclusion: The military requirements for a command post communications system (LAN) probably are best met by two different systems, unless considering other concepts based on airborne (or satellite) relays. The most mobile users will have to confine with systems in the UHF or lower SHF band with a data rate (well) below 1 Mbps. But with omnidirectional antennas such a system offers high mobility combined with an acceptable range (typically 1 - 2 km at a data rate of approximately 100 kbps). Users requiring higher data rates are well served by our proposed system based on directional antennas at 10-20 GHz, offering several Mbps but supporting lower mobility. In addition to WLAN, much of the concept may also be applied to LOS links for wide area system, and a combined system is not unthinkable.

References

[1] CCIR study programme IA/5, "Influence of terrain irregularities and vegetation on tropospheric propagation". Report-236-6, 1986

[2] WEISSBERGER, M.A.: "An initial critical summary of models for predicting the attenuation of radio waves by foliage", ECAC-TR 81 101, Electromagnetic Compatibility Center, Annapolis, USA, 1981

[3] AL-NUAIMI, M.O. and STEPHENS, R.B.L.: "Measurements and prediction model optimisation for signal attenuation in vegetation media at centimetre wave frequencies". IEE Proc.-Microw. Antennas Propag., Vol. 145, No. 3, June 1998

[4] GRØNNERUD, O.K. and OMTVEIT, B.: "Propagation measurements in the UHF band", FFI-NOTAT-93/7043 (In Norwegian)

[5] OMTVEIT, B.: "Wideband propagation measurements in the UHF band", FFI-NOTAT-94/01720 (In Norwegian)

[6] BERG, T.J., "The use of 802.11 DSSS LAN for command post communications", To be published by FFI, 1999 (In Norwegian)

[7] JENSEN, I., HOLE, B.S. and HOLTER, B.: "Antenna study for wireless command posts", STF40 F99008, January 1999, (In Norwegian)

Spread spectrum and Mobile/Tactical - Radio/Wireless Communications

Dr. Davras Yavuz
NATO C3 Agency, Netherlands (Formerly SHAPE Technical Center)
Communications Systems Division - Radio Branch
P.O.Box 174, Den Haag 2501 CD, The Netherlands
Tel: 31 70 314 2422, Fax: 31 70 314 2176, yavuz@nc3a.nato.int

Abstract/Summary

Spread spectrum techniques are one of the primary technologies that have driven the proliferation of radio (or wireless) communications. However, these spread spectrum techniques were originally developed for overcoming noise/interference effects. Subsequently they were the enabling technology of for space communications providing processing gains essential for long distance space communications. More recently these techniques have provided, and continue to provide, one of the essential ingredients for the proliferation of mobile and multiple-access communications.

Shannon's fundamental work laying the foundations of information theory has as the corner-stone spread spectrum theory and we will first cover briefly bandwidth expansion bounds for spread spectrum systems derived from Shannon's capacity limits. These show the exponential improvements possible through expanding the bandwidth of a baseband signal for transmission. In spite of the fact that they do not provide any hints on how that could be achieved, these results are significant for testing known systems. An interesting conclusion from these is that, for example, single-side-band suppressed carrier systems are optimum from this perspective.

There are many electromagnetic propagation related effects; reflection, diffraction, scattering, multipath, large & small-scale fades, delay spread effects. These effect all radio/wireless communications and specifically mobile links (=at least one end of the link is rapidly deployable and/or moving). With mobility additional issues also arise from limitations in antenna size, pointing stability if directionality is required, absorption/scattering effects in forest/foliage, extreme changes in local ground conductivity conditions, characteristics of the mobile platform/vehicle,... etc. Spread spectrum techniques, in general, reduce the detrimental effects of most of these, to some degree, through primarily the averaging effects inherent in the use of a large bandwidth for transmission. Some these will be exemplified and elaborated in the context of tactical/mobile radio systems, including conventional frequency hopping systems and such systems as JTIDS, GSM, TETRA, Iridium, wireless LANs, ...

Introduction

The first "spread spectrum concept" is generally credited to Armstrong for his frequency modulation invention in the thirties. The idea that one would intentionally increase the frequency spectrum of a signal was considered "strange" at the time but eventually the wisdom was recognized. Some authors use the term "spread spectrum" to refer only to digital coded spreading techniques. We will use it generically to any system that increases the bandwidth of the information signal for transmission. Spread spectrum techniques were originally developed for overcoming noise/ interference effects. Subsequently they were the enabling technology of for space communications providing the large processing gains essential for long distance space communications. More recently these techniques have provided, and continue to provide, one of the essential ingredients for the proliferation of mobile and multiple-access communications.

Shannon's fundamental work laying the foundations of information theory has as the corner-stone spread spectrum theory and this will be covered briefly first. These indicate the exponential improvements

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

possible through expanding the bandwidth of a baseband signal for transmission.

Shannon: Bandwidth Expansion for S/N Improvement[1]

Claude E. Shannon, with his classic work in the fifties, not only laid the foundations of what we now call ***"information theory"*** but also illuminated the way for many related technological innovations. One of the most important of these is related to his well known ***capacity theorem*** which basically states that, for a given channel, a coding scheme exists such that information can be transmitted with no errors if the transmission rate is below the (Shannon) capacity. From this fundamental result we can derive a relationship between ***bandwidth expansion and signal to noise ratio (S/N or SNR)*** which has subsequently formed the basis for spread spectrum implementations. Although as mentioned above, others before advanced the idea that S/N improvement can be obtained by intentionally expanding the transmission bandwidth of a system, it was his fundamental mathematical analysis that really set the stage. Derivation of these results is quite simple (based on the assumption of additive Gaussian noise as disrupting effect of the channel) by considering the capacities of two channels with bandwidth fs for the information signal and BW for the transmission channel. The resulting relationship is as follows and is shown in Figure 1.

$$
\begin{aligned}
SNR_o &= \left[1 + SNR_i\right]^{\gamma} - 1 \\
&= \left[1 + SNR'_i \,\frac{f_s}{B}\right]^{\gamma} - 1 \\
&= \left[1 + SNR'_i \,/\gamma\right]^{\gamma} - 1
\end{aligned}
$$

Where SNR_o is the output signal to noise ratio, SNR_i and SNR'_i are the actual and normalized input signal to noise ratios and γ the bandwidth expansion factor (BWEF) B/f_s, with f_s denoting the information signal bandwidth and B denoting the transmission bandwidth. We should note the exponential improvement with BWEF. From Figure 1 we note the extremely large possible improvements of output S/N ratio as the bandwidth expansion factor is increased for the same input S/N. For example we note that for a BWEF of 20 (top dotted curve) we can have a system that would provide an output S/R of 30 dB (1000 times) for an input S/R of 0 dB (1). It should be noted that this result is an "existence" proof just like the Shannon capacity result on which it is based on; it simply states that such an improvement with bandwidth expansion is possible but does not indicate how that can be achieved. However, with the passage of more than forty years since the original publication of Shannon's results and enormous developments in information/coding theory, we are now in a position to construct codes that approach very closely the Shannon Limit and therefore the improvements shown in these curves.

There is a simple intuitive explanation for this improvement for the direct sequence spreading (DSS, also called DS-SS) case. With DSS, to increase the bandwidth of a low data rate signal we multiply it by a high rate code or chipping signal. Take for example the information space as 8 bits (i.e. $2^8 = 256$ symbols as for example in the ASCII set); we spread each symbol, when transmitting, to 16 chips with a suitable code, doubling our bandwidth. There are $2^{16} = 65536$ code words of length 16, of these only 256 correspond to the symbols of our alphabet. By choosing this very small set of 256 codewords (0.4 %) from the large space of 65536 judiciously, i.e., well separated in code space (= large Hamming distance), our correlator receiver can tolerate a lot of errors and still decide on the correct received symbols. Note that we could add another orthogonal symbol set of 256 characters and still only be using 0.8 % of 16 bit/chip[2] codewords. This is the basis for code division multiple access (CDMA) which is intimately related to Shannon's results. In practice, expansion factors used are usually even greater. For example in the Joint Tactical Information Distribution System (JTIDS) in use with all AWACS/NAEW aircraft and related ground systems since mid eighties (and soon to be introduced into other platforms), the DSS expansion is from 5 bits to 32 bits/chips. The system also provides up to 128 different orthogonal CDMA nets (alphabets of 5 bit codes). This corresponds to $2^{(5+7)} = 2^{11} =$

[1] This section of the paper is from a presentation by the author at the AFCEA 1997 Ankara Conference

[2] "Chip" is normally used for *bits in actual transmission* and "bits" for *symbol alphabet*, but they are also used interchangeably. The context should clarify.

2048 which is < 0.00005 % of all possible 32 bit/chip sequences. The Global Position System (GPS) implements such a large level of DSS that signals 30 dB below noise level can actually be utilized. All space communications are also based on these DSS techniques where spreading factor even larger are used in communications with spacecraft exploring the solar system.

Direct Sequence Spreading and Frequency Hopping

Direct sequence spreading (DSS) and frequency hopping (FH) are the two basic methods of expanding the bandwidth of an information signal. Both techniques, individually and together provide significant advantages from the viewpoint of propagation in mobile scenarios. Great majority of multipath effects are relatively narrow-band and both these techniques can greatly reduce their effect. Frequency hopping has an additional advantage that may sometimes be useful; if the frequency separation between hops is sufficiently large it can provide improvements that are due to the "frequency diversity" effects.

Hybrid systems which combine the two also exist (JTIDS) and much current work in ongoing in this area. The results presented above are, strictly speaking valid only for the DSS case. With pure FH a basic narrow band signal is hopped over many frequencies to gain multiple access facilities and some advantages against interference, jamming. DSS was affordable only for expensive space and military applications until recently due to the costs of high-speed correlator receivers. With the PC revolution driving the technological developments required by the market place this is no longer true and correlators for many hundreds of megachip rates are now readily available. Digital signal processing (DSP) chips have also improved tremendously so that such correlation operations can be easily implemented. DSS has one significant disadvantage and that is the **near-far problem** in the CDMA application. In order for a correlator receiver to work efficiently the received signals must be approximately the same level. A strong signal from a nearby transmitter, even though it does not have the correct spreading code for the receiver net in question, can swamp the desired signals. For this reason, CDMA applications for mobile operation must have some sort of power management (also called dynamic power control/management DPC/M) for satisfactory service. This added complexity is the primary cause of the continuing debate comparing time division and code division multiple access (TDMA/CDMA) technologies. TDMA is simpler to implement as it does not, from a receiver operation viewpoint, require dynamic power management but has a lower theoretical capacity. However, DPC/M eases frequency reuse, lowers hand-set power requi-rements (something TDMA systems also have to do to make efficient us of small batteries) and improves privacy by setting the power levels to that required for a good performance and no more.

Propagation Mechanisms

There are a number of basic electromagnetic energy propagation mechanisms. Spread spectrum techniques reduce/remove detrimental effects that can exist in most of these as we shall discuss.

a) Free space propagation (perfect line-of-sight - LOS)
b) Surface wave propagation (energy follows a surface which has some conductivity e.g. sea water)
c) Reflection (propagating waves impinge on objects that are much larger than the wavelength – ionosphere, land masses, mountains/hills, buildings, walls, ...
d) Diffraction (radio path obstructed by a surface with sharp, irregular edges – waves bend around even with no LOS)
e) Scattering (objects smaller than the wavelength – foliage, fence/lamp posts, ...) Also troposcatter, meteor scatter and ionoscatter phenomena which generally result from natural phenomena (man-made ionospheric heating is an exception).

When at least one end of the radio link is moving we then also have effects due to the changing nature of the path. In fact, even when both ends are fixed there are many similar effects due to the movement of electromagnetically significant elements in the path such as the ionosphere, meteor trails, airplanes, trucks, cars, foliage movement in wind, ...etc. In general when mobility is involved the channel can vary greatly with time and location resulting in very complex propagation modes. There will be significant multipath effects due to signals travelling over differing paths (also multipath scattering from relatively nearby objects) and many other shadowing, diffraction and attenuation effects. These can result in large and rapid fluctuations in the received power levels. These fluctuations can also be described as fast fading although this term is more appropriately used for long distance, for example ionospheric multipath effects. Random frequency modulation due to varying Doppler shifts over different multipath signals and time dispersed versions of signals (echoes) are some other significant phenomena that must also be taken into account. Spread spectrum techniques, especially when combined (DSS & FH) are highly effective for overcoming most of the detrimental effects of the above.

The ranges over which the delays of the multipath signals occur are termed delay spread, Figure 2. In a typical mobile cell-phone (e.g. GSM) scenario the delay spread can range from about 0.2 μsec for open rural areas to 4 μsec in highly constructed urban/city centers).

A simplified, generic look at outdoor propagation in terms of attenuation is given in Figure 3. The path loss between the transmitter and receiver is proportional to the square of the distance in the simplest free space (LOS) case and can increase up to typically the fourth or even 4.5 th power of the distance.

Indoor propagation[3] introduces additional attenuation and complications due to the propagation path crossing walls, floors and scattering, ducting effects. Path loss is in general proportional to the distance between transmitter and receiver, number of floors and walls the signal traverses (including loss per wall/floor factors which may or may not be the same), other metallic scattering/reflection effects. Some relatively simple formulas for estimating first cut path losses exists but measurements are generally necessary for good overall system performance especially for marginal locations. The tendency to use larger powers in, for example, office wireless LANs to get good performance has to balanced with possible hostile interception threat in a nearby parking lot or other location.

Penetration of signals into buildings (outdoor to indoor transmission) is a complex function of many factors; frequency, building parameters (materials, orientation, layout, height, window area and type of glass, ...). Received signals generally improve with height (assuming everything else is the same) both due to antenna cleaqrance effects and reduction of urban clutter at upper floors. In general, normal glass windows offer about 6 dB less loss as compared to walls but coatings/films for security and ultraviolet screening can reduce this figure.
People moving around in buidings effect propagation much more than in outdoors and multipath induced attenuations of 10 dB due to such indoor movements have been measured. Overall values for the rate of signal attenution from outdoor transmitter to indoor receiver could be between 3 to 6.5 th inverse power of distance.

Some examples of models and results will be given in the presentation (Internet is an excellent source for up to date information: an Altavista search for GSM and "propagation models" results in 160 hits most of which are relevant).

Performance

As with most digital systems, the most significant performance effects are due to fades that can induce synchronization loss and various doppler effects. As the on-the-air bit period approaches delay spread error rate can increase significantly. Average duration of fades are first order function of frequency, speed of mobile and fade depth. For example at 900 MHz (e.g. GSM) with a mobile speed of 2 Km/hr (slow walk) and fade depth of 20 dB the average duration of fades is typically around 25 ms, if the mobile moves at 50 Km/hr the average fade duration becomes around 1 ms. In general, for digital voice communications bit error rates (BER) of 10^{-3} or better are sufficient, for data communications 10^{-6} or lower, is nominally accepted as sufficient.

From electronic protection measures (EPM)[4] viewpoint the two spreading techniques, FH and DSS, have significant differences. In principle, both

3 Indoor to indoor : Both transmitter and receiver indoors or Outdoor to indoor : Transmitter outdoors receiver indoors

4 We use EPM to cover both electronic counter measures (ECM) and electronic counter counter measures (ECCM).

force the adversary to spread the available jamming signal over a much wider band, reducing its effectiveness. However, with FH, if the hopping rate is slow and there are a small number of hopping nets in a given LOS volume, following the hopping signals (per net) with narrow band follower-jammers may be possible. With DSS signals it may be possible to focus on vulnerable parts of the signal spectrum in terms of sychronization of preambles, ... etc. However, all these complicate the task of the adversary and in most tactical scenarios simply delaying the "denial of service" type attacks is sufficient.

Conclusions

The fundamental aspects of spread spectrum techniques originated with Shannon's pioneering work about half a century ago. The two basic class of techniques, frequency hopping (FH) and direct sequence spreading (DSS) provide significant advantages, not only in the well known processing gain sense, but also from radio propagation aspect, especially in mobile scenarios. The primary advantage is in terms of reducing the effects of multipath and related propagation anomalies. There can also be some beneficial diversity effects, especially for channels which are highly time varying and difficult to characterize. Both FH and DSS offer such advantages but through rather different technical routes. Hybrid systems that combine FH and DSS can potentially offer a wider range of advantages both in terms of overcoming propagation anomalies and denial of service (e.g. jamming) attacks.

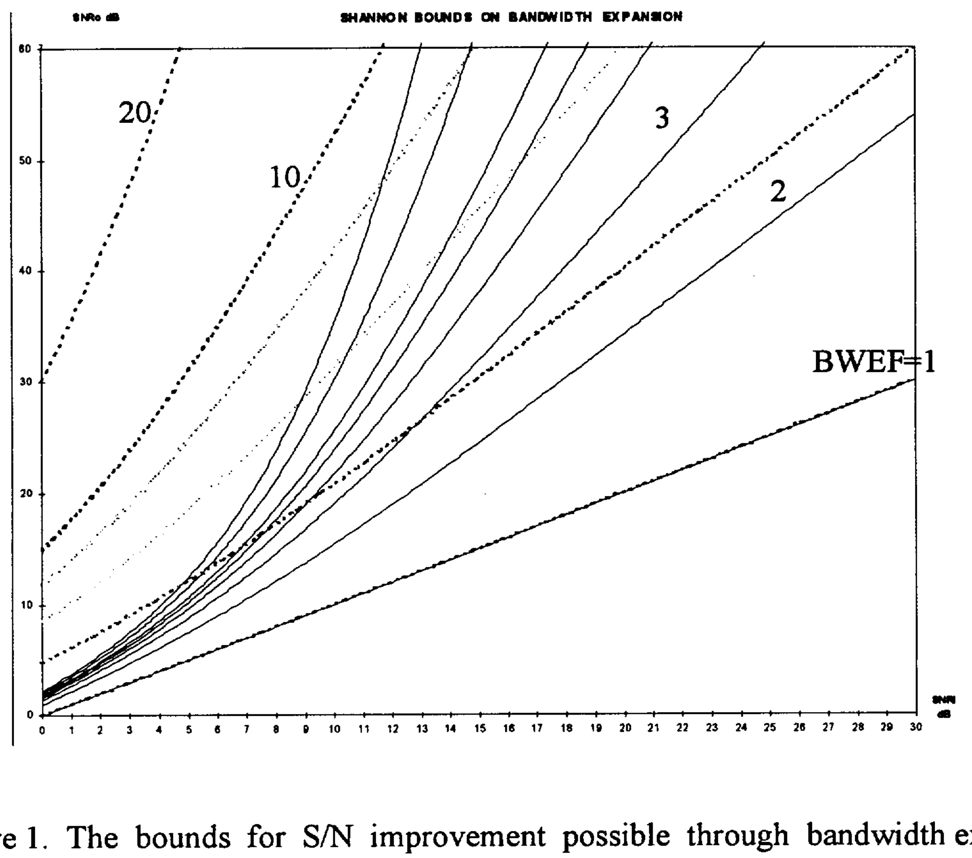


Figure 1. The bounds for S/N improvement possible through bandwidth expansion derived from Shannon's capacity theorem. The straight line is for BWEF =1 followed by curves for 2,3,4,.... Dotted lines are for actual S/R (SNR_i), solid lines for normalized S/N (SNR'_i).Note that horizontal axis is SNR_i for solid curves and SNR'_i for dotted.

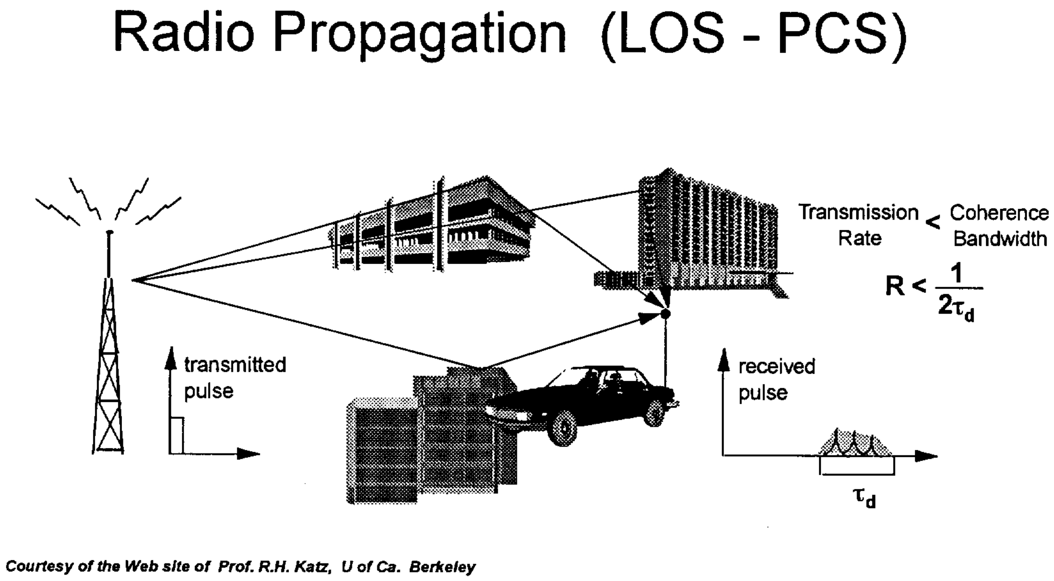


Figure 2. A representative "line-of-sight" (LOS) type scenario applicable to PCS system in an urban setting indicating the time spreading effects of multipath. With motion (mobiles and/or moving propagation impact elements) doppler effects can also be present.

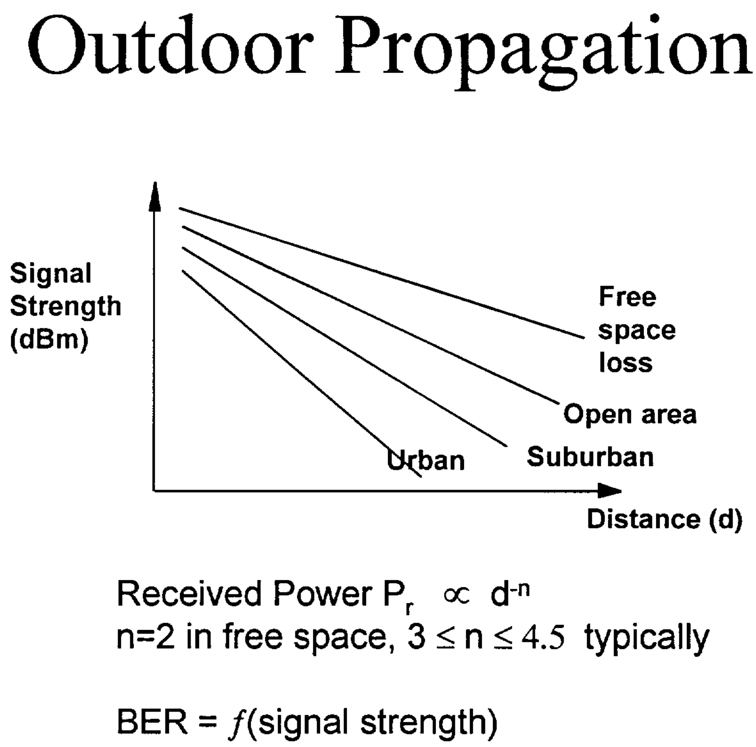


Figure 3. Typical (simplified) attenuation characteristics ranging from the well known inverse squared free-space loss characteristics to possibly as high as inverse 4.5th power in urban environments.

Tactical Mobile Radiocommunications – A Challenge for Military Frequency Management

P. Bingel, E. Trautmann and T.J. Taylor
NATO HQ C3 Staff / Frequency Management Branch
Boulevard Leopold III, B-1110 Brussels, Belgium

INTRODUCTION

The increasing demand for information at all levels of the military places a great demand on tactical mobile radiocommunications and the supporting radio frequency spectrum resources. Military frequency managers have been able to operate using agreed civil/military provisions of the frequency spectrum in close coordination with Civil Authorities. It appears, however, that civilian radiocommunications users are demanding more and more spectrum for high value commercial purposes. Civilian operators have been challenging the reservation of spectrum for critical military systems for several years and this became particularly evident at World Radiocommunication Conferences (WRC). The military community can expect more of such challenges in the future. These challenges often mean seeking specific agreements for spectrum based on specific systems. This leaves little room for new military technology growth and even less for increases in spectrum demand to meet growing information transfer requirements. This challenge increases rapidly when considering that efficient operations are dependent on flexibility, mobility and interoperability while being employed across international borders. Embedded in these challenges, however, are opportunities to counter deficient military spectrum requirements through the utilisation of civilian commercial systems.

NATO's Military Application

The applications of North Atlantic Treaty Organisation's (NATO's) military force both today and for the 21st Century are defined. The threat of Global Warfare has diminished. Although collective defence against an attack on any Alliance member remains NATO's core mission, the potential to conduct Peacekeeping, Peace support and other non-article 5 military commitments has expanded and is a lot more likely. The difficulties in the employment of military power are compounded, as reductions in size are actualised due to the reduced threat of global warfare. NATO however, has developed new doctrine and tools to accomplish its new tasks with fewer resources. The new tools range from the new Combined Joint Task Force (CJTF) doctrine to new command control and information systems and ultimately, also, to new weapons. These new tools are highly dependent on access to the radio frequency spectrum to achieve success.

The CJTF is a multinational (Combined) and multi-service (Joint) task force. The CJTF must have the capability to exercise command and control of any mission assigned to a unit of up to the size of a Corps. NATO's planning calls for the ability to employ two land component CJTF's and one Naval CJTF simultaneously. The "combined" challenge to frequency management is encountered through the coordination of frequencies across thousands of kilometres and several international borders to accommodate multi-access theatre-wide operations. This challenge is further magnified as now these combined forces are not restricted to just NATO nations as NATO is now "combining" with Partnership for Peace (PfP) Nations. The challenge to the frequency manager is evident, as frequencies that have not

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

previously been harmonised with NATO must be incorporated into the force plan.

The "Joint" benefit of the CJTF is that it creates synergy through the selective application in the capabilities of the "joining" of forces to effect an overall greater result than the single services could generate independently. This "joint" flexibility enables a commander to select more or less of any type service dependent upon each mission. This also places a high demand for flexibility within the frequency manager's tools as the radio frequency needs of army, air force and navy components differ significantly.

It is evident that command and control, and co-ordination is required by the multifaceted CJTF. The CJTF Commander is highly dependent on seamless information flow, which in turn enables the force to be flexible, mobile, and have interoperability. Additionally it helps limit own force vulnerability. Remarkably, it is through the radio frequency spectrum that audio and video information is transmitted in the form of analogue and digital data, that sensors (radar) gather new information, navigation data is received, electronic warfare is conducted and that weapons systems are employed.

NATO's Military Radio Frequency Spectrum Requirements

Military radio frequency spectrum requirements have been defined to ensure NATO forces have the capability to execute the full range of Alliance missions. The diversity of the military spectrum requirements provides a great challenge for the military frequency manager.

Military requirements vary between activities and countries. In some countries military spectrum requirements may exceed those of the NATO military harmonised bands published in NATO's Joint Civil and Military Frequency Agreement (NJFA). In other countries, for the time being, military requirements do not exceed the spectrum range of the current harmonised bands defined in the NJFA. In general, the harmonised military bands provide the common military frequency resource which:

- Allows systems to operate in a common framework.
- Facilitates combined/joint exercises and training.
- Allows allied forces to operate their equipment on foreign NATO territories.
- Includes frequency spectrum for day-to-day employment, training, exercise, and combat readiness.
- Supports electronic warfare training without interfering with civil frequency applications on own or other nations territory.

When NATO exercises or projects its military forces, even with the assistance of highly sophisticated computer aided planning tools, the coordination required for military frequency management remains a huge task which requires unique understanding of each service.

Army CIS planning principally addresses the Corps level; in some cases detailed requirements are described down to lower echelons. Tactical and operational exchange of information between NATO land forces and coalition or PfP army forces are to be based on having available spectrum. Although the same for navy and air force, the army communications architecture requirements must address:

- Command and Control (C2) within and between NATO and PfP forces.
- Border coordination.
- Non-contiguous operations.

Additionally, the **Air Force** units accommodate:

- Flying units such as squadrons and wings with 20-40 combat aircraft with organisational and technical support required.
- Ground mobile air defence units, especially anti-air artillery with surface to air missiles of various types and ranges.
- Various means of air transportation and long range logistics support.

Maritime forces can comprise a variety of command, transport and warships which, in a tri-service scenario or amphibious operation, would operate rather closer to the shore than independently on open sea. Maritime forces may also add all the requirements of the air force and army forces into their operations.

Assessment of Frequency Requirements for CJTF

Taking into account multinational employments of NATO and non-NATO forces, under worst case conditions composed of two land and one sea CJTF simultaneously. The following standard frequency requirements are assumed for a single CJTF:

1. To accommodate a variety of HF/short wave communications, using a standard 3 kHz channel with 1 kW ERP, employment of up to 800 HF channels is required.

HF frequencies used for long distances require international registration in accordance with International Telecommunications Union (ITU) regulations. Also protection would be desirable under the umbrella of national civil administrations/frequency management authorities. However, the highly mobile nature of the CJTF and breadth of potential force application makes such planning difficult, as location of system employment is seldom predictable.

Further to this the availability of pre-coordinated, multinational frequency pools contributes to spectrum economy and supports the readiness of forces as well. For this purpose nations have agreed on a common NATO HF land pool, which provides HF frequencies for NATO and participating Partnership for Peace countries.

2. The second typical CJTF spectrum requirement is in direct support of all tasks in and around the deployment area, known as Combat Net Radio (CNR) VHF applications, which increasingly include ground/air communications.

The standard version of CNR has a highly mobile character with single channel or adaptive, automated channel selection and frequency-hopping features on the basis of 25 kHz bandwidth. In total this requirement would comprise 1000 single radio channels.

3. CJTF spectrum is also dedicated to support Air/Ground/Air (A/G/A) functions and various navy communication requirements. In NATO, these essential A/G/A requirements are met in the first place, in the military frequency band 225-400 MHz, also known as the NATO UHF band. For a CJTF deployment the present allocation which serves day-to-day training and readiness, however, would not be sufficient to support the additional communications requirements

In fact, large scale air operations, even without the incorporation of ground manoeuvre or helicopter assets rely on the full availability of that harmonised military band. Future expansion by civil systems into this band would severely constrain military operations.

4. CJTF operations are also reliant upon military satellite applications. It is not difficult to predict that the frequency requirements for communications, navigation, identification and intelligence satellite services are increasing strategically, operationally and tactically in utilisation. These systems are not only becoming more numerous but also designed to handle ever increasing "stove pipes" of specialised information.

5. Terrestrial Radionavigation, Identification and Distress and Safety functions also exist in radio frequency spectrum ranges that have encountered extremely high density of use. The availability of European or NATO wide TACAN-channels and identification subbands (or frequencies), the unconstrained use of the Global Navigation Satellite System (GNSS), in particular of the Global Positioning System (GPS; Subsystem to GNSS), the Instrument Landing System (ILS) and radionavigation radars cannot be renounced.

6. The sixth CJTF standard requirement supports radiolocation services. A considerable number of modern radar, exist that are stationary, mobile and airborne and use the typical radar bands between 1 GHz and 36 GHz. Due to the broad radio frequency spectrum range of these systems, maintaining harmonised military radio frequency bands without sharing is nearly impossible. In these cases military frequency managers must ensure new equipment is designed and built to be flexible to regional frequency requirements by having alternate modes of operation.

7. The last but not least, Telemetry and Data Link applications, such as Multifunctional Information Distribution System (MIDS), as well as connectivity with NATO Airborne Early Warning, Tactical Reconnaissance Aircraft and Remotely Piloted Vehicles (RPV) are increasingly important for NATO air operations.

A consideration that is not directly related to the CJTF is the incorporation of new technologies into our weapons and weapon systems. This enables our military forces to achieve their objectives with minimal combat casualties and with little or no collateral damage. With few exceptions nearly every modern piece of military equipment depends on access to the radio frequency spectrum. Adequate radio frequency spectrum usage is critical to maintaining and improving our military capability.

CJTF Military Requirements -- Synopsis

For NATO the standard military equipment currently in use will remain the backbone of military operations for most of the next several decades. This military equipment provides everything from rapid feed back for tactical support to logistics. System design makes them practical for routine frequency management. In most cases the radio frequency spectrum that they operate under is deeply imbedded in the traditional harmonised NATO frequency bands. Nevertheless frequency management for these systems, as described earlier remains a key task. The new technologies that are making the modern military forces successful from the command and control perspective are radically different from standard military communications. These systems require large spectrum allocations (bandwidths) to accommodate everything from state-of-the-art data links to video teleconferencing capabilities. The advantage to frequency management that these new systems offer is that they transmit using commercial radio frequency spectrum. As the systems were designed commercially the frequency management is, in general, coordinated prior to military adaptation of the system.

Military frequency managers must maintain as their key responsibility, for periods while home-based, diligence towards internationally harmonising spectrum access. When successful their efforts will further be included in the NJFA and the European Table of Frequency Allocation and Utilisation and form the foundation of NATO's position for World Radiocommunication Conferences.

Additional Consideration – The Fallacy of Peace Dividend

There is a fallacy regarding radio frequency spectrum requirements of NATO forces. Since many countries have downsized their forces many people believe that proportionally less spectrum is required to conduct operations. This opinion is false for primarily three reasons:

First, fewer operators do not equate to a reduced number of systems required to conduct operations. The draw down in numbers of personnel is absorbed, in part by reducing the number of personnel in everything from a combat squad to a Corps. Although there are reductions of total number of personnel the number of fighting units may remain the same. Therefore the number of radio networks required to relay information remains the same. In the air force this reduction can be seen in squadrons downsizing from 18 aircraft to 12 aircraft. However, the number of tactical communications frequencies required remains unchanged.

Second, in today's military, technology is used to maintain the fighting advantage with equipment replacing personnel for performing missions. These technologies, as previously discussed, are heavily dependent on traditional and new military frequency bands to operate.

Third reason is that the radio frequency spectrum required for large deployments and exercises has not changed from that which has been historically demonstrated. It is arguably true that day-to-day need for training is not as demanding on available spectrum as previously demonstrated. Similar to reason 1, above, however, the final training phases of joint/combined operations requires the same interference free access to harmonised military spectrum as ever.

It is acknowledged that some of the radio frequency bands may see less utilisation on a day-to-day basis. The equipment that utilises those bands and the operators that train on those systems however, still require the same spectrum access. As stated, radio frequency planning for modern military coalition forces must support highly mobile, deployable multinational major force units which are able to deploy throughout NATO Europe or beyond with their Air, Land and Maritime components. Training and exercising each of the systems used by the Air, Land and Maritime components is required for the respective units' survival.

Challenge Brought on by the Civil Radio Telecommunication Technologies

Many international organisations, commercial and civil are using the electromagnetic spectrum to provide valuable services worldwide. The mobile communications and broadcasting markets are exploding into the information age. Access to the spectrum is considered critical to enable the technologies of these two markets, in particular, to expand.

The trend is quite clear and industry as well as network providers are ready to follow it. One forecast, based on sound statistics, indicates almost an explosion of mobile communications. It is said that within the next 3 years Global System For Mobile Communication (GSM)-networks will increase to 80 million subscribers in Europe with a further augmentation of up to 180 million subscribers by the year 2005. The expansion of GSM is predicted to reach 700 million subscribers worldwide within the next 8 years. It is obvious that these developments will certainly have displacement effects on the frequency spectrum, as current allocations for GSM will be barely sufficient to keep pace with the rapid growth of the subscribers. Already attempts have been made to free spectrum for the next generation of systems at the expense of military allocations. I am happy to say that for the time being these attempts failed. But I have my doubts as to whether the military will be successful in maintaining its position.

Re-emphasising from above: mobile telecommunications and multimedia will definitely govern a large portion of the radio frequency spectrum of tomorrow. This is a consolidated expert's view leaving no room for any doubt. And GSM, the example above, is only one kind of application out of a large range. During a recent international symposium a representative from industry described the forthcoming situation in world-wide telecommunications with the following key elements:

- Integrated and mixed terrestrial and satellite based mobile telecommunications.
- Multimedia-communications spanning across an unlimited number of different networks independent from a specific location or organisation.
- Global satellite communications with 2 – 5 million subscribers.
- An enlarged number of GPS-based services, such as public travel guidance and road navigation, fleet management of transportation services (trucks, taxis etc.), theft prevention of privately owned vehicles, etc.
- Internet in the sky.

- More than a total of 300 GSM networks installed in more than 200 countries.
- Ratio between mobile and stationary communications reaching 50 %.
- Mobile communication assets will become "least cost" mass products.

What this development means in terms of additional spectrum is easy to estimate and, in fact, first approaches to this end have already been observed for some time. Maybe WRC-95 could be regarded as the beginning of this new frequency battle. At that time the United States first undertook to seek a majority for a resolution in support of their new satellite communications system called TELEDESIC which is presently under development by U.S. industries. It was certainly a surprise to the WRC participants when the US request for additional frequency spectrum for TELEDESIC was suddenly a key item at the Conference although it was not scheduled on the agenda. Just to give you an idea of the system's magnitude: If the development schedule and industrial planning remains unchanged about 288 satellites, so-called «Little LEOS» (Low Earth Orbit Satellites), will be placed in orbit in a constellation of 12 planes around the earth. They will provide communications services all over the world. Looking at the communications networks in the Third World countries, particularly the developing countries, satellite communication systems might offer them a reasonable alternative to their current terrestrial systems. These terrestrial systems have insufficient capacity and therefore need very expensive extensions and technological upgrades to match their growing communications needs. I believe that was why the U.S. received broad support from the Third World although the matter was not pre-coordinated nor was it based on a common proposal.

As expected, the need for more spectrum to accommodate new Big, Little and Mega Low Earth Orbit (LEO) systems was also one of the most important, and controversial, items on the agenda of the subsequent WRC in 1997. The controversy was caused partly by the strongly felt need of many delegations to protect existing services occupying the bands targeted for additional allocations, and partly by competition within the mobile and fixed satellite service operators. Each of the fixed satellite service operators were keen to gain an operational advantage for their own particular system. As a matter of fact, the provisions for space services, in particular enhancements to the Mobile Satellite Services (MSS), also affected a number of frequency bands with military applications, mainly those used for tactical radio relay systems below 1 GHz and for GPS navigation in the area 1.5 –2.5 GHz.

The challenge by civil radio-telecommunications technologies to the military is brought about due to competition for scarce spectrum resources and a sharp rise in the amount of radio frequency spectrum congestion. Given the GPS – MSS example, above, it is clear to see that in some radio frequency spectrum bands demand exceeds supply. The "supply and demand" competition in some spectrum regions has generated competition for radio frequency spectrum, which is resulting in competitive spectrum pricing.[1] Although spectrum pricing has not been brought to every country it is considered a trend that is inevitable and will be the next challenge to military radio frequency spectrum as it is driving civil users to search for new and less expensive spectrum ranges.

Regardless of national implementation, spectrum pricing has brought with it a high political profile. This high political profile has gained international recognition in the ITU, the Conference of European Posts and Telecommunications Administration (CEPT) and the Inter-American Telecommunication Commission (CITEL). These organisations, along with the spectrum pricing efforts, are attempting to alleviate the stress of spectrum congestion by finding additional ways to bring

[1] Spectrum pricing, Licensing and Spectrum Auctioning are technically different types of spectrum management tools; however, for the sake of this paper they will all be referred to spectrum pricing.

the supply of radio frequency spectrum into balance with the demand. This is done by:

- Encouraging the development and use of more spectrum efficient equipment.
- Encouraging the return of spectrum that is not required, as a cost saving measure.
- Motivating consumers to move to less congested areas of the radio frequency spectrum.

As a matter of fact they all apply also to military use of the spectrum. The last bullet in particular provides a great challenge to the military frequency manager. It is also closely tied to returning spectrum that is not required. Due to the often irregular use of the radio frequency spectrum in some military frequency ranges, commercial mobile communications and broadcasters that are searching for less congested radio frequency spectrum will eventually turn to military frequency bands and seek government support for broader application of their use.

Future Trends in "Going Mobile" - Opportunities and Challenges

Global satellite communications systems such as GSM, Iridium and IMT-2000 offer great capability for expanding communication to the benefit of society. The military are also taking advantage of these systems as much as possible. The tactical advantage of getting immediate communications with advanced units is considerable, even down to the individual soldier who can provide instantaneous status reports to a command post that is located far behind the visible horizon.

The trend of "Going Mobile" however, is at the heart of commercial encroachment into harmonised military bands. The encroachment is a threat to military systems such as Radio Relay, Air/Ground/Air, Air Defence Radar/Navigation Systems. Civil emergency services, such as Trans European Trunked Radio (TETRA)/TETRAPOL, are already established in the NATO UHF band.

Encroachment on military GPS/Radionavigation Satellite Service (RNSS) by the MSS system is a risk. During the last WRC an attempt was made to expand MSS. The proposed expansions would not only provide overlap with the already implemented GPS system but also interfere with the proposed expansion of GPS.

As expected during the preparations for the WRC and highlighted in the NATO Military Position Paper, discussions at the WRC 97 really became controversial on proposed reallocations of spectrum to MSS in the 1.5 GHz range currently reserved for radionavigation services such as the GPS. GPS provides military and civil capabilities, which are crucial to NATO's current military engagement in their peacekeeping mission in the Former Yugoslavia and also in the future. Deep concern was expressed about the proposal to allocate further services in this band, which would create severe degradation of GPS functions. When the controversy reached its peak, the United States intervened at highest level possible. In a letter from the White House, a strong appeal was made to Heads of States and Governments of all NATO nations to join them in opposing the spectrum sharing proposal. Finally a compromise was reached at WRC 97 that no allocation to Mobile Satellite Service was included in the Table of Frequency Allocations but, based on a Resolution of the Conference, further electromagnetic compatibility studies would be undertaken in order to resume the issue at the next WRC.

It is clear that as the civil community uses up their available spectrum they will turn to traditional military bands for new service opportunities. The challenge presented to the military frequency manager is to justify sole service priority for use of the frequency band but also the additional challenges for deployment as nations established sharing exceptions based on their unique frequency spectrum needs. It is the sincere individual efforts on the national level within the NATO Nations, which works against the frequency manager attempting large-scale coordination efforts. National sharing locally results in forced civil-military sharing internationally which in the long run leads to loss of "on demand" spectrum access as required by a highly mobile international military force.

Summary

The challenges of military frequency management are great and change is happening quickly. NATO's Frequency Management Sub-Committee is working hard to address the issues. As the legal situation for frequency management is different in every country the influence of military frequency experts also varies significantly. This is the reason why direct cooperation between civil and military frequency managers is so important in the work of the NATO Frequency Management Sub-Committee and its Working Groups.

Additionally, liberalisation of telecommunications, the globalisation of frequency demands, and technological developments make it more and more difficult for the military to maintain adequate access to the spectrum. We are competing with the growing number of commercial customers asking for new assignments in order to accommodate their extended services. There is also an imminent danger that military power is regarded as secondary in nature as a consequence of developments in the security environment which lead to cooperation instead of confrontation and a strong feeling that the risk of war is over. NATO's current engagement in the SFOR-mission to preserve peace in the former Yugoslavia, however, reveals that the risk of war is not over. Nevertheless the call for a "peace dividend" is becoming louder and louder in many countries. Frequency bands used for military purposes are sometimes looked upon as gold mines, which offer sufficient reserves to satisfy the growing demands for additional frequencies, needed for civil applications. To counter this trend it is absolutely necessary that military frequency managers demonstrate in the most proficient manner possible that frequencies allotted for military use are not wasted nor that they exceed the absolute minimum required to maintain an appropriate level of security for a nation. In the future military authorities will have to justify their property rights on the spectrum and that military occupation of frequencies does not hamper civil users unless strictly unavoidable.

In other words: in order to be able to credibly defend their position vis-à-vis the challenges of the liberalised market military frequency management must be as proficient as the civil administration, with the same degree of expertise and equipped with the same modern frequency management tools as their civil colleagues. Military frequency usage must follow the same rules and procedures in terms of efficiency and flexibility. Sharing of frequency resources without commercially damaging interference will also be mandatory for military applications. The exclusive use of a frequency band for military applications will be the exception. The standard rule will call for flexible and economic sharing of frequency bands.

To deal effectively with this situation military frequency management will have to cope with similar tasks as their civil colleagues and follow the same pattern in their work, namely:

- Definition of minimum frequency requirements for the various military functions
- Development and maintenance of a flexible frequency assignment concept leaving enough room for civil applications during peace time
- Ensuring electromagnetic compatibility and non-interference of military radio transmissions
- International frequency harmonisation through cooperation within NATO and with CEPT / EC / ITU
- Open contributions to and qualified advisory support during the development of frequency allocation tables
- Control and administration of assigned frequencies
- Supervision and monitoring of frequency usage.

The military frequency manager must understand that there is relatively less challenge to the military frequency management when war occurs, as in the case of NATO Article 5 operations. In this situation the military would, in accordance with national legislation, assume control of the radio frequency spectrum required. Even with the full understanding that civil radiocommunications would have to be restored, similar to the restoration of water and electrical services, frequency management would be controlled by the needs of the military. As well the commercial systems may be vital to the military communications. It is in non-Article 5 operations, however, such as peacekeeping where the challenges arise. In the non-article 5 operations coordination prior to the utilisation of any system is mandatory to avoid harmful interference with existing services, often unregistered, including civilian radio and television broadcasting stations. The IFOR/SFOR theatre stands as the most significant example with over 44,000 frequency coordinations required to be performed since inception. Therefore, it is critical that the military frequency manager remains vigilant in his quest to harmonise key military bands internationally and continually strive to maintain expertise.

Digital Communication and Multipath Propagation

Christoph Zehntner, Christian Patry, Markus Maier
Defence Procurement Agency, Electronics and Optronics Division
Stauffacherstrasse 65, CH-3000 BERN 22

Abstract

Multipath propagation with tactical VHF radios is investigated in a mountainous terrain. Depending on transmitting frequency, echo time delay and receiver signal bandwidth, distortion in analog and intersymbol interference in digital systems is caused. The behavior of digital fixed frequency (FF) and frequency hopping (FH) radios is examined.

Methods are described to measure multipath propagation in the terrain as well as to simulate it in the laboratory. The analysis of the multipath performance of the new Swiss tactical radio SE-235 led to a modification to improve the synchronization characteristics.

1. Introduction

With the change from analog to digital VHF communication systems, the awareness of multipath propagation in the Swiss central plain and alpine terrain grew continuously. Whereas analog Frequency Modulation (FM) radios only are impaired by distortion of the voice, with digital radios high bit error rate (BER) and synchronization problems are caused.

Considerable work has been done worldwide to investigate multipath behavior for cordless and wireless telephone systems as the Global System for Mobile communication (GSM). The paper presented here mainly deals with tactical VHF radio propagation.

Measurements of the propagation paths in the real countryside have to be carried out. This is a basic requirement in order to define and to describe the operational situation. Simulation is needed to analyze systematically the multipath effects on the radios as well as to perform comparative measurements.

2. Multipath Propagation

Multipath signals result from obstacles in the path of the signal propagation, i.e. obstacles in the terrain where the radios are used. If an obstacle has dimensions in the order of the transmitted wavelength it can interfere with the signal propagation. VHF-obstacles are e.g. large buildings, woods, rocks, hills and mountains. They attenuate, bend and/or reflect the transmitted signal. The reflected signal is here called **echo** in accordance with the acoustic echo. **Fig. 2.1** displays the general situation.

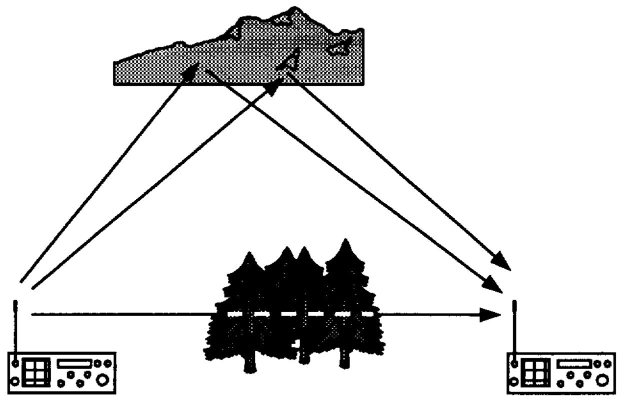


Fig. 2.1: Transmitter, Receiver, woods in the direct path, mountain as obstacle.

At the receiver site the signal is composed of several single signals that result from the various propagation paths. The combining signals have different amplitudes and phase differences among one another. So the resulting signal varies from slightly more (signal reinforcement) to considerably less (signal cancellation) compared to a single signal.

Echo Geometry. In order to estimate the possible echo signal delays it is assumed that the distance r between two radio sets is between 2 and 20km, the distance h from the baseline to the obstacle varies 2...30km, **Fig. 2.2**. The detour of the echo signal is then calculated to about 2...60km.

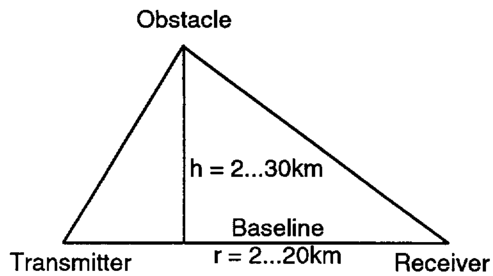


Fig. 2.2: Echo Geometry

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

This theoretical assumption corresponds to a time delay of the echo to the direct signal of 3...200μs.

Echo signal properties. What are the properties of an echo that can disturb a radio signal at the receiver site? The echo signal is added or subtracted from the direct signal, depending on its phase relative to the direct signal. With the assumption of constant amplitudes of the direct and echo signal as well as a constant time delay of the echo signal the superposition of the two signals at the receiver site is only dependent on the transmitting frequency:

- A signal maximum is obtained when the two signals are in phase, i.e. when the delay time τ is a multiple of the signal period: $\tau = n*1/f$, n = 1, 2, 3, ... or **f *τ = n, frequency-delay product.**
- A signal minimum occurs when the signals are of opposite phase: $\tau = 0.5*n*1/f$, n = 1, 3, 5, ...

Therefore the signal of echo propagation is periodic in frequency over the whole frequency range with a period of $1/\tau$. In a real situation the propagation condition changes with frequency and the periodic property of the signal is only valid in a partial frequency range. Of course, the multipath contributions form a fading pattern not only in frequency but also in space. **Fig. 2.3** shows the frequency dependence of a signal disturbed by one echo at different relative echo levels. This is the diagram a spectrum analyzer would produce at the receiver site.

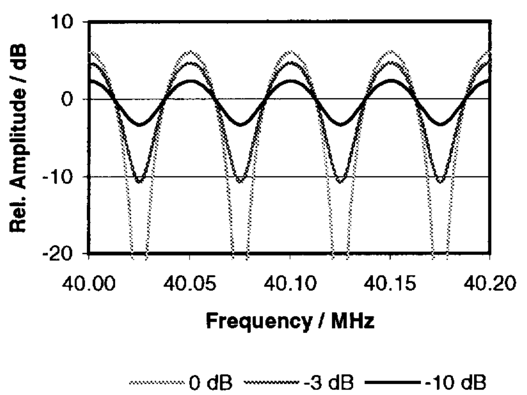


Fig. 2.3: Frequency dependence of a signal disturbed by an echo, delay time = 20μs

The signal amplitude at the receiver depends on the amplitude relation of the echo to the direct signal, the echo level. Usually only echo levels ≥ -10dB disturb the communication link.

How much a radio communication is disturbed by an echo signal depends first of all on the relation of the receiver bandwidth to the echo period ($1/\tau$).

A **short-term echo** has an echo delay time τ that is short compared to the reciprocal of the receiver bandwidth B: $\tau << 1/B$. With bandwidth B = 25kHz, τ << 40μs or $\tau \leq$ 4μs. It can also be said that the coherence bandwidth of the channel is wider than the receiver bandwidth. The minimum notch or hole in the received signal is broad compared to the transmitter bandwidth. Such echoes originate from near obstacles as large buildings or hills.

Fig. 2.4 shows an example of a 4μs echo. A radio working in channel A will not, in channel B will be disturbed or disrupted. The echo signal nearly cancels the direct signal entirely.

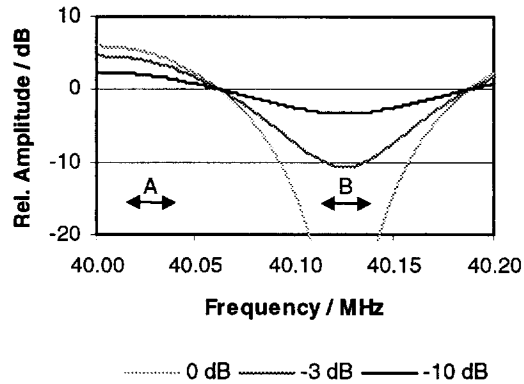


Fig. 2.4: Short-term echo, delay time = 4μs

In a fixed frequency (**FF**) link the short-term echo in antiphase (minimum) attenuates the signal whereas the in-phase echo does not disturb, independently of the modulation principle, both analog and digital. As the entire channel is attenuated or amplified it is arbitrary whether the channel is usable or not. In order to evade the signal cancellation the frequency-delay-product can be changed by a factor of 0.5. This can be done by displacing the location of the antenna by half the transmitting wavelength (1.7 - 5m for a frequency range of 88 - 30MHz) or by changing the transmitting frequency by half the reciprocal of the delay time (125kHz for 4μs).

When either the transmitter or receiver or both are in moving vehicles, the relative phases of the multipath signal components change continuously and therefore so does the signal level. The signal fluctuates with a Rayleigh distribution.

With frequency hopping (**FH**) the signal hops constantly between good and bad frequencies. Some hops get lost but there will normally remain sufficient undisturbed channels. The effect of multipath is averaged out and FH is generally not susceptible to short-term echoes.

A **long-term echo** has an echo delay time τ that is long compared to the reciprocal of the receiver bandwidth B: $\tau > 1/B$. With bandwidth B = 25kHz, $\tau > 40\mu s$, e.g. $\tau = 100\mu s$. Now several minima will fall into the receiver bandwidth (RBW) and the signal will be distorted, **Fig. 2.5**. These echoes originate from large, distant obstacles as mountain chains. The coherence bandwidth of the channel is smaller than the receiver bandwidth, the delayed signals are decorrelated and do not contribute to the fade but to co-channel and intersymbol interference.

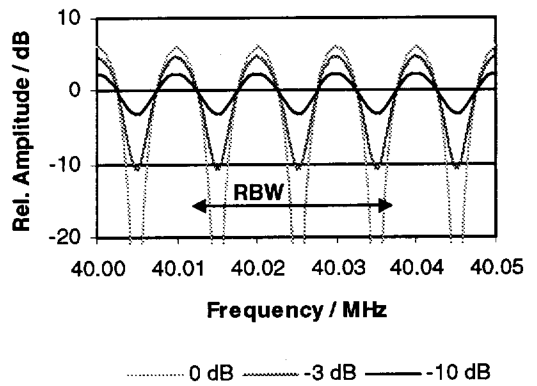


Fig. 2.5: Long-term echo, delay time = 100µs

This is usually a minor problem for voice communication over analog FM (frequency modulation) **FF** channels. The audio frequencies are well below the influence of these delays and the human ear is not susceptible to such distortions. With digital communication systems the bit detection is disturbed or impossible.

With **FH**, only the channels are disturbed that fall into the affected frequency range.

All these considerations are applicable when both, the direct ***and*** delayed signals are present.

Direct *or* delayed signals. The attenuation of the direct path as well as the delay time of the echo path are frequency dependent. Consequently at the receiver site for some frequencies the direct path may be dominant, for others an echo path. This is not a problem for FF systems. On the other hand with FH systems this leads to bit and/or synchronization loss when the difference between the delay times of different frequencies is greater than half the bit length of the transmitted signal. E.g. with a carrier bit rate of 25kBit/s (bit length = 40µs) and a delay time difference of greater than 20µs the delayed hops cannot be detected: the synchronization of the bit detection is out of phase. The impairment of the FH link is dependent on the number of delayed hops.

3. Measurement of Multipath Signals in the Terrain

In order to understand the effect of multipath propagation in practice it is important to know what happens in reality: what are the technical parameters of the echoes in our real terrain. These have to be measured if possible in the tactical environment where the communication systems are used.

The technical requirements of such an echo measuring system can be derived from the parameters of the communication system:

- Frequency range 30–88MHz (tactical VHF radio)
- Bit rate 20...25kBit/s → bit length ≥ 40µs: delay time resolution ≤ 20µs.
- Maximum time delay difference ≥ 200µs (according to experience).

Several **methods** to measure a real propagation channel are known, e.g.:

- Impulse channel sounder: Carrier frequency impulses allow a direct analysis of the echo signals in the time domain.
- The frequency and phase dependence of the transfer function of the propagation channel is measured with a continuous wave sweep and after a Fourier transformation the analysis in the time domain is possible.
- The linear detection of transmitted signals allows preserving the amplitude and time relationship of the channel. The correlation function of a suitable data word represents the impulse response of the propagation channel.

All methods have been used, most recently the last mentioned as spin-off of transmission trials with Direct Sequence Spread Spectrum signals. The measurement setup is made of off-the-shelf laboratory instruments as displayed in **Fig. 3.1**. The carrier frequency of a RF synthesizer is phase modulated with a defined data word and amplified to 40W. The broadband rods of tactical VHF radios are used as transmitting and receiving antennas. At the receiver location the inphase and quadrature (I/Q) components are measured with a vector signal analyzer. The known transmitted data word is then correlated with the demodulated signal sample. This yields the signal amplitudes and relative delay times of the different propagation paths. A computer on both sides controls the instruments and does all the calculation off-line.

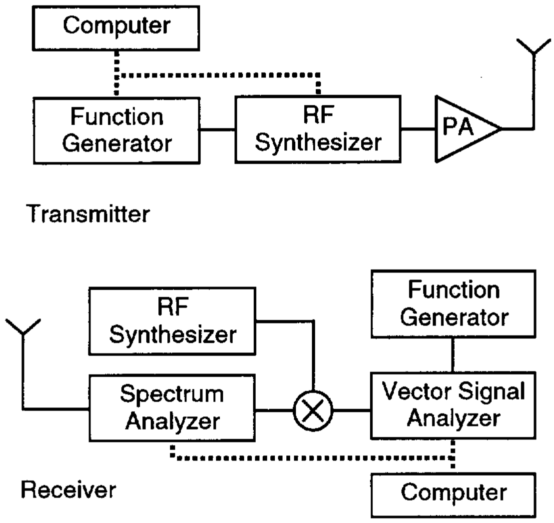


Fig. 3.1: Setup for multipath measurements in the terrain.

The transmitted data word has a length of 1024 bit for good autocorrelation properties (i.e. one dominant maximum). This determines the dynamic range of the measuring system, which is about 20dB. The data rate is 250kBit/s and thus the time resolution is 4µs (1/[data rate]). Though the sample rate of the receiver is 1MHz the resolution is not augmented but the graph is smoothed.

The system does not measure absolute time delays between transmitter and receiver location but delay differences. The maximum measurable time delay between two paths is 2ms (0.5 * [word length] / [data rate]). This is far enough, 200µs being the maximum delay that has been measured up to now.

The **result of the measurement** is shown in a relative amplitude / time diagram for each frequency channel, **Fig. 3.2**. The first arriving signal is considered as the direct signal path and automatically set to 0µs. The echo signals appear subsequently as peaks with a height corresponding to their relative amplitude. The largest of the direct and echo signals is set to 0dB. In this example two large echoes with a delay time of 63µs and 69µs are registered.

In the next example, **Fig. 3.3**, the echoes for three frequencies are displayed, the transmitter and receiver sites being the same. It is impressive to discover how the relative amplitudes of the direct and the echo signals change according to the working frequency.

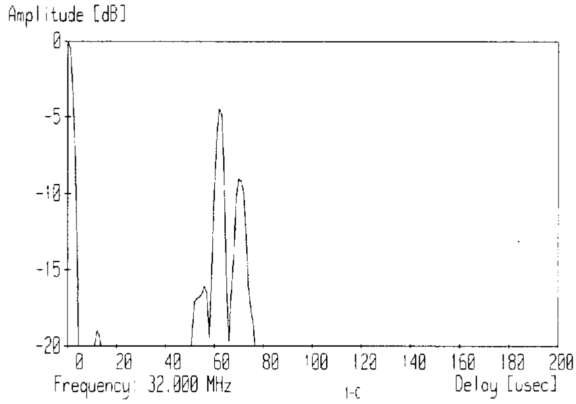


Fig. 3.2: Relative signal amplitude / delay time diagram for a single frequency channel.

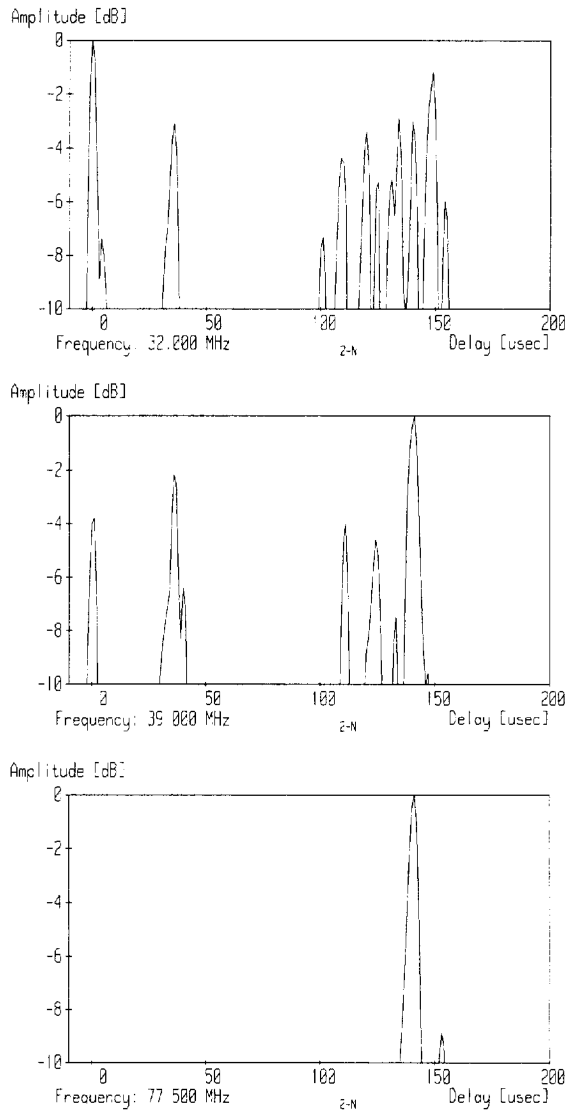


Fig.3.3: Direct and echo signals at different frequencies for the same transmitter – receiver location.

With **FH** radios the propagation path interests in a whole frequency band rather than in a single channel. Therefore the measurement equipment can be tuned to subsequent frequencies automatically. Start- and stop-frequency as well as frequency step can be defined. On each frequency the channel propagation is measured and the result is stored. The measurement of one frequency takes 6s; the whole band from 30 to 88MHz with half MHz steps, 12 minutes.

The result is presented in a delay time / frequency chart: For each frequency the signal with the largest amplitude is displayed with "x" and the signals with relative amplitudes within -5dB to the largest, with "o". The absolute amplitudes of the largest signal at each frequency vary in the displayed examples between 10 and 30dBμV.

Fig. 3.4 shows an example of a communication link of 3.6km. The direct path (delay time = 0μs) is dominant in the frequency range of 33...36Mhz. In the whole frequency range large echo signals occur with a delay time of about 60μs. Above 40MHz the direct path is practically no more existent. The echo signals are usually not reflected from one single spot in the terrain but from a mountain chain. Therefore the reflection components appear in the diagram as clouds, while the direct path as a line. The gap around 50MHz is due to a TV station transmitting near the test site.

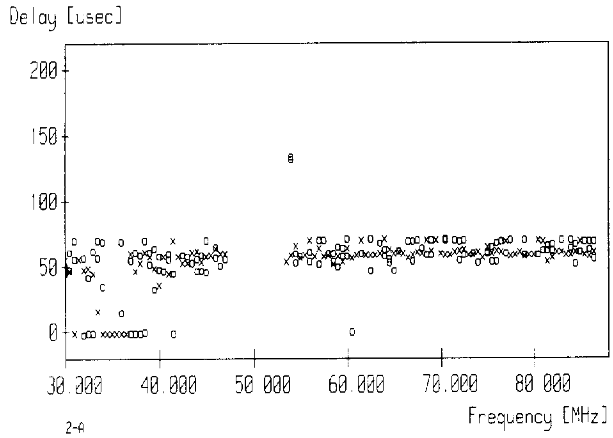


Fig. 3.4: Two path propagation in VHF band, "x" = largest signal per frequency, "o" = within -5dB.

Fig. 3.5 shows an extreme example of multipath propagation. It is situated in central Switzerland, near Lucerne, with the two radio stations at a distance of 2.3km. Six different propagation paths can be figured out with delay times of up to 200μs.

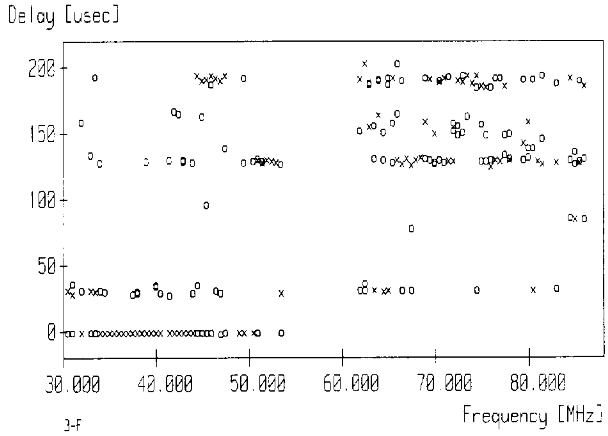


Fig. 3.5: Multipath situation with 6 distinct propagation paths, delay times up to 200μs, "x" = largest signal per frequency, "o" = within -5dB.

An other impressive situation is presented in **Fig. 3.6**: Here the strongest signal occurs in wide frequency ranges to equal parts on the direct as well as on the echo paths. Under these circumstances FH communication is not possible because half of the hops can not be demodulated synchronically.

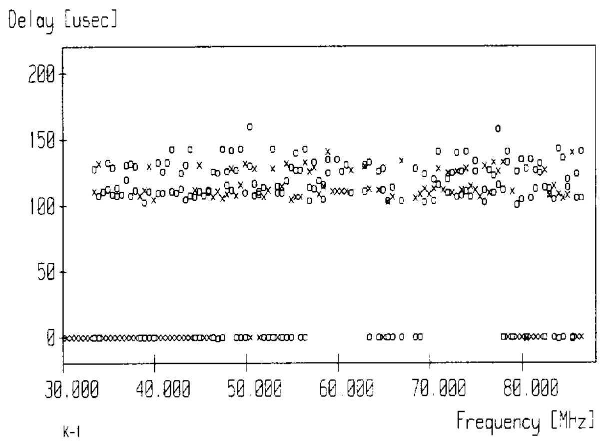


Fig. 3.6: Multipath with equally distributed signal amplitudes on direct and echo paths, "x" = largest signal per frequency, "o" = within -5dB.

4. Multipath Simulation in the Laboratory

For the understanding of the communication system performance in a multipath condition environment, a laboratory simulation in a mock-up is necessary. The following advantages are obvious:

- Reproducibility: it allows making repeated trials under the same conditions.

- The comparison of equipment of different manufacturers, types or operational conditions is possible.
- Modifications and improvements in hard- and software can easily be controlled.

The RF channel simulator system used - the Echo Simulator - has been custom-developed by the Swiss company CIR in 1991. The Echo Simulator has four modes: HF, RAYLEIGH, V/UHF and FH. It allows simulating propagation channels with reflection, fading and Doppler properties. Up to three propagation paths in a FF channel or two paths in a FH band can be defined. The channel properties are digitally processed in a baseband with a sampling rate of 2MHz and a resolution of 12 bit. The technical data of the V/UHF and FH mode are:

- Bandwidth 300kHz
- Delay time 0 ... 1020μs
- Delay resolution 0.5μs
- Relative level -46 ... +20dB
- FH: change of channel settings synchronized to hopping signal (max hop rate: 2000Hop/s)

The block diagram of the Echo Simulator is shown in **Fig. 4.1**.

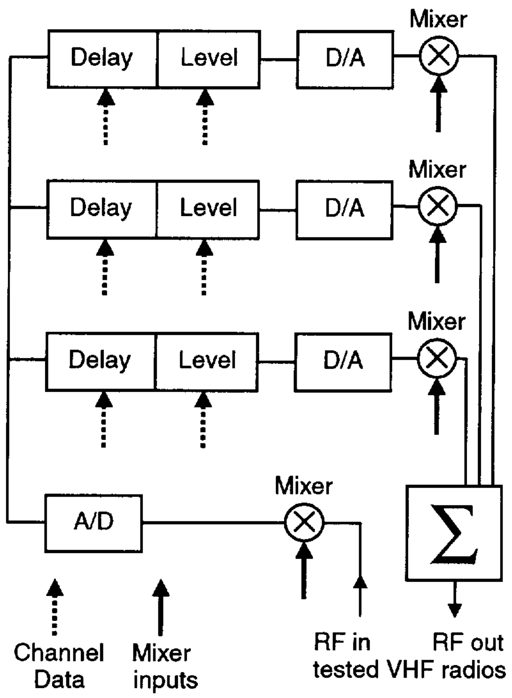


Fig. 4.1: Multipath Echo Simulator

Today there are also commercial systems available that allow the simulation of propagation channels for FF and FH modes.

The parameter settings are menu driven and can be stored for subsequent use. The frequency mixing is derived from a single source for stability reasons. Appropriate filtering to avoid interference products may be indicated.

In order to get reproducible results, the tuning position of the simulated multipath signal has to be controlled relative to the tested radio bandwidth. Usually the signal maximum (**Tuning Maximum**) or the signal notch (**Tuning Minimum**) is set to the center of the bandwidth.

The measured parameters to evaluate the multipath performance of a communication equipment are normally bit error rate (BER), synchronization capability as well as signal distortion or speech intelligibility.

Of course simulation does not replace a test in the terrain but it is a great help in understanding the results of the field tests.

5. Multipath Performance Analysis of the Swiss Tactical Radio SE-235

As is well known, the Swiss country is quite hilly and mountainous. The central plain stretches between the Alps and the Jura Mountain chain. These rocky and limestone obstacles are very well suited to reflect radio signals and to create multipath propagation.

During the evaluation of the new tactical VHF radio SE-235 of the Swiss army, multipath performance was a criterion. With laboratory and field tests the radios of different manufacturers have been investigated. The finally chosen model - the French Thomson-CSF PR4G - has been modified to improve the synchronization performance under multipath conditions.

6. Conclusion

The understanding of multipath propagation in the real terrain is mandatory for the construction and operation of VHF tactical radios. This is even more true in the future where higher data rates and multiple use of the frequency spectrum will be inevitable.

EMC Analysis on VHF Combat Net Radio (CNR) in Airborne Applications

K. S. Kho
NATO HQ C3 Staff
Frequency Management Branch
1110 Brussels, Belgium

1. Introduction

Usage of VHF Combat Net Radios (CNR) on airborne platforms will be contrary to the present co-ordination agreements with some NATO nations. For instance, in Germany, the military VHF band can be used without further co-ordination with the German Civil authority, provided that the restriction to e.g. the antenna height of 5 m is not violated. As a concession, use in helicopters is also permitted, but usage in fixed-wing aircraft requires additional co-ordination with the national authorities.

This paper is made to contribute to discussions on an appropriate balance between the interference risk from airborne use of CNR and the operational requirement. The analysis will first be based on co-channel interference. The results will then be extrapolated to cases where frequency separation plays a role and to operation in Free Channel Search (FCS) and Frequency Hopping (FH) modes.

2. Characteristics of a hypothetical CNR radio

The characteristics of ground or airborne CNR do not differ so much except that the output power and antenna gain of the airborne platform are lower than the ground CNR. The output power for the airborne is chosen to be 4 Watts and the antenna gain including feeder loss -15 dBi.

The relevant characteristics of ground CNR are as follows:

a. Transmitter:	
Frequency (used in the calculations)	: 44 MHz
Output power (vehicle (30 km) set)	: 50 W
b. Receiver:	
Frequency (used in the calculations)	: 44 MHz
Effective IF Channel bandwidth	: 25 KHz
Noise Figure	: 5 dB
Man-made noise level above KToB	: 10 dB
Selectivity:	
1st adjacent channel rejection	: 55 dB
2nd adjacent channel rejection	: 65 dB
4% f_c separation	: 120 dB
c. Modulation technique	: non-coherent FSK
d. Antennae:	
Gain of the antenna system	: 0 dBi
Height : 2 m	
e. Transmission bit rate	: 25 kb/s

3. Analysis

(1) Total noise level of a ground based and airborne CNR receiver

In order to assess the degree of interference, it is essential to calculate the total noise level at the CNR receiver. The receiver noise level is: 10 log (k.To.B) + Nf (dB) = - 125 dBm.

In case of:

a. Ground based CNR:

The man-made noise level is 10 log (k.To.B) + 10 = -120 dBm. The total noise level at receiver is then -119 dBm.

b. Airborne CNR:

The man-made noise level is 10 log (k.To.B) + 15 = -115 dBm. The total noise level at receiver is then -114 dBm.

(2) Propagation model

The Egli model was considered for the Ground - Ground (G/G) communications and the IF-77 Electromagnetic wave Propagation model for Air - Ground (A/G) and Ground - Air (G/A).

a. For G/G communications

$$\text{Pathloss(dB)} = 48 + 20 \log f(\text{MHz}) + 40 \log D(\text{km}) + Ht(\text{dB}) + Hr(\text{dB}).$$

where:

$Ht,r(dB) = 10 - 10 \log h(m)$ for $h <= 10$ m
$Ht,r(dB) = 20 - 20 \log h(m)$ for $h >= 10$ m
f = frequency in MHz
D = distance between Tx and Rx in km and
h = Tx and Rx antenna heights in m.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

b. For A/G and G/A communications

The IF-77 program can produce prediction of transmission losses in the format of Free Space Loss, IF-77 50%, 5% and 95% probability propagation losses. The values presented in this paper were obtained by using this IF-77 program. Only the 50% values are used. The 95% values do not differ much from the 50% values.

(3) Various Criteria of Reception

Various criteria are used in this report. The catagory of the criteria are as follows:

a. C/(N+I) = 20 dB

C/(N+I) = 20 dB represent a minimal C/(N+I) ratio for a good CNR reception. The value of 20 dB includes already some dB margin for antenna, antenna polarisation and additional propagation losses. The propagation loss used in this report is based on the 50% locations propagation loss. The threshold of a good reception for 16 kb/s CVSD signal in CNR is determined to be 10% BER. The relationship between BER and C/(N+I) ratio based on a non Cohehent FSK detection is shown in the following figure 3.1.

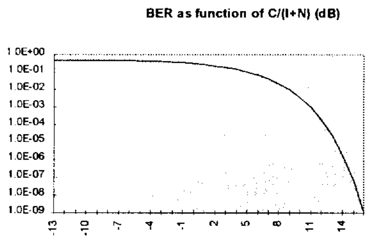


Figure 3.1 BER as function of C/(N+I) ratio.

b. C/(N+I) = 0 dB

C/(N+I) = 0 dB represents situations where the level of the interference signal is equal to the level of the wanted signal. In this case, the victim will suffer harmful interference.

c. C/(N+I) = 3 dB

C/(N+I) = 3 dB represent situations where the BER is still severe, of the order of 20 %. At this level of interference, the reception is still not acceptable.

(4) Radio Horizon Distances

The propagation loss in the zone where the communications are Line of Sight (LOS), follows the IF-77 model. The propagation loss is a much higher in the beyond LOS zone. Therefore, Radio Horizon Distance (RHD) is a parameter of interest in determining the propagation model applied in the calculations.

Radio Horizon Distance (RHD):

SQR(RHD(km)) = 2*k*R*h
RHD = 4.12 * SQRT(h)

where h = effective antenna height (m)
R = 6370 km -radius of the Earth (km)
k_{factor}= 4/3 (for west Europe)

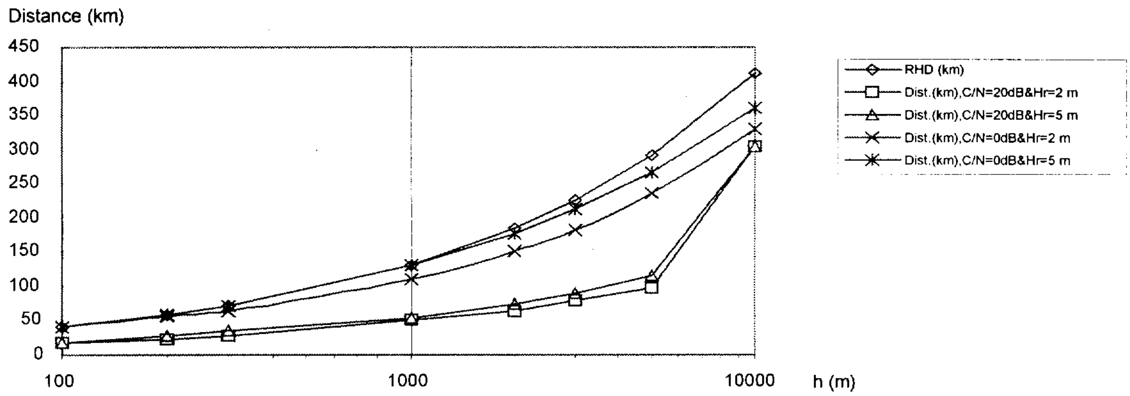
(5) Relations between antenna height and Radio Horizon Distance (RHD)

The probability that a CNR will interfere with another CNR link operated in the beyond LOS zone is assumed to be negligible. Using the above formulae, the following figure 3.2 shows the relationship between antenna height (aircraft altitude) and Radio Horizon Distance. In addition, the practical distances where the airborne CNR signal is still 20 dB above the noise level were calculated. The noise level of an airborne CNR and C/(N+I) of 20 dB were taken as basis.

$$Pr = Pt + Gt + Gr - Loss_{IF\text{-}77}$$
$$= 21 - Loss_{IF\text{-}77}\ \mathrm{dBm}$$
Noise level = - 119 dBm

C/(N+I) = 20 dB
$$21 - Loss_{IF\text{-}77} - (-119) = 20$$
$$Loss_{IF\text{-}77} = 125\ \mathrm{dB}$$

Using the IF-77 software, the corresponding distances for various altitude and Rx antenna heights were calculated. In a similar manner, the distances where the signal level is equal to the noise level (C/(N+I) = 0) were calculated as well. These results are displayed in figure 3.2 at the next page.



CNRs in airborne platforms are mainly used in helicopters at the altitudes up to 300 m.

Sub-Conclusion:

a. The area interfered by an airborne CNR at 1 km altitude is a circle with a radius of 130 km which is comparable to the surface of a battle field and it is much greater than the area interfered by ground CNRs.
b. If the link availability of the CNR nets should be 100%, the fixed frequencies used by airborne CNRs at I km altitude should be exclusive to the fixed frequencies used by the ground CNRs within the RHD of the airborne CNR used in the same operational area.

(6) Comparison of interference level versus noise and received signals

It can be seen in figure 3.2 that an airborne CNR transmitting at 4 Watts output power will produce a signal level greater than the noise level of the ground CNRs up to the radio horizon. Therefore, it will always degrade the sensitivity of terrestrial CNR links working on the same channel (co-channel), even from distances up to 360 km depending on the flying altitude. Therefore, frequency management between the airborne CNRs and the CNRs used by the land forces in the combat zone is essential.

(7) The interference from ground VHF CNR users to airborne VHF CNRs

The interference from ground CNR to the airborne CNR could be more harmful due to the fact that an airborne Rx will receive interfering signals from all interfering sources (many) within its Radio Horizon Distance.

a. On an airborne Rx receiving another airborne CNR

In the case of an air-air signals, the situation is even worst because the interfering ground CNR output power is 16 - 50 Watts and the wanted airborne CNR is only 4 Watts with a low antenna gain.

b. On an airborne Rx receiving a ground CNR

Using Free Space Loss, in the case of a single interferer, the co-channel interference from a ground CNR will start to degrade the airborne CNR reception (C/(N+I) < 20 dB) if:

Du / Dd < 10 (Degradation)

and the interference will be harmful if the C/(N+I) < 3dB which means if:

Du / Dd < 1.4 (Harmful interference)

Where: Du = distance unwanted CNR
Dd = distance wanted CNR

The above formulae are valid if the output power of the wanted transmitter equals the output power of the interferer. In principle both output power of the CNRs discussed above are equal, since they are both ground CNRs.

4. Results

The relation between the quality of a wanted link and the distances of the wanted and unwanted links are calculated with the above mentioned formulae. The output power of airborne CNRs are assumed to be 4 Watts. The possible situations are as follows:

a. Airborne CNR interfering G/G CNR communications

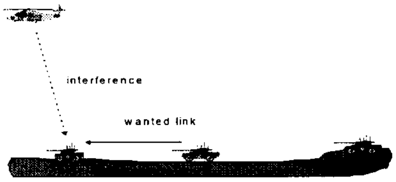


Figure 4.1 Schematic diagram of an airborne CNR interfering a land combat zone CNR communication

b. Ground CNR interfering A/A CNR communications

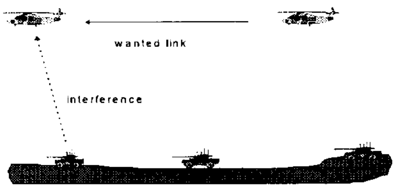


Figure 4.2 Schematic diagram of a ground CNR interfering an A/A CNR communications

c. Ground CNR interfering A/G CNR communication

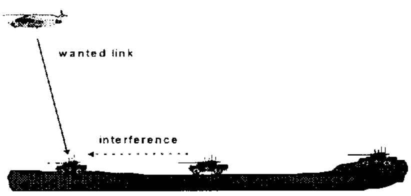


Figure 4.3 Schematic diagram of a ground CNR interfering an A/G CNR communications

d. Ground CNR interfering G/A CNR communications

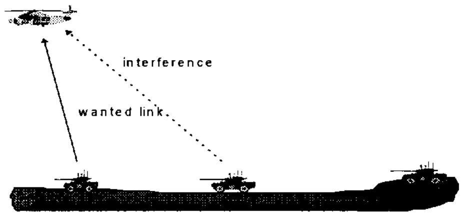


Figure 4.4 Schematic diagram of a ground CNR interfering an G/A CNR communications

(1) Airborne CNR interfering G/G CNR communications

NOTE: In the following figures of paragraphs 4.1, 4.2, 4.3 and 4.4 the geographical separation for various combinations of victims and interferers are calculated for C/(N+I) = 20 dB, the point where degradation starts and C/(N+I)I = 3 dB, the point where interference is deemed to be harmful. Therefore, the areas where the communications are interference-free or where they suffer harmful interference can be derived from the geographical separations displayed.

The required minimum distance d_{ag} resulting in C/(N+I) ratio of 3 and 20 dB for an airborne CNR interfering with G/G CNR communications is as follows :

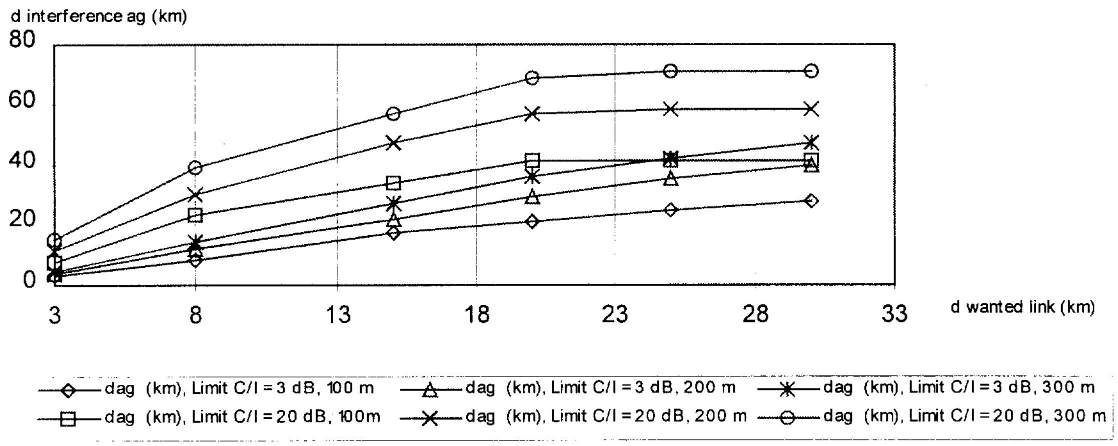


Figure 4.5 The minimum distance of an airborne CNR as function of the length of the wanted link of the G/G CNR communications in order to maintain a link quality better than 3 and 20 dB C/(N+I) ratio for flying altitudes of 100, 200 and 300 m.

(2) Ground CNR interfering A/A CNR communications

The required minimum distance d_{ga} resulting in C/(N+I) ratio of 3 dB and 20 dB for a ground CNR interfering with A/A CNR communications is as follows:

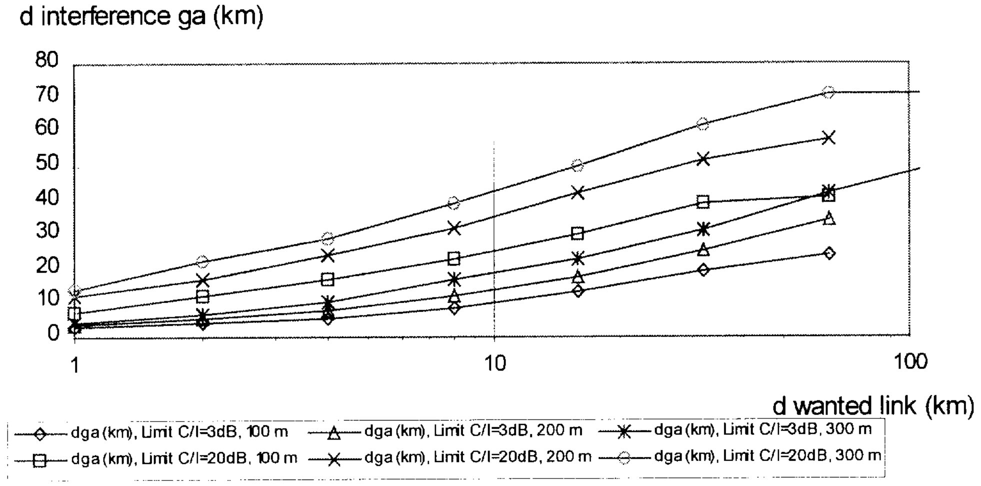


Figure 4.6 The minimum distance of a ground CNR as function of the length of the wanted link of the A/A CNR communications in order to maintain a link quality better than 3 and 20 dB C/(N+I) ratio for flying altitudes of 100, 200 and 300 m.

(3) Ground CNR interfering A/G CNR communications

The required minimum Egli distance d_{gg} resulting in C/(N+I) ratio of 3 and 20 dB for a ground CNR interfering with A/G CNR communications is as follows:

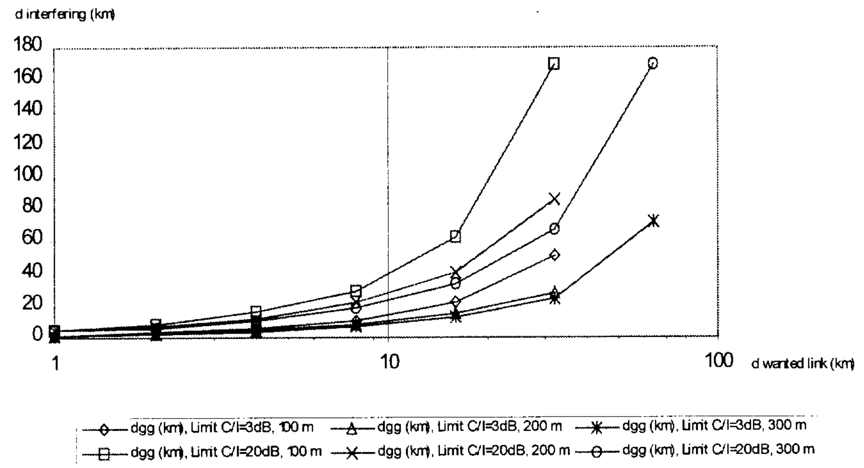


Figure 4.7 The minimum Egli distance of a ground CNR as function of the length of the wanted link of the A/G CNR communications in order to maintain a link quality better than 3 and 20 dB C/(N+I) ratio for flying altitudes of 100, 200 and 300 m.

(4) Ground CNR interfering G/A CNR communications

The required minimum Egli distance d_{gg} resulting in C/(N+I) ratio of 3 dB and 20 dB for a ground CNR interfering with G/A CNR communications is as follows:

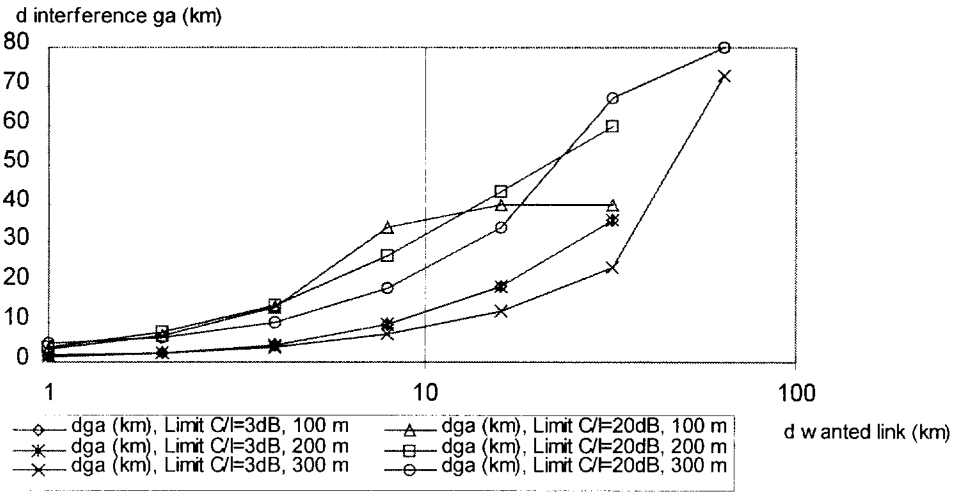


Figure 4.8 The minimum Egli distance of a ground CNR as function of the length of the wanted link of the G/A CNR communications in order to maintain a link quality better than 3 and 20 dB C/(N+I) ratio for flying altitudes of 100, 200 and 300 m.

Sub-Conclusions / Observations

The areas where frequency co-ordination of the ground CNRs and airborne CNRs is required, could still be large. The size is depending on the length of the wanted link. The discussed area is relatively limited in the case of ground CNRs interfering with Air to Ground (A/G) reception.

5. Other Parameters To Be Considered

The degree of interference actually depends on the following three frequency management parameters:

a. Geographical Separation
b. Frequency Separation
c. Time Sharing

The following analysis in paragraphs 5.1 and 5.2 is performed based on the requirement that the link availability should be 100%.

(1) Geographical Separation

As it was discussed above, in airborne applications, frequency re-use will only be possible if the other CNR is almost beyond the RHD. In practice, the RHD covers a wide area. Therefore, frequency re-use is not practical in this matter.

(2) Frequency Separation

Based on the CNR parameters considered, a frequency separation of 25 kHz will provide an isolation between two nets as high as 55 dB. This frequency separation is especially useful if the CNRs are operated in the fixed frequency mode. This will ensure that a single interferer on a different platform (not cosited; geographical separation > 500m) separated by 25 kHz will not significantly interfere with the wanted link.

(3) Time Sharing

The utilization of the nets is normally low. In accordance with the questionnaire responses, the utilization rate is between 10 - 20 %. The probability that two nets using the same frequency in the same area will interfere each other is in the order of 1% to 4% (10%*10% to 20%*20%). In case of analogue CNR, this probability will be even much lower due the possibility for the users to listen first whether the channel is free or not before using it. However, manual time sharing by the users is not effective. It should be done automatically by the system. The **Free Channel Search (FCS)** mode of CNRs is taking advantage of the fact that the utilization rate of CNRs is low. Therefore, it is a spectrum efficient mode allowing sharing frequencies without interference.

6. Frequency Management Aspects

Based on the discussions in the previous paragraphs, it is concluded that it is necessary to manage the usage of CNR frequencies on airborne platforms with dedicated rules.

In case of operation in the Fixed Frequency mode:

The VHF CNR nets having airborne components should be separated by 25 kHz from the other nets operated in the same area.

In case of operation in the Free Channel Search (FCS) mode:

The FCS bundle allotment/assignment algorithm used for ground VHF CNR can be applied.

In case of operation in the Frequency Hopping mode:

Ideally, it would be necessary to apply hopsets with exclusive frequencies with respect to the frequencies used for the Fixed Frequency and the Free Channel Search operated Land based CNRs. Due to very limited CNR spectrum resources, the number of exclusive frequencies for the hopset in question should also be very limited. There is therefore a choice between using a common hopset with as many frequencies as possible or an exclusive hopset with a small number of frequencies. The influence of this second strategy to the Processing Gain of the Frequency Hopping Electronic Protective Measure (EPM, ex ECCM) is analyzed as follows:

Theoretically, the processing gain of a Frequency Hopping system is defined as:

a. the ratio between the number of frequencies used for frequency hopping and the single frequency used by fixed frequency radios

or

b. the ratio between the total bandwidth used for hopping and the bandwidth of the base band signal.

However, in practice, the following factors apply:

a. To jam e.g. 100 randomly distributed channels, it would be not practical to build a jammer which will be able to jam simultaneously 100 distinct single channels.

b. To employ several single channel jammers to jam the 100 channels would be very costly because of the number of equipment, logistical support and man power required.

Therefore, in the example where 100 frequencies are available for hopping, providing that they are well spread over a spectrum width required for 1000 channels, the Processing Gain will not be only 20 dB but it will be close to 30 dB. In other words, the OPERATIONAL Processing Gain produced by e.g. 100 frequencies used for frequency hopping, is very close to the Processing Gain produced by 1000 consecutive frequencies, provided that the 100 frequencies are well spread over the whole CNR band.

7. Sharing Of Frequency Resources Between CNRs Operating In Different Modes

Efficient usage of frequency resources could also be imposed by sharing frequency resources used by CNRs working in different modes in a particular area e.g. in a corps area. The situations are analyzed in the following sub-paragraphs:

(1) Sharing between Fixed Frequency and Free Channel Search modes

NOTE: In all situations outlined below, it is assumed that both systems are active and the geographical separation is such that the C/(N+I) ratio at the victim Rx is much less than 20 dB, such that co-channel interference would occur if the same fixed frequency was used.

If a Fixed Frequency CNR net uses a frequency of a frequency bundle of Free Channel Search CNR nets the following situation will occur:

If the shared frequency is being used by a Fixed Frequency CNR, the Free Channel Search algorithm will choose another free frequency in the bundle. However, if the shared frequency is being used by a Free Channel Search CNR, two situations can occur with respect to a Fixed Frequency user:

a. The Fixed Frequency CNR will hear that his shared frequency is in use and will only start to operate his push to talk switch after the frequency in question become free or

b. The Fixed Frequency CNR user is ignorant of the status of his frequency and just will directly use his Fixed Frequency CNR.

In the case of situation a, the user should probably wait for about 15 - 30 seconds before he can use his Fixed Frequency CNR. In case of b, of course he will interfere with the on going Free Channel Search CNR communications and also he will notice interference on his communications. It is up to the users to determine whether the draw back of sharing frequencies between Fixed Frequency and Free Channel Search CNR users is acceptable. Interference could be minimized by procedural means developed in exercises.

(2) Sharing between Fixed Frequency and Frequency Hopping modes

In the case where a Fixed Frequency CNR uses a frequency in the hopset of a Frequency Hopping CNR, burst errors with a duration of the dwell time of the Frequency Hopping system and an interval of size of hopset/hoprate will occur in both systems.

For example consider:

Hopset = 1000 frequencies
Hoprate = 100 hops/sec

If both systems of CNR are on the air:

a. A single Fixed Frequency CNR vs a single Frequency Hopping CNR

At the Fixed Frequency as well as the Frequency Hopping CNR, a burst error of about 10 ms will occur every 10 seconds. The average BER will be about $\mathbf{0.5*10^{-3}}$.

b. **n** Fixed Frequency CNRs vs a single Frequency Hopping CNR
(**n** = number of Fixed Frequency CNRs using n different frequencies in the hopset of the Frequency Hopping CNR)

At each Fixed Frequency CNR a burst error of 10 ms will occur every 10 seconds. The average BER at each Fixed Frequency CNR will be about $0.5*10^{-3}$. However, the Frequency Hopping CNR will suffer n times burst errors of about 10 ms each every 10 seconds. The average BER will be then $\mathbf{0.5*n*10^{-3}}$.

c. A single Fixed Frequency CNR vs **n** Frequency Hopping CNRs
(**n** = number of Frequency Hopping CNRs of different nets using the same hopset)

Each Frequency Hopping CNR will suffer a burst error of about 10 ms every 10 seconds. The average BER at each Frequency Hopping CNR will be then $0.5*10^{-3}$. However, at the Fixed Frequency CNR, n error bursts of about 10 ms each will occur every 10 seconds. The average BER will be about $\mathbf{0.5*n*10^{-3}}$.

d. **n** Fixed Frequency CNR vs **m** Frequency Hopping CNRs
(**n** = number of Fixed Frequency CNRs using n different frequencies in the hopset used by **m** Frequency Hopping CNR in different nets using the same hopset)

At each Fixed Frequency CNR **m** burst errors each of 10 ms will occur every 10 seconds. The average BER at each Fixed Frequency CNR will be about $\mathbf{0.5*m*10^{-3}}$. The Frequency Hopping CNR will suffer **n** burst errors of about 10 ms each every 10 seconds. The average BER will be then about $\mathbf{0.5*n*10^{-3}}$.

(3) Sharing between Free Channel Search and Frequency Hopping modes

The effect of sharing of frequencies used by a Free Channel Search bundle and a Frequency Hopping hopset, can be derived from and is in principle similar to the effects described in sub-para 7.2. An additional aspect is, if the interval is too short (small hopset & high hoprate), the Free Channel Search CNR may have problems in finding a free channel.

(4) Sharing between Fixed Frequency, Free Channel Search and Frequency Hopping modes

The effect of sharing the frequencies of a Free Channel Search bundle, a hopset of Frequency Hopping and also used by Fixed Frequency CNR can also be derived from and in principle similar to the effects described in sub-para 7.2. Like in para 7.3, if the interval is too short (small hopset & high hoprate), the Free Channel Search CNR may have problems in finding a free channel. The frequencies used by Fixed Frequency CNR will only reduce the number of free channels in frequency bundle of the Free Channel Search CNR.

(5) Sharing between Frequency Hopping CNRs operated on two different hopsets sharing some frequencies

The Frequency Hopping CNRs discussed in this paragraph, due to the nature of their operational usage are considered to be on different platforms. Therefore, the interference mechanism is the co-channel interference mechanism. Further, it is assumed that the Frequency Hopping CNRs are made based on a single synchronised CNR system, so that partial overlap of dwells do not need to be considered.

Frequency Hopping CNR 1:
Hopset1 = **n1** frequencies
Frequency Hopping CNR 2:
Hopset2 = **n2** frequencies
Number of common frequencies = **m**
Hoprate = **r** hops/sec

Interference on Frequency Hopping CNR 1 as well as CNR 2:

Bit Error Rate (BER) = **m * 1/n1 * 1/n2 * 50%**
Every error is a burst error with a duration of **1/r** seconds.

NOTE: The design of a modern radio has already taken into account the statistical distribution of the hits and bit errors by providing adequate interleaving and data protection schemes to protect against error rates of about 10^{-3} (modern CNR up to 10^{-1}) and burst lengths of 10 msec. However, to comply with the request of TWG of FMSC, the following is also provided on the possibility of longer error burst.

In case of frequency hopping, generally, the instantaneous frequency is selected (pseudo) randomly from the pool of frequencies in the hopset.

The average probability that a frequency leading to a collision is selected is: m*1/n1*1/n2 = **m/(n1*n2)**
For example: If n1 = n2 = m = 1000, the probability of collision is 10^{-3}.

The average probability of 2 consecutive collisions is: $\mathbf{(m/(n1*n2))^2}$

In general:
The average probability of p consecutive collisions is: $\mathbf{(m/(n1*n2))^p}$

The character of the distribution of the burst lengths is Poisson distributed.

Sub-conclusion:

a. The nature of the errors is pseudo-random burst errors.
b. The length of the burst errors in number of bits is corresponding to the number of bits in a dwell period. The length of the burst decreases if the hoprate increases. (The length of burst in number of bits defines the size of the required buffer and the time delay introduced by the buffer.)
c. The average BER increases with the number of shared frequencies and decreases with the increase of the size of each hopset.

(6) Two Frequency Hopping CNRs with different hoprates operated on two different hopsets sharing some frequencies

The Frequency Hopping CNR systems discussed in this paragraph, due to the nature of their operational usage are considered to be on different platforms. Therefore, the interference mechanism is the co-channel interference mechanism.

Frequency Hopping CNR 1:
Hopset1 = **n1** frequencies
Hoprate1 = **r1** hops/s
Frequency Hopping CNR 2:
Hopset2 = **n2** frequencies
Hoprate2 = **r2** hops/s
Number of common frequencies: **m** frequencies
The average time where all of the frequencies are used once is:
for CNR 1: **n1/r1** sec
for CNR 2: **n2/r2** sec

Interference on Frequency Hopping CNR 1 as well as CNR 2:
Errors = **m * 1/n1 * 1/n2 * 50%**
Every error is a burst error with a duration of the smallest between [1/r1 and 1/r2] seconds.

Sub-conclusion:

a. The nature of the errors is pseudo-random errors.
b. Each error is a burst error given by the shorter dwell time. The length of the burst decreases if the highest hoprate increases.
c. The BER increases with the number of shared frequencies and decreases with the increase of the size of each hopset.

8. Summary

Summary of the results are as follows:

(1) Operational aspects

The optimal selection of frequencies for CNRs depends on the scenario of the deployment of the nets and on the radio modes used, since each mode has different interference characteristics. The military doctrine for the change of mode is also an important input parameter.

(2) Coverage areas

a. The LOS coverage area of airborne CNRs at the lower altitudes typical of Close Air Support is similar to the equivalent area under the influence of ground CNRs in the VHF frequency range.

b. The LOS coverage area of airborne CNRs at higher altitudes is larger than the equivalent area under the influence of ground CNRs. The area interfered by an airborne CNR at 1 km altitude is a circle with a radius of 130 km which is comparable to the surface of a battle field. However, due to the larger number of interfering sources (ground CNRs) seen by an airborne CNR, the airborne CNR is principally a victim of interference rather than a source, when hopping on frequencies used by fixed frequency ground CNRs or sharing hopset frequencies with ground CNRs.

(3) Frequency resources

a. If the link availability of the CNR nets should be 100%, the fixed frequencies used by airborne CNRs should be exclusive to the fixed frequencies used by the ground CNRs within the Radio Horizon Distance (RHD) of the airborne CNR used in the same operational area.

b. Ideally, in the case of simultaneous use of both hopping and non hopping systems, the frequencies should be mutually exclusive in order to preclude any degradation of the link quality. Since frequency resources are quite often limited in comparison to user requirements, it is anticipated that this approach will limit the number of frequencies available for hopsets.

c. In the case of simultaneous use of hopping and non-hopping where overlapping of frequencies can not reasonably be avoided because of insufficient frequency resources, the acceptable level of link degradation, duty cycle and priority of the CNR nets should be included in the frequency management considerations.

d. If the airborne and ground CNRs are both hopping, the usage of a single hopset containing as many frequencies as possible will be beneficial.

e. In case of the simultaneous usage of non-exclusive frequencies for the operation of fixed frequency and Free Channel Search with frequency hopping CNR, some interference is expected. Large hopsets will minimise the degree of interference.

f. A principal consideration for use of airborne CNRs at higher altitudes (i.e., above about 300 meters) will be the tolerable level of link degradation for the airborne user. Some modern radios have been designed such that the likely interference environment may be acceptable. An additional effort related to the use of airborne CNRs at higher altitudes is the potential need to co-ordinate its use with adjacent operations (or cross-border, if applicable).

g. Frequency sharing advantage can be made from the fact that the utilisation rate of VHF CNR is only up to 10% - 20% e.g. by employing Free Channel Search mode.

(4) Frequency management aspects

a. In case of operation in the Fixed Frequency mode:

The VHF CNR nets having airborne components should be separated by 25 kHz from the other nets operated in the same area.

b. In case of operation in the Free Channel Search (FCS) mode:

The FCS bundle allotment/assignment algorithm used for ground VHF CNR can be applied.

(5) Effectiveness of Electromagnetic Protection Measure (EPM)

If a hopset with a limited number of frequencies should be used, the frequencies should be well spread over the VHF Combat Net Radio band in order to maximise the EPM effectiveness of the frequency hopping .

9. References

a. ECAC-CR-85-019 January 1985
Frequency Management Considerations for the US ARMY Airborne SINGGARS

b. Mischa Schwartz
Information transmission, modulation, and noise
Third edition, 1980, McGraw-Hill, Kogakusha.

c. J. Egli, Proceedings of IRE, October 1957 pp 1383-1391

d. DOT/FAA/ES-83/3, Sep 1983
G.D. Gierhart and M.E. Johnson
IF-77 Electromagnetic Wave Propagation Model

e. For NATO Community
ARFA(E)-C(97)-0449L-SPA/152, dated 06 May 1997
K.S. Kho
EMC Analysis on VHF (20-108 MHz) in Airborne Application

Noise Pre-Processing for Tactical Secure Voice Communications

John S. Collura

National Security Agency

9800 Savage Rd, STE 6516, Ft. Meade, MD 20755-6516, USA

Phone: (301) 688-0269 Fax: (301) 688-0289 E-mail: jscollu@alpha.ncsc.mil

ABSTRACT

Recent advances in speech enhancement and noise pre-processing algorithms have dramatically improved the quality and intelligibility of speech signals, both in the presence of acoustic background noise and in more benign environments. The use of speech enhancement algorithms combined with voice coding algorithms and applied to secure wireless communications systems is an area of increasing importance to the tactical user community. This paper will present the experience of the US. Government with respect to the selection and subsequent improvement of the 2.4Kbps Mixed Excitation Linear Prediction (MELP) speech coding algorithm. A new 1.2Kbps version of the MELP algorithm intended for the tactical user will also be discussed.

1. INTRODUCTION

NATO tactical forces rely on secure voice equipment provided by the 19 member nations to satisfy joint interoperable secure communications requirements. NATO is currently undergoing a competition among the nations to select the next generation secure voice coding STANAG for the foreseeable future.

NATO legacy tactical secure voice communications systems use either the 2.4Kbps Linear Predictive Coding (LPC10e) algorithm known as STANAG 4198, or the 16Kbps Continuously Variable Slope Delta Modulation (CVSD) algorithm for speech compression. These algorithms were considered the state of the art in narrow band speech coding when introduced 20 to 30 years ago. In fact, it's a testimony to their effectiveness that until 1996 there were no alternative narrow band speech coding algorithms available to the tactical user.

In March of 1996, the U.S. government DoD Digital Voice Processing Consortium (DDVPC) announced the selection of the 2.4Kbps Mixed Excitation Linear Prediction (MELP) voice coding algorithm as the next standard for narrow band secure voice coding products and systems [1]. The selection of the MELP voice-coding algorithm represented a dramatic improvement in both speech quality and intelligibility at the 2.4Kbps data rate. The concept for this new voice coding algorithm was to provide a single high performance speech coding algorithm for seamless interoperability across and between the strategic, tactical and SATCOMs communications domains.

One of the driving forces behind the selection of the 2.4Kbps MELP algorithm was operation in harsh acoustic noise environments such as HMMWV's, helicopters and tanks. In these harsh acoustic environments, the MELP algorithm, while providing superior performance over legacy systems, exhibited somewhat degraded performance, thus demonstrating a need for further improvement. Such improvement was addressed by the development and ongoing integration of a speech enhancement algorithm into the front end of the MELP speech coding algorithm.

There are a number of important transmission channels which don't support ***robust*** speech coding transmission at or above 2.4Kbps. These channels include survivable strategic SATCOMs, high latitude HF links and covert operations, among others. To accommodate these requirements, a 1.2Kbps MELP algorithm was developed which shares the 2.4Kbps core algorithmic paradigm, but with an alternate quantization scheme for the parameters. The resulting combined 1.2Kbps/2.4Kbps MELP algorithm also incorporates an integrated speech enhancement algorithm as a front-end process to provide superior performance in these harsh acoustic noise environments. This paper discusses the resulting combined 1.2Kbps/2.4Kbps MELP algorithm suite including the integrated speech enhancement front end. The 1.2Kbps/2.4Kbps MELP algorithm coupled with the speech enhancement algorithm will be referred to in this paper as the MELPn algorithmic suite.

The limited performance of the current generation voice coding systems (including the MELP algorithm as selected in 1996) in harsh

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

acoustic noise environments has given rise to the idea of enhancing the voice signal prior to compression. This paper discusses the combination of one such noise pre-processing algorithm developed at AT&T Labs Research in conjunction with the US government. Performance data presented in this paper will include quality and intelligibility testing, with and without speech enhancement, and under two main scenarios. The first scenario tested the system performance in the quiet environment, while the second tested the performance in the HMMWV and CH-47 harsh acoustic noise environments. The HMMWV is a heavy-duty four-wheeled drive vehicle used for troop transport. Due to the low gear ratios, and four-wheeled drive operations, the acoustic character of the HMMWV background noise is dominated by non-stationary low frequency rumbling with an average (over six speakers) speech SNR of approximately 12.6 dB. The CH-47 is a turbine driven tandem rotor heavy lift helicopter. This acoustic environment is characterized by both the beat frequency of the rotors (quasi-stationary noise components) and the Gaussian type noise generated by the turbine operation. The CH-47 has an average (over six speakers) speech SNR of approximately 13.75 dB.

2. SYSTEM CONFIGURATION AND DESCRIPTION

2.1 Speech Enhancement Algorithm

Previous and current generation noise pre-processing algorithms have generally been effective in noisy environments at the expense of introducing objectionable structured musical artifacts in the enhanced signal. An additional drawback was the need to manually switch the noise pre-processing algorithm depending upon the nature of the acoustic environment. This switching was required due to the degradation inherent with the application of noise pre-processing algorithms to benign acoustic environments. These older generation noise pre-processing algorithms tended to improve the quality, but degrade the intelligibility of the noisy speech signal.

The AT&T speech enhancement/noise pre-processing algorithm is the culmination of several years of research, [2][3][4]. This speech enhancement algorithm is based upon the following heuristic operational description. First, the speech signal is divided into time slices of 32 ms (and later 22.5ms) in length, and an FFT is applied to provide access to the spectral information contained in the signal. Next, an estimation algorithm is used to model the noise during frames in which speech is absent. This part of the algorithm uses a voice activity detector to enable the algorithm to distinguish when the signal is composed of speech + noise, and when it is noise only. Separate noise models are maintained for each frequency bin based upon those portions of the signal, which is noise only. The algorithm minimizes the mean squared error, (MMSE) of the log spectral amplitudes. It tracks the probability of speech presence and applies an additional gain factor based upon these probabilities. These probabilities are also used to update the noise power spectral density during speech. Once all of the frequency bins have been adjusted, the resulting modified magnitude spectrum is recombined with the original phase spectrum using an Inverse FFT to recover the enhanced speech. A synchronized overlap and add, (SOLA) technique, which helps to eliminate many of the artifacts is then applied to the reconstructed signal segments to recover the composite enhanced signal. This enhanced speech signal is then used as the input to the speech coding algorithm, in this case the MELP algorithm. Work is proceeding on the integration of the speech enhancement algorithm into the front end of the MELP algorithm resulting in a new algorithmic designation of MELPn for the combined algorithm. The AT&T speech enhancement algorithms used for the bulk of this paper are versions 5 & 7, and will be designated as NPP5 and NPP7 in this paper. The main difference between these versions lies in the frame update rate. NPP5 operates on a 32ms frame while NPP7 was made to synchronize with the MELP voice coder basic frame rate of 22.5ms. To accommodate the smaller frame size, the parameters of NPP7 were retuned to maintain system performance. This tuning process actually resulted in a higher level of overall performance.

Initial quality and intelligibility testing of the AT&T speech enhancement algorithm used in conjunction with the MELP voice coders indicates that a fundamental breakthrough has been achieved in the operation of noise pre-processing algorithms. Testing shows that speech quality and intelligibility improve in both harsh acoustic noise environments as well as the benign acoustic environments. The need for an algorithmic switch for the noise pre-processor will be eliminated once a more extensive test verifies these initial test results.

2.2 2.4Kbps MELP Speech Coding Algorithm

The 2.4Kbps MELP speech coding algorithm was selected by the US Government in March of 1996 to be the primary voice coding algorithm for seamlessly interoperable narrow band secure voice communications for strategic, tactical and SATCOMs applications. The selection of 2.4Kbps MELP as the standard voice coding algorithm provides the first enabling technology for seamless interoperability across the domains, regardless of the points of presence of the users in the communications networks. In fact, the end users probably will not be aware of the intermediate routing (tactical radio signals patched through SATCOM links which then feeds the signal through Internet Protocols, etc...) used to establish the end-to-end communications link.

The 2.4Kbps MELP algorithm divides the 8KHz 16 bit linear sampled speech signal into 22.5ms frames for analysis. Table 1 provides a breakdown of the parameters used by the MELP algorithm with the number of bits per frame needed to quantize each parameter. A complete description of the MELP algorithmic paradigm can be found in the MELP FIPS draft standards publication at http://www.plh.af.mil/ddvpc/frontpage.html.

Table 1: 2.4Kbps MELP Parameter breakdown

PARAMETER:	NUMBER OF BITS/FRAME
LSF's	25 Bit Multistage VQ
Fourier Magnitudes	8 Bits VQ
Pitch	7 Bits
Bandpass Voicing	4 Bits
Gain	2 x 4 Bits
Aperiodic Pulse	1 Bit
Sync	1 Bit

3. TESTING

3.1 Test Description

The object of this work was to measure any performance gains or degradations for communications equipment when the 2.4Kbps MELP voice coding algorithm is preceded by a speech enhancement algorithm. To do this objectively, the test included the most benign acoustic environment possible, that of the quiet background, and two of the harsher acoustic noise environments, the HMMWV, and the CH47 helicopter.

Two subjective measures were used to evaluate coder performance. The Diagnostic Rhyme Test (DRT) [5] was used to measure speech intelligibility, and the Diagnostic Acceptability Measure (DAM) [6] was used to measure speech quality. Both of the tests have been used extensively in previous US Government voice coder selections. Each evaluation was made with six speakers: three males and three females.

The DRT is a two choice intelligibility test based upon the principle that the intelligibility relevant information in speech is carried by a small number of distinctive features. The DRT was designed to measure how well information as to the states of six binary distinctive features (voicing, nasality, sustension, sibilation, graveness, and compactness) have been preserved by the communications system under test. The DRT uses a suite of 96 rhyming word pairs (192 items per speaker), in which the initial consonants of the two words of each pair differ only with respect to one of the distinctive features. The listener must select which of the two rhyming words was spoken. With a carefully selected and monitored panel of eight listeners, the DRT has extremely high resolving power and test-retest reliability

The DAM is a proprietary test developed and administered by Dynastat, Inc. in Austin, Texas. The DAM requires the listeners to judge the detectability of a diversity of elementary and complex perceptual qualities of the signal itself, and of the background environment. The qualities, which are evaluated, have been experimentally shown to determine a listener's judgements of speech acceptability. The DAM thus provides multiple direct and indirect estimates of a communication system. The DAM is designed for use with small (12-16) crews of listeners, who are rigorously screened, trained, calibrated, and monitored to ensure their collective response very closely approximates that of the typical or normative listener.

3.2 Test Results

Several tests were run to exercise the performance of the MELP algorithms both with and without the speech enhancement front end. Additionally, the performance of the noise pre-processor version 5 (NPP5) without speech coding was contrasted against that of the unprocessed 128Kbps PCM source material for both quiet and HMMWV acoustic noise environments. Tables two through four present performance data for three acoustic noise environments quiet, HMMWV, and the CH47

"Chinook" heavy lift helicopter. It should be noted that the data for Tables two through four reflects the latest available version of the speech enhancement software for the condition tested. In some cases, comparisons will be made with differing versions of the enhancement algorithm. Reporting of the 2.4Kbps MELP coder results reflects speech enhancement algorithm version 7, (fully optimized with 22.5ms delay) which was implemented in a separate pre-processing stage for this paper. In an actual implementation, the speech enhancement process will be integrated into the speech coding algorithm. Reporting for the 1.2Kbps MELP coder results reflects an incomplete integration of the speech enhancement front end. In this instance, the original 32ms frame rate was changed to 22.5ms and integrated into the speech coder, but the parameters of the enhancement algorithm were not re-optimized. For this reason, the tests on the 1.2Kbps version of the MELP algorithm quoted below are considered preliminary in nature. After the integration of the re-optimized version of the speech enhancement algorithm has been accomplished, additional testing will be performed.

TABLE 2:QUIET ENVIRONMENT		
TEST CONDITION	**DRT/SE**	**DAM/SE**
Source Material	**97.8/0.23**	**84.1/2.1**
Noise Pre-Processing Only	**96.7/0.52**	**85.1/1.3**
2.4 Kbps MELPn (NPP7)	**93.8/0.65**	**66.3/0.9**
2.4 Kbps MELP Coding Only	**93.0/0.89**	**64.9/1.0**
1.2 Kbps MELPn (NPP5)	**92.6/0.65**	**63.4/0.9**

TABLE 3: HMMWV ENVIRONMENT		
TEST CONDITION	**DRT/SE**	**DAM/SE**
Source Material	**91.0/0.37**	**45.0/1.2**
Noise Pre-Processing Only	**80.6/0.66**	**54.6/1.2**
2.4 Kbps MELPn (NPP7)	**74.4/0.83**	**52.6/0.9**
2.4 K bps MELP Coding Only	**67.3/0.8**	**38.9/1.1**
1.2 K bps MELPn (NPP5)	**67.8/0.77**	**52.1/0.6**

TABLE 4: CH47 HELICOPTER ENVIRONMENT	
TEST CONDITION	**DRT/SE**
2.4 Kbps MELPn (NPP7)	**76.9/0.77**
2.4 Kbps MELP Coding Only	**66.7/0.61**
1.2 Kbps MELP Coding Only	**69.1/0.81**

TABLE 5: HISTORY OF 2.4KBPS MELP W/ NPP (HMMWV ACOUSTIC NOISE CONDITIONS)		
TEST CONDITION (test date)	**DRT/SE**	**DAM/SE**
Original 2.4 Kbps MELP (3/96)	**63.1/0.72**	**N/A**
Updated MELP (1/98)	**67.3/0.8**	**38.9/1.1**
MELP + NPP version 3 (6/98)	**72.3/0.92**	**48.4/0.6**
MELP + NPP version 5 (10/98)	**72.0/0.64**	**52.0/0.9**
MELP + NPP version 7 (1/99)	**74.4/0.83**	**52.6/0.8**

3.3 Interpretation of Test Results

Tables two and three cover the same test conditions, but in different acoustic noise environments, quiet and HMMWV respectively. The first two test conditions provide a reference for the other conditions by contrasting the performance of the original source material with that of NPP5 without the additional influence of speech coding. Then, the data for the 2.4Kbps MELP algorithm with and without NPP7 is presented.

Table two presents data for testing performed in the benign quiet acoustic noise environment. This data indicates that noise pre-processing alone introduces minimal degradation with respect to both speech intelligibility and quality. When the performance of the 2.4Kbps MELP algorithm is compared with that of the 2.4Kbps MELPn algorithm, the difference is borderline positive for the intelligibility measurement with a clear benefit for the quality measurement. Of particular interest, is the limited degradation with respect to quality and intelligibility for the 1.2Kbps MELPn algorithm as compared to both the original 2.4Kbps MELP and 2.4Kbps MELPn versions.

Table three presents data for testing performed in the harsh HMMWV acoustic noise environment. Most notable in this data is the simultaneous decrease in the intelligibility and increase in quality scores for the NPP5 algorithm without speech coding as contrasted against the unprocessed source material. Both the 1.2Kbps and 2.4Kbps MELPn algorithms demonstrate higher overall voice quality than the unprocessed source material. Though the intelligibility scores are considerably degraded for 2.4kbps MELPn when compared with the unprocessed source material, they are likewise considerably higher than that of the original 2.4Kbps MELP algorithm. The performance of the 1.2Kbps MELPn algorithm out-performs the original 2.4Kbps MELP algorithm for both quality and intelligibility, although the intelligibility scores are within the standard error of the test.

Table four provides intelligibility scores for the CH47 helicopter environment. This information provides a simple verification that the performance gains exhibited in both the HMMWV and quiet environments were applicable to other acoustic environments. This corroboration of the test results demonstrates a consistency of operation regardless of the type of acoustic noise. Most notable in this data is the superior performance of the 1.2Kbps MELPn over original 2.4Kbps MELP.

Table five presents the improvements which have been achieved with the MELP algorithm in the HMMWV acoustic noise environment from the initial selection in the spring of 1996 to the current MELPn algorithm using NPP7. Two items are most significant to note, first that a steady improvement was achieved in the operation of the speech enhancement algorithm over the course of development. Second that these test scores demonstrate an intelligibility improvement of 7 to 10 points and a quality improvement of up to 14 points. Given standard errors of between 0.6 to 1.2 points, an improvement from one condition to another of 10 points is extraordinary! These results are further corroborated by the CH47 helicopter scores.

4. DISCUSSION

4.1 General introduction

NATO currently depends upon two voice coding algorithms for interoperable tactical secure voice communications, 2.4Kbps LPC10 and 16Kbps CVSD. This section discusses the test data for the new 1.2Kbps/2.4Kbps MELPn algorithm in the context of the current NATO voice coders. The MELPn speech coding algorithm(s) represents only one example of several algorithms which are currently competing to become the next NATO primary voice coding algorithm for interoperable strategic, tactical and SATCOMs communications.

Speech enhancement algorithms have greatly improved both the quality and intelligibility of state of the art voice coding algorithms. In the case of the MELPn algorithm, the introduction of speech enhancement techniques increased both the quality and intelligibility of the reconstructed speech by at least 7 to 10 points in the harsh HMMWV and Helicopter noise environments. The MELPn test results for the HMMWV and Helicopter environments demonstrate that the speech enhancement algorithm is effective on both stationary (helicopter noise) and non-stationary (HMMWV engine & gear noise) types of acoustic background noise.

Table six provides test results for the two NATO voice coders in the three representative acoustic environments of quiet, HMMWV & CH-47. This data allows for a direct comparison between the 1.2Kbps/2.4Kbps MELPn technology and the 2 NATO voice coders.

4.2 CVSD

The 16Kbps CVSD voice coding algorithm is predominately used in applications such as VHF frequency hopping combat net radios. Two examples of this type of radio are the US Single Channel Ground Air Radio (SINCGARS) and the French PR4G. The most significant attribute associated with the CVSD algorithm is that it provides relatively high intelligibility in the presence of harsh background acoustic noise. A prime example of this is the performance of CVSD in the CH-47 helicopter. Even with speech enhancement, the 2.4Kbps MELPn algorithm only matches the intelligibility of CVSD, a true testimony to its effectiveness. A second data point for harsh noise environments is that of the HMMWV. In this case however, the performance of 2.4Kbps MELPn far exceeds that of CVSD for both quality (~20 points) and intelligibility (5 points).

Performance of the CVSD algorithm in benign acoustic environments is at least seven points lower for intelligibility and 11 points lower for quality when compared to the 2.4Kbps MELPn algorithm. In fact, the 2.4Kbps MELPn algorithm exceeded the performance of the CVSD algorithm in two of the three test conditions. It is believed that the helicopter acoustic background environment will remain the only environment for which the CVSD algorithm performs on par with the new generation of 1.2Kbps/2.4Kbps speech coding algorithms. This conjecture is supported by the test data from the 1996 US Government test which originally selected the 2.4Kbps MELP algorithm.

TABLE 6: NATO LEGACY VOICE CODING		
QUIET BACKGROUND	DRT/SE	DAM/SE
2.4 Kbps LPC10e	86.2/0.6	48.9/0.8
16 Kbps CVSD	88.5/0.85	55.1/1.0
HMMWV BACKGROUND		
2.4 Kbps LPC10e	31.7/2.3	28.5/
16 Kbps CVSD	69.3/1.33	31.5/
Helicopter BACKGROUND		
2.4 Kbps LPC10e	47.6/1.2	N/A
16 Kbps CVSD	77.2/0.78	N/A

4.3 Linear Predictive Coding (LPC)

Linear Predictive Coding (LPC) is a 2.4Kbps speech coding algorithm which is probably the single most widely used speech coding algorithm for secure interoperable government wide voice communications systems (strategic, tactical and SATCOMs applications). The reduced data rate requirements for the LPC algorithm as compared to CVSD (6.6667 to 1) allows the LPC algorithm to operate in far more restrictive transmission channels. Applications such as MILSATCOMs, HF radio, strategic secure voice, etc. use the LPC algorithm at the core of their systems. The 2.4Kbps data rate has become pervasive for secure voice systems. LPC currently provides a minimum level of interoperability throughout NATO and US communications systems.

The arguments against the LPC algorithm mirror those of the CVSD algorithm, except that the performance data for the LPC algorithm are uniformly lower than that for CVSD. A good example is in the HMMWV harsh acoustic environment, the intelligibility score for LPC is 42.7 points lower than that of the 2.4Kbps MELPn algorithm. Likewise for the quality measurement, the LPC score is 44.1 points lower than that of the 2.4Kbps MELPn algorithm. These scores are so low in fact, that for this noise condition, the communications link would be virtually unusable. While the scores for LPC in the quiet are acceptable, they are still considerably lower than those for the MELPn algorithm.

5. CONCLUSIONS

Given the limitations of the current generation secure voice algorithms, and the availability of a new generation of 1.2 Kbps and 2.4 Kbps voice coding algorithms using speech enhancement, (the MELPn algorithm is a good example) future generations of NATO secure voice products and systems will provide a greatly improved quality of service to the users. When the NATO voice coding selection is complete, the algorithm which is ultimately selected will form the core of a seamlessly interoperable (with end-to-end security) secure voice communications system which spans the domains of strategic, tactical and SATCOMs operations. The end result is that NATO will have a speech coding algorithm(s) which exhibit higher overall performance, (quality and intelligibility) a lower data rate (1.2Kbps) and greater robustness to the acoustic environment for the military user.

6. ACKNOWLEDGMENTS

The author wishes to thank Dr. Richard. Cox and Dr. Rainer Martin of AT&T Labs Research for their work on the Speech Enhancement algorithm and Dr. Vladimir Cuperman and Dr. Allen Gersho and the staff at SignalCom Inc. for their work on the 1.2Kbps MELP algorithm.

7. REFERENCES

[1] Supplee Lynn M., Cohn Ronald P., Collura John S., McCree Alan V., "*MELP: The New Federal Standard at 2400 bps*", IEEE ICASSP-97 Conference, Munich Germany, pp.1591-1594.

[2] Malah David, Cox Richard V., Accardi Anthony J., "*Tracking Speech Presence Uncertainty to Improve Speech Enhancement in Non-Stationary Noise Environments.*" IEEE ICASSP-99 Conference, Phoenix AZ.

[3] Accardi Anthony J. and Cox Richard V., "*A Modular Approach to Speech Enhancement with an Application to Speech Coding*". IEEE ICASSP-99 Conference, Phoenix AZ.

[4] Martin Rainer, Cox Richard V., "*New Speech Enhancement Techniques for Low Bit Rate Speech Coding.*" IEEE Speech Coding Workshop-99, Porvoo Finland.

[5] Voiers William D., "*Diagnostic Evaluation of Speech Intelligibility.*" In M.E. Hawley, Ed, Speech Intelligibility and Speaker Recognition (Dowden, Huchinson, and Ross; Stroudsburg, PA 1977)

[6] Voiers William D., "*Diagnostic Acceptability Measure (DAM): A Method for Measuring the Acceptability of Speech over Communications Systems.*" Dynastat, Inc.; Austin Texas

ENHANCED ERROR CORRECTION OF THE US FEDERAL STANDARD MELP VOCODER EMPLOYING RESIDUAL REDUNDANCY FOR HARSH TACTICAL APPLICATIONS

Douglas J. Rahikka
National Security Agency, 9800 Savage Rd, STE 6516, Ft. Meade, MD 20755 USA
Ph: (301) 688-0269 Fax: (301) 688-0289 e-mail: djrahik@alpha.ncsc.mil

Thomas E. Fuja
University of Notre Dame, Notre Dame, IN 46556 USA
Ph: (219) 631-7244 Fax: (219) 631-4393 e-mail: tfuja@poisson.ee.nd.edu

Tahereh Fazel
University of Maryland, College Park, MD 20742 USA
e-mail: fazel@eng.umd.edu

ABSTRACT

The U.S. government has developed, adopted, and promulgated a new Federal Standard vocoder algorithm which operates at 2400 bps and is called MELP - Mixed Excitation Linear Prediction [reference 1]. This vocoder has reasonably good voice quality under benign error channel conditions. However, when subjected to high error conditions as may be experienced in vehicular applications, correction techniques may be employed which take advantage of the inherent inter-frame residual redundancy of the MELP parameters. This paper describes experiments conducted on the MELP vocoder algorithm in combination with Viterbi convolutional error decoding, and enhanced with Maximum A Posteriori techniques which capitalize on the redundancy statistics. Both hard and soft Viterbi decoding implementations are investigated.

INTRODUCTION

Over the past two decades, a series of vocoder algorithms have been developed and standardized for application in U.S. government applications with strict interoperability requirements. These include FS1016 CELP at 4800 bps and a new vocoder at 2400 bps called MELP. Work has been undertaken to enhance the performance of CELP when applied over high error rate channels [reference 2], and similar efforts have been recently pursued with MELP. Channels that have been addressed include those impaired by high levels of Gaussian noise and severe Rayleigh fading as experienced in typical urban vehicular applications. The raw channel error rates of interest are in the vicinity of 5-13%. This would be the rate subjected to the 4800 bps composite bit stream of 2400 bps MELP and an associated rate 1/2 Viterbi convolutional error coder (FEC Forward Error Correction overhead of 2400 bps).MELP is a frame-oriented parametric voice coding algorithm. The MELP parameter set bit allocation (for voiced frames) is described in Table 1. These include Pitch, 2 Gains, 4 MSVQ (Multi Stage Vector Quantizer) stages which characterize the LPC coefficient line spectral frequencies (LSFs), FM Fourier Magnitude, BP Bandpass Voicing, AF Aperiodic Flag, and Sync. The results described in this paper are in a system with rate 1/ 2 error correction applied fully to all of the MELP bits. In situations where channel capacity limits one to protecting only half of the MELP bits (i.e., 27 of 54 per 22.5 msec frame), one would only protect the Class1 most significant frame bits as shown in Table1. Fortunately, results for this condition approach that of full protection, clearly indicating the varying importance of parameter bits even for such a low rate vocoding algorithm.

Parameter	Class1 MSBs (27/54)	Total Bits
Pitch	7	7
Gain1	4	5
Gain2		3
MSVQ1	7	7
MSVQ2	4	6
MSVQ3	1	6
MSVQ4	1	6
FM		8
BP	1	4
AF flag	1	1
Sync bit	1	1

Table 1. MELP Parameters

ERROR CORRECTION METHODS

As a first step to optimizing the Viterbi convolutional decoding error correction of MELP, we investigated the inherent redundancy characteristics of the MELP

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

bitstream. As would be expected, due to the non-random nature of the human speech signal, and due to the framing structure of MELP, the MELP bitstream is not memoryless, equiprobable, nor uniformly distributed. We observed the 3 high-order bits in the 7 MELP parameters shown in Table 2 over the course of a long (192,014 frames, or 72 minutes) training sequence, in order to create a first-order (inter-frame) 8-state Markov model for each parameter. The entropy rate H and residual redundancy (3-H) were then computed. Note that the redundancy of the pitch and gain parameters exceeds 50%, and MSVQ1 approaches 20%. Since the amount of residual redundancy in the remaining parameters is minimal, it was decided to concentrate efforts on P, G1, G2 and MSVQ1. Expending additional computation resources on the other parameters would be wasteful.

Parameter	Redundancy, R	Entropy, H
Pitch P	1.62 (54%)	1.38
Gain1 G1	1.80 (60%)	1.20
Gain2 G2	1.70 (57%)	1.30
MSVQ1	0.58 (19%)	2.42
MSVQ2	0.28 (9%)	2.72
MSVQ3	0.08 (3%)	2.92
MSVQ4	0.03 (1%)	2.97
Total	6.09	14.91

Table 2. Residual Redundancy of MELP

In order to evaluate and compare the potential benefit that could be made of the redundancy in MELP, a core convolutional Viterbi error correction system was implemented. All of the results shown are for a 64-state convolutional code (constraint length or encoder memory = 6) and generator matrix of $G = [1 + D^2 + D^3 + D^5 + D^6, 1 + D + D^2 + D^3 + D^6]$, with free Hamming distance of ten. Four error correction schemes, of differing complexities and external requirements, were implemented:

- Hard ML - Traditional Maximum Likelihood Viterbi decoding algorithm that minimizes cumulative decoder trellis path Hamming distance. Hard-decision quantized bit output from demodulator.

- Hard MAP - Maximum A Posteriori (MAP) technique in which hard demod outputs are used in conjunction with the *a priori* Markov model transition probabilities.

- Soft ML - Traditional ML Viterbi algorithm minimizing cumulative decoder trellis path squared Euclidean distance. Soft-decision unquantized matched filter output from demodulator.

- Soft MAP - Maximum A Posteriori (MAP) technique in which soft demod outputs are used in conjunction with the Markov model transition probabilities to select the valid sequence minimizing the "MAP metric" - i.e., $f_{\mathbf{X}|\mathbf{Y}}(\mathbf{y}|\mathbf{x})$, where **X** and **Y** are the channel input and output sequences.

Note that both soft algorithms require that the receiving system's demodulator provide the Viterbi algorithm with bit "quality" information in addition to the hard decoded bits. Depending upon overall system design, soft quality metric information may or may not be available.

SIMULATION RESULTS

In order to determine how well the error correction techniques performed, we employed three evaluation schemes: (1) Informal subjective listening tests, (2) Objective speech quality metrics, and (3) Formal speech testing DAM (quality) and DRT (intelligibility) scores. All of the simulation result data shown in this paper were for a typical cellular radio channel with independent Rayleigh fading (under noise) or with just pure Gaussian Noise GN, BPSK receiver demodulation, and were based upon the processing of approximately 3 minutes (7909 frames) of representative speech. The additional system implementation delay demanded by the MAP algorithms (in excess of that required by the strictly ML algorithms) is one frame (22.5 msec) in order to observe interframe redundancy.

A subset of the objective speech metrics employed is shown in Figures 1 through 6. The Frequency Weighted Spectral Distortion referred to by Figures 1-2 employs the same weighting described by McCree *et al* in [reference 3]. The distortion of noiseless "pure" MELP was found to be approximately 1.1 dB in the 3 minute test sequence. The MELP algorithm encodes gain twice per frame. As an indicator of how well the first of these performed, we computed the mean square error (MSE) between the receiving decoder's estimate and the unquantized value observed at the transmitting encoder. For pure MELP and our sequence, the MSE is about 5.7. The results for the Gain 2 parameter were very similar to that of Gain 1, as would be heuristically concluded. The gain parameter effects were most notable in the corollary informal subjective listening tests, where uncorrected errors in the most significant bits caused very annoying "blasting" to occur. Indeed, it was most fortunate that the results of our residual redundancy measurement foundation experiment found so much lack of interframe entropy in Gain 1 and Gain 2. Finally, we similarly computed the MSE of the difference between the transmitting encoder's unquantized pitch and that value provided by the receiver's error correction process. For pure MELP, that MSE is about 52.1.

Independent Rayleigh Fading Results :

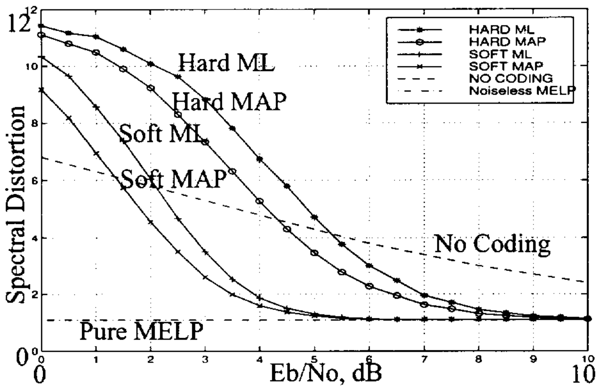


Figure 1. Spectral Distortion Results (Fading)

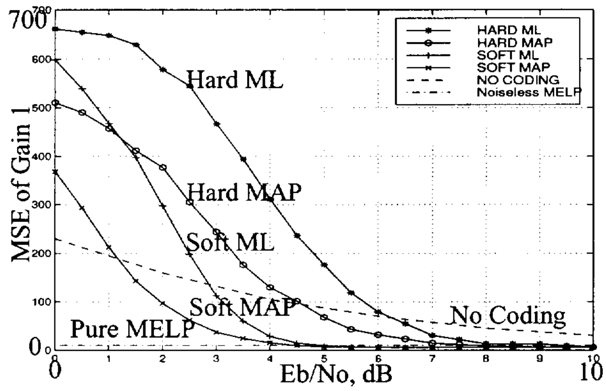


Figure 3. Gain Parameter Results (Fading)

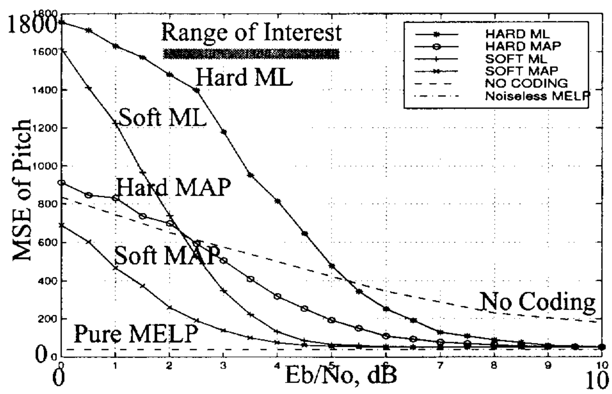


Figure 5. Pitch Parameter Results (Fading)

Gaussian Noise GN Results :

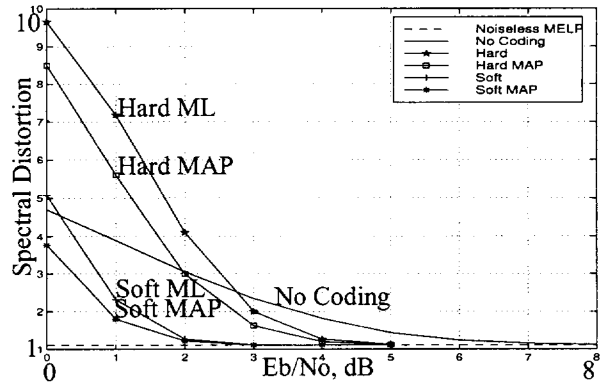


Figure 2. Spectral Distortion Results (GN)

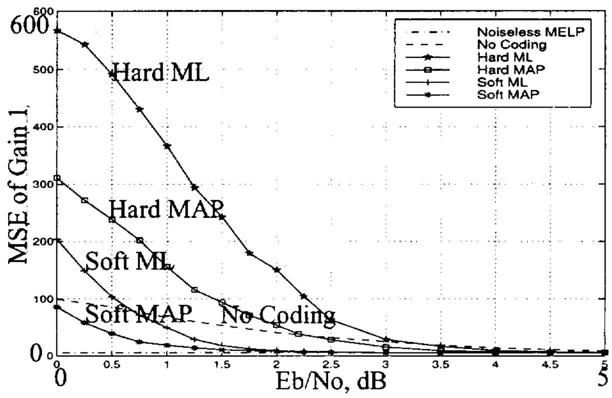


Figure 4. Gain Parameter Results (GN)

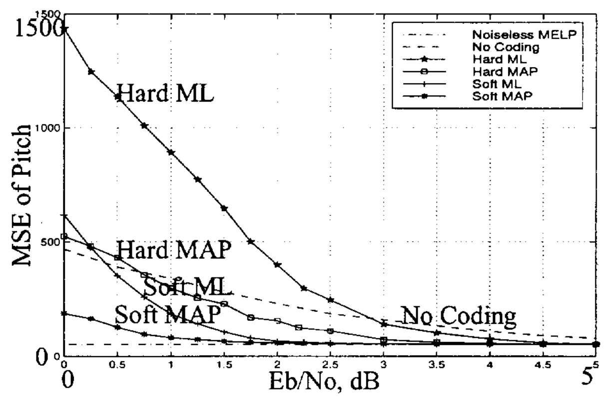


Figure 6. Pitch Parameter Results (GN)

INTERPRETATION & ANALYSIS

Of course, there are no ideal objective measures of speech quality. The performance criteria of spectral distortion, Gain MSE, and Pitch MSE can however be useful indicators and discriminators, especially for system diagnostics.

In all cases, it was somewhat gratifying to see an improvement of 1-2 dB of coding gain provided by the Hard MAP algorithm when compared against the Hard ML, especially for Pitch and Gain. This may be somewhat misleading, however, in that our Eb/No region of interest is 2-5 dB, where raw channel error rates are on the order of 10%. In this region, both methods yield worse performance than no coding at all. This is due to the classic cross-over effect of hard Viterbi convolutional error correction actually doing more harm than good (actually increasing the error rate) when the channel is noisy enough. At only higher Eb/No values do these respective curves plummet below the no coding case, at which level informal listening tests reveal far fewer truly annoying impairments such as frequent blasting. In our 2-5 dB Eb/No region of most interest, it is gratifying to see the improvement demonstrated by the Soft MAP over the Soft ML algorithm, again on the order of a 1 dB coding gain. This improvement is definitely noticeable in informal listening tests and corroborated by DAM and DRT scores.

An alternate method for objectively evaluating the performance of our MAP-enhanced algorithms would be to compute the error rates of the various bit classes, depending on which algorithm they are subjected to, at the output of the channel decoder. For example, Figure 7 shows this sort of decomposition. Here we see the raw transmission channel error rate range between about 20% at 0 dB Eb/No and 10% at 5 dB Eb/No. The "All ML" curve represents the average error rate of all parameters combined when the error correction algorithm is Soft ML applied to all bits with no MAP processing done at all. The "ML bits" curve represents the error rate of those bits (*other* than P, G1, G2, or MSVQ1) not specifically getting the MAP treatment but during application of the defined Soft MAP algorithm. Note the ancillary error rate reducing side-benefit "placebo effect" which results. This is due to the even more reduced error rate of the "MAP bits" themselves reflecting through the joint and intermingled Viterbi decoding algorithm and giving a positive effect on all (basking in the glow of the "good neighbor effect"). For example, at Eb/No of 3 dB, the resulting MAP bit BERs are bounded by Pitch at 3.39% and Gain 2 at 4.06%. This reflects positively back to an "ML bits" BER of 5.37%, which is considerably lower than the corresponding "All ML" bit error rate of 8.98%.

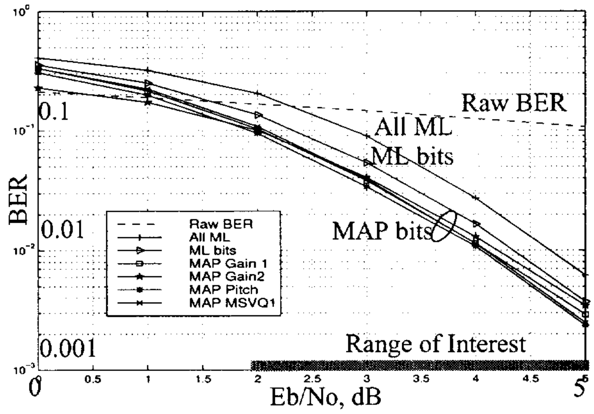


Figure 7. Parametric BER Analysis

Finally, informal subjective listening tests are often the true arbiter of whether a vocoder algorithm is operating well or discernibly different from another. We conducted listening tests of the algorithms at raw channel error rates of 5%, 10%, 13%, and 15%. Table 3 shows the subjective results with the following key - U = Unusable, P = Poor, F = Fair, G = Good. The DAM Quality and DRT Intelligibility scores are to be compared with those of error-free MELP (DAM = 66.1, DRT = 92.5). Note the DAM improvement by more than 6 points (from ML's 45.6 to MAP's 52.0) in the 13% error condition. The Diagnostic Acceptability Measure DAM and Diagnostic Rhyme Test DRT numeric results shown in Table 3 are the outputs of very recent extensive and controlled testing evaluations conducted with trained listening crews and do not appear in a companion brief workshop paper [reference 4].

Channel BER	5%	10%	13%	15%
Uncoded	Poor	U	U	U
Hard ML	Good	Poor	U	U
Hard MAP	Good	Poor	U	U
Soft ML (DAM/ DRT)	Good 66.0/ 93.3	Good 64.0/ 93.0	Fair 45.6/ 87.7	Poor
Soft MAP (DAM/ DRT)	Good 66.2/ 92.3	Good 65.5/ 92.4	Good 52.0/ 90.3	Fair

Table 3. Subjective Listening Tests

As was stated earlier, all results shown in this paper are for a simulation with Viterbi constraint length equal to 6 and with all of the 54 bits per MELP frame protected. Additional comparative results for constraint lengths of 3, 4, and 5 are shown in [reference 4], indicating that constraint length is not a strong influence. Also shown in [reference 5] is some data which indicates the relatively small price to be paid for protecting only the 27 most important of the 54 MELP bits per frame.

TURBO CODE EXPERIMENT

There has been much recent research done by others in the area of Turbo coding as applied to data communications [reference 6]. The Turbo coder that we tested is shown in Figure 8, comprised of a pair of rate-1/2 convolutional 16-state codes with puncturing in order to bring the ensemble code rate up to 1/2 from the unpunctured total rate of 1/3 (so that we can compare with the prior Soft MAP results). The iterative turbo decoding process was used at the receiver five times.

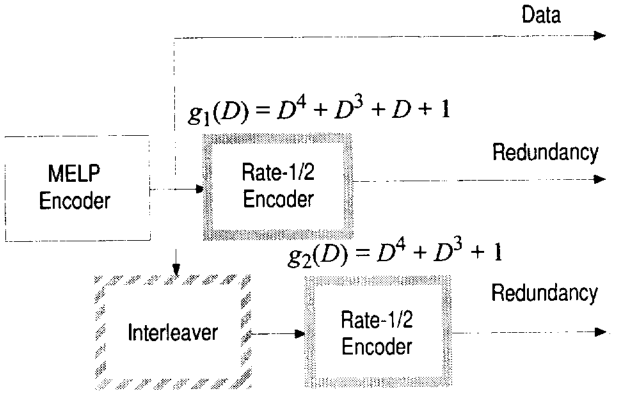


Figure 8. Turbo Coding Transmitter

We conducted simulations using plain Turbo coding and Turbo coding enhanced with MELP's residual redundancy information. A "classic" Turbo decoder produces for each data bit b a "log-likelihood" ratio $L(b)$:

$$L(b) = \log(P(b = 1|\mathbf{r}) / P(b = 0|\mathbf{r}))$$

Here, $\mathbf{r}$ is the received sequence. In order to keep complexity manageable, our (sub-optimal) approach is to use these log-likelihood ratios produced by the Turbo decoder in conjunction with the *a priori* probabilities obtained from the training sequence to estimate the MELP sequence (rather than using the *a priori* probabilities to compute the log-likelihood ratios). Given a random binary 3-tuple $\mathbf{b} = [b_0 b_1 b_2]$, we associate the following reliability metric with $\mathbf{b}$:

$$L(\mathbf{b}) = \sum_{i=0}^{2} \log[P(b_i = 1|\mathbf{r}) / P(b_i = 0|\mathbf{r})]$$

Now assume a first order Markov relationship among a sequence of binary triples - i.e., assume the statistical model we have been using for the high-order bits in the MELP gains, pitch, and MSVQ1. Let $\{\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2, ...\}$ be the sequence of binary triples, then the metric that we employ in our simplified ad hoc approach is

$$M(\mathbf{b}_i) = L(\mathbf{b}_i) + \log\left[\frac{P(\mathbf{b}_i \mid \mathbf{b}_{i-1})}{P(000 \mid \mathbf{b}_{i-1})}\right]$$

Here, $L(\mathbf{b}_i)$ is the contribution computed from the Turbo decoder's log-likelihood values and the second term accounts for the *a priori* transition probabilities computed from the training sequence.

Figure 9 shows the results obtained from the Turbo coding experiments in the case of the mean square error of the Gain 1 parameter. The testing conditions were under independent fading or the same as for Figures 1,3, and 5, and are thus directly comparable. In the region above 4 dB Eb/No (i.e., from 4 to 7 dB), we see that the Turbo coding performs about the same (even slightly better than the Soft MAP decoding). However, in our prime region of interest of 2 to 5 db Eb/No, neither the plain vanilla Turbo nor the enhanced Turbo decoding (using the MELP redundancy features) perform as well as the Soft MAP process. It is only in the region of less than 1 dB Eb/No where Enhanced Turbo outperforms Soft MAP. It is generally gratifying to see the divergence behavior of the Enhanced Turbo cluster of curves below that of the plain vanilla Turbo curve cluster, in addition to the fact that the Turbo decoder is not overly sensitive to frame delay required at the decoder.

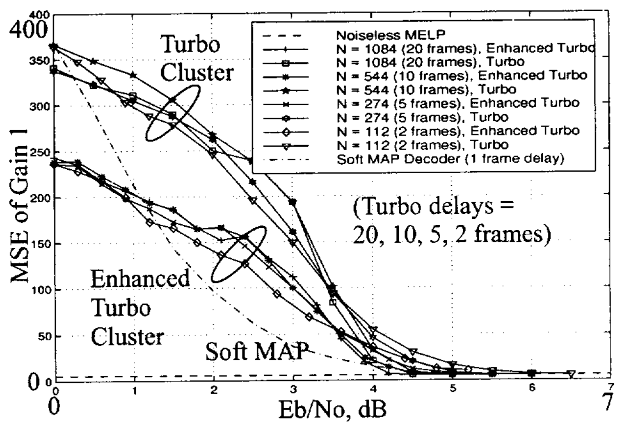


Figure 9. Gain1 Results of Turbo Coding in Fading

Figure 10 shows the results obtained from the Turbo coding experiments in the case of the mean square error of the Pitch parameter. Notice that the divergence of the two curve clusters is even more pronounced than in the case of the Gain 1 parameter.

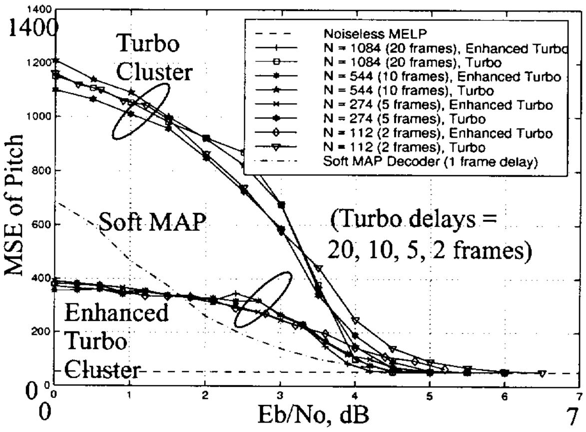


Figure 10. Pitch Parameter Results (GN)

In summary, Turbo codes use iterative applications of two a posteriori probability (APP) decoders, working with the two constituent codes - one for the original data and another for the randomly interleaved data. These two decoders share "extrinsic information" to gradually improve their estimates of the transmitted data. In the new approach, the decoder working on the un-interleaved data makes use of the transition probability estimates derived from the training sequence to improve its extrinsic information. These preliminary results indicate to us that Turbo decoding is an area for further research and potential added gains in performance.

SUMMARY & CONCLUSIONS

It has been demonstrated that the inherent residual redundancy characteristics of the Fed STD MELP parameters may be utilized to enhance the performance of MELP over very poor transmission tactical channels, at the cost of one additional frame delay. MELP will have application in many government communication equipments, including a variety of tactical military systems which will operate over terrestrial and satellite radio frequency channels. These include HF channels and VHF/UHF cellular-like channels, potentially under jamming and interference effects. One potential application scenario would be in noisy fringe reception areas with deep and rapid Rayleigh fading due to fast vehicle motion. Demonstration of MELP over these such channels was shown to be feasible, when sophisticated Soft MAP error correction schemes were applied. If additional frames of delay are allowed for interleaving, then any rate of fading can be made to appear like the independent Rayleigh fading addressed in the described experiments. The efficacy of the developed algorithm has been shown via objective metrics, BER performance, and subjective listening.

Possible other techniques that may be applied to MELP error correction include Turbo codes, which require a considerable amount of additional delay (on the order of up to 500 msec) to perform well, depending on conditions. However, this may not be problematic in Half Duplex operational situations. For Full Duplex (simultaneous two-way) short delay communications, the Maximum A Posteriori MAP technique shows promise.

REFERENCES

(1) Supplee, L.M., Cohn, R.P., Collura, J.S., McCree, A.V., *"MELP: The New Federal Standard at 2400 bps"*, IEEE ICASSP-97 Conference, Munich Germany, pp.1591-1594.

(2) Alajaji, F., Phamdo, N., Fuja, T., *"Channel Codes That Exploit the Residual Redundancy in CELP-Encoded Speech"*, IEEE Transactions on Speech and Audio Processing, Vol.4 No.5, September 1996, pp. 325-336.

(3) McCree, A., Truong, K., George, E., Barnwell, T., Viswanathan, V., *"A 2.4 kbit/s MELP Coder Candidate for the New U.S. Federal Standard"*, IEEE ICASSP-96 Conference, Atlanta Georgia USA, pp. 200-203.

(4) Rahikka, D.J., Fuja, T.E., Fazel, T., *"Optimized Error Correction Of MELP Speech Parameters Via Maximum A Posteriori (MAP) Techniques"*, IEEE SCW-99 Speech Coding Workshop, Porvoo Finland, June 1999.

(5) Rahikka, D.J., Fuja, T.E., Fazel, T., *"Channel Coding Of Fed Std MELP Vocoder Utilizing Residual Redundancy Features"*, IEEE VTC-99 Vehicular Technology Conference, Houston Texas, May 1999.

(6) Berrou, C., Glavieux, A., Thitimajshima, P., *"Near Shannon Limit Error Correcting Coding and Decoding: Turbo Codes"*, IEEE ICC-93 Conference, pp.1064-1070.

STILL IMAGE TRANSMISSION IN A LINK 16 NETWORK USING EMBEDDED ZEROTREE WAVELET (EZW) ALGORITHM

MANUEL MARTINEZ-RUIZ. -C.I.D.A.- (M.O.D). Arturo Soria St 28033 -MADRID -SPAIN

ANTONIO ARTÉS-RODRIGUEZ.-UNIVERSIDAD DE ALCALÁ.-MADRID -SPAIN

SECTION I: SUMMARY

The transmission of tactical images via Link 16 standard is meant to be of a great operational importance.

It is necessary to reach acceptable image compression rates to optimize the Link 16 network resources: the time slots corresponding to the TDMA architecture. These time-slots are also shared by the Link 16 tactical messages in accordance to the STANAG 5516.

This paper addresses an application to image transmission via Link 16 Standard for tactical data communication based on the Embedded Zerotree Wavelet algorithm, sending with higher anti-jamming and error detection and correction (EDC) protection the parts of the image with higher impact on image quality. The Link 16 transmission protocol as well as the Embedded Image Coding allow the development of a combined algorithm to take advantage of both concepts.

We present a Link 16 messages type allocation algorithm matched to the multirresolution representation of images obtained by the EZW algorithm.

A table is shown with the different messages packing limit and their corresponding data rate. It is also shown both the different anti-jamming and the EDC capabilities. Each packing type requires an accurate data rate and this is in line with the EZW algorithm characteristics.

The paper organization is as follows:

the operational aspects for image transmission over any data link system are presented in Section II.

Section III dicusses the EZW algorithm. A hierarchical organization of the information is desirable in order to obtain a multirresolution representation of the still image to be transmitted.

Section IV discusses the features of the Link 16 transmission. The Link 16 TDMA architecture is based on a time slot structure. Several data rates are allowed in a slot-by-slot basis. The above mentioned presents operational involvement in both the anti-jamming protection and the System ECD features.

The operational range is also related with the Link 16 message structure and therefore it depends on both the anti-jamming capability and the data rate..

Section V discusses the allocation algorithm of the different data rates to the image's hierarchical representation, based on the EZW approach.

Finally, a conclusion and several areas of further research are presented.

SECTION II. OPERATIONAL ASPECTS OF IMAGE TRANSMISSIONS OVER A DATA LINK SYSTEM

Data link systems have been designed to convey tactical information between command and control systems. Special functions such as voice and video imagery might be implemented if the data link system had enough resources. Link 16 Systems can transmit digitized voice and as well as both formatted and free text messages over the channel.

The quality of an image depends on the following parameters:

1.-The size of the image

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

2.-The "gray scale" or "color" features.

3.-The compression ratio. We need to compress the image in order to maximize the data link performance.

4.-the parameters associated with the data link protocol such as time slots allocated to image transmission, error detection and protection features (EDC) , packing structure and time slots assigned to each platform.

The operational aspects of image transmissions over a data link channel also depend on the time and antijamming requirements.

The "transmission time" is one of the most important constraints imposed by operational application to the data link system. Real-time or near real-time image transmission is meant to be a strong requirement.

The anti-jamming capability of the data link system is extremely important because most of the present and near future operations are used to take place in a jammer environment.

The image progressive transmission also allows to implement layered protection schemes.

.SECTION III. EMBEDDED ZEROTREE WAVELET ALGORITHM.

For rapid transmission from a data base, it should be desirable to quickly provide a coarse signal approximation that is progressively enhanced as more bits are transmitted. The embedded coders offer this flexibility by grouping the bits in order of significance. This means that the coefficients are sorted and the first bits of the largest coefficients are sent first. An image approximation can be reconstructed at any time, from the bits already transmitted.

These coders can take advantage of any prior information on the location of large versus small coefficients. Such information is available for natural images decomposed on wavelets bases.

In a typical operational environment it should be desirable to address the problem of obtaining the best image quality for a given bit rate in an embedded fashion.

This approach is also applicable to the transmissions over a noisy channel (jamming environments) in the sense that the ordering of bits in order of importance leads naturally to layered protection schemes.

The EZW algorithm presents also the following properties:

-Discrete Wavelet Transform (DWT) or hierarchical sub-band decomposition which decorrelates most image sources and provides a multirresolution representation of the image.

-Zerotree coding which provides significant coding gains. To improve the compression of significance maps of wavelet coefficients a new data structure is defined called zerotree.The Zerotree is based on the hypothesis that if a wavelet coefficient at a coarse scale is insignificant with respect to a given threshold, then all the wavelets coefficients of the same orientation in the same spatial location at finer scales are likely to be insignificant with respect to the threshold.

With this approach, the transformed coefficients can be coded as: positive symbol (PS), negative symbol (NS), isolated zero symbol (IZS), and Zerotree root symbol (ZRS).

-Entropy-coded successive approximation quantization to perform the embedded coding. A losless scheme can be achieved

-Prioritization protocol to establish the order of data processing. The EZW algorithm has the property that the bits in the bit stream are generated in order of importance

By using an embedded coding algorithm an encoder can terminate the encoding process at any given time. For this reason a target rate and a target distortion can be met exactly. The decoder can cease decoding at any point in the bit stream and still produces image with the quality reached by the decoded bit stream.

This is one of the most important features used by our algorithm because an exact bit rate is required in a Link 16 network to transmit the information in a time-slot (TDMA) architecture

SECTION IV. NATO LINK 16 TRANSMISSION CHARACTERISTICS

Link 16 uses the principle of time division multiple access (TDMA) . The TDMA architecture uses time interlacing to provide multiple and apparently simultaneous communication nets. All MIDS units are preassigned sets of time slots in which to transmit their data and to receive data from other units. Each time slot is 1/128 second in duration (7.8125 msec). There are 51 frequencies available for MIDS transmissions and the frequency is changed rapidly during each time slot, according to a predetermined pseudo-random pattern (frequency hopping spread spectrum technique).

In link 16 either 3,6,12 link 16 words can be transmitted in a time slot, depending on the data packing structure that is being used. Each link 16 word comprises 70 bits of data and a link 16 message is composed of a variable number of words (1,2,3, and more). Each message consists of a message header and message data. The header is not considered a part of the message structure and specifies the type of data and the source track number of the transmitting MIDS/JTIDS terminal. There are four types of messages:

-Fixed format: J-series messages as specified on Stanag 5516

-Variable format: user-defined messages

-Free-text: unformatted messages used for link 16 digitized voice

-RTT : used for synchronization

We are using free-text messages because these are unformatted and there is no parity process associated to associated them. They may or may not be Reed-Solomon encoded

The throughput, range and anti jamming features of link 16 transmissions depend on the packing structure

The throughput is determined by the density of message packing: 3, 6 or 12 words.

Range can be normal (300 nm) or extended (500 nm).

The normal range can be accommodated by all the packing structures, but the extended range can only be used with Std and P2SP. In the extended range mode, the jitter dead-time is shortened to allow a longer period for propagation of the signal. Some AJ capability is lost with extended range. The AJ capability of the link 16 transmissions depends on the jitter and redundancy (DP) . For this reason, the amount of AJ margin decreases as throughput increases. With the P4-SP packing structure the system provides the greatest throughput with the least amount of AJ margin.

The MIDS terminal selects the packing structure in the following order of preference: STD-P2SP-P2DP-P4.

The Packing limit initialization parameter specifies how much AJ margin can be given up during any time slot. The host system can override this limit for specific time slots by means of functional input messages (Initialization data changes). The message structures are the following:

-STD-DP: Standard double pulse. It represents 225 bits of coded information or 465 bits of uncooked data. The redundancy of this packing structure makes it the most reliable way of transmitting tactical information.

-P2-SP: Packed-2 Single Pulse. This transmission mode is not as jam-resistant as Std-DP because the data is transmitted only one time (Single pulse). As a result, this packing structure allows to double the data capacity.

-P2-DP: Packed-2 Double pulse. Data redundancy is like Standard-DP mode and data capacity is like P2-SP. The jitter is eliminated so the transmission could be jammed by synchronized jammers.

-P4: Packed-4. This packing structure is the least jam-resistant of all.

The following Table shows the different parameters involved in a Link 16 transaction. The antijamming capability is coded as follow:

-Double pulse→ (+)

This A-J technique take advantage of the Link 16 frequency hopping spread spectrum capability and it is used to repeat the same piece of information at a different frequency.

-Jitter→ (+)

This is an A-J technique against the jammers synchronized with the time-architecture of a Link 16 network.

-Error Detection and Correction (EDC) features: Reed Solomon Coding→(+)

The following Table presents the antijamming capability in accordance with the above mentioned code:

Time Slot Structure	Free Text/Fixed Format	EDC Coded	Data bits/Time Slot	Data Rate (bps)	Anti Jamming Capability	Range Mode
STD	FF	Y	225	28,800	+++	N/E
STD	**FT**	**Y**	**225**	**28,800**	**+++**	**N/E**
STD	**FT**	**N**	**450**	**57,600**	**++**	**N/E**
P2SP	FF	Y	450	57,600	++	N/E
P2SP	**FT**	**Y**	**450**	**57,600**	**++**	**N/E**
P2SP	**FT**	**N**	**930**	**119,040**	**+**	**N/E**
P2DP	FF	Y	450	57,600	++	N
P2DP	**FT**	**Y**	**450**	**57,600**	**++**	**N**
P2DP	**FT**	**N**	**930**	**119,040**	**+**	**N**
P4	FF	Y	900	115,200	+	N
P4	**FT**	**Y**	**900**	**115,200**	**+**	**N**
P4	**FT**	**N**	**1,800**	**238,080**	**-**	**N**

The Link 16 features can be summarized as follows:

-Architecture

Time Division Mode Access (TDMA) architecture.

No critical nodes. It means that a "Net Control Station" is no longer needed

Message and transmission encryption

Spread spectrum techniques via frequency hopping and direct sequencing (CCSK)

UHF and Line of sight (LOS)

-Message types

Fixed Format

Free text

Variable Format

-Operational Use. The link 16 tactical information is classified in the so called "Network Participating Groups" (NPGs). This is a useful way to design the Link 16 networks. The most common NPGs are:

Surveillance

Electronic warfare

Mission management

Weapon coordination

Air control

Fighter to Fighter

Secure voice

Navigation

Positive identification

SECTION V. TIME SLOT'S TO IMAGE ALLOCATION ALGORITHM BASED ON LINK 16 TDMA PROTOCOL

The Link 16 TDMA architecture is based on a time slot structure. Several data rates are allowed in a slot-by-slot basis. The last means operational involvement in both the anti-jamming protection and the System error correction and detection (EDC) features.

The algorithm we present is based on this protocol and it allows:

1.-To transmit firstly the "trends" of images at a given data rate and with maximum noise and ECM protection.

2.-To transmit secondly the "anomalies" at a higher data rate and with lower ECM protection.

3.-To match the coding data rate to the Link 16 network bit rate.

A decision has to be taken to match our algorithm to the Link 16 protocol, for instance:

-Free-Format messages Vs Fixed-Format messages.

-Different "Packing Limit" to transmit "trends" and "anomalies". The message package format types chosen are the following

6(0): F-T std with EDC (225 bits per time slot)

6(1): F-T p2sp with EDC (450 bits per time slot)

2(0): F-T p2dp with EDC (450 bits per time slot)

2(1): F-T p4 with EDC (900 bits per time slot)

0(0): F-T std without EDC (465 bits per time slot)

0(1): F-T p2sp without EDC (930 bits per time slot)

1(0): F-T p2dp without EDC (450 bits per time slot)

1(1): F-T p4 without EDC (900 bits per time slot)

-Different Anti jamming protection to each piece of image information. As we have mentioned in section IV, the message package format types provide different antijamming protection. Our algorithm will be focussed on providing both the maximum protection to these parts of the information that contains trends and the minimum protection and maximum data rate to the piece of information that contents anomalies.

-Range Mode:Normal (300 miles) Vs Extended (500 miles). The range mode is imposed by the message package format. The p2dp message structure only allows normal range operation (300 miles), but it provides more anti jamming capability than the p2sp, due to the fact that the same piece of information is repeated at different frequencies. In our algorithm a extended range operation is allowed in order to use all the free text message package formats.

-Error Detection and Correction features: Y/N. The bit error rate (BER) of the channel is measured considering the systems error detection and correction features. If we initialize our system to implement EDC techniques the B.E.R of the channel is lower than without EDC techniques.

Due to the fact that the most important part of the information is contained in the bit significance map (see section III), we are using message structures with EDC features in the allocation of the dominant pass. The subordinate pass contains refinement information about the magnitudes of the significant coefficients and this can be transmitted without EDC techniques in order to send more bits per time slot.

The next step is to obtain the number of "time slots" needed to transmit a still image. Both the parameters showed in Table I and the time constraint imposed by the operational application shall be taken into account.

The optimization criteria to obtain the "best" performance of our application is as follows:

To minimize the number of time slots required. The fact that the time slots left to image transmission is suppose to be minimum must be taken into account.

To maximize the anti-jamming protection to the overall image.

To establish a time constraint that would be imposed by the operational environment. This means that it is necessary to establish a time limit to receive all the image information.

Therefore, the allocation algorithm requires the following inputs:

Percentage of the Link 16 resources than will be used for transmission of images. This will have impact in the amount of time slots allocated for tactical data information.

Time limit to transmit the compressed image that is clearly an operational input. In our algorithm we are using 12 seconds because this is the Link 16 frame time-unit and it is very useful for network designers.

The steps of the allocation algorithm is as follows(see flow chart in annex 1):

Step-1. Input the percentage of time slots assigned to images transmission.

Step-2. Input the time limit to transmit the compressed image

Step-3. input the image to transmit.

Step-4. Run the EZW algorithm. As a result the following information is needed to initialize our allocation algorithm:

4.1. Number of EZW algorithm passes: (e.g.: 1..10)

4.2. Successive-approximation quantization (SAQ) Maximum Threshold.

4.2.-For each EZW pass, save the number of bits of both the dominant and the subordinate lists.

Dominant pass 1 (1P): n1_ bits

Subordinate pass1 (1S): n2_bits

Dominant pass2 (2P): m1_bits

Subordinate pass2 (2S): m2_bits, and so on...

At the end of step 4 we have 2 files per pass ("one data file " for the dominant and "one data file" for the subordinate). The algorithm computes and saves the size of these files. This piece of information is needed to initialize our time-slot allocation algorithm.

Step-5. Create the control template. The packing structure allocation algorithm is initialized by means of a table to control the message structure assignments to the EZW encoded files. The control template contents the maximum percentage of time slots dedicated to convey data with a given packing structure.

This table is based on the size of the files obtained in step 4.2. The percentage of time slots allocated to images are further subdivided in accordance with the control template. For example:

x% of the T_S assigned to image transmission ➔packing limit: stander

y% of the T_S assigned to image transmission ➔packing limit: packed 2 single pulse

z% of the T_S assigned to image transmission ➔packing limit: packed 2 double pulse

k%of the T_S assigned to image transmission ➔packing limit: packed 4

Step-6. Implement the "packing structure" allocation algorithm.

The allocation algorithm is based on the link 16 system capability to code the messages with or without error correction and detection features in a time-slot by time-slot basis.

The link 16 system meets a communication performance requirements for both the error rates and the jamming resistance.

For the error rates, the system B.E.R for a channel with EDC techniques is lower than the B.E.R. for a channel without these features. Due to the fact that the most important part of the information is conveyed in the dominant pass, this pass is coded with the EDC features provided by the system.

For the anti jamming resistance, this shall be reduced as the EZW encoder codes the successive iterations. The images trend information is obtained in the first pass, therefore it is extremely important to provide the most anti jamming protection to the data conveyed in this iteration. As it has been mentioned before, the trade-off between the antijamming techniques is the data rate. Therefore, the first pass file is sent at the minimum data rate (28,8 KBPS). The subsequent iterations, conveyed image 'anomalies', are sent at higher data rates with less anti-jamming protection.

The "packing structure" allocation algorithm (step-6) can be summarized as follows:

-Dominant passes P: send the data over message package format types with EDC features:

6(0), 6(1), 2(0), 2(1)

-Subordinate passes S: : send the data over message package format types without EDC features:

0(0), 0(1), 1(0), 1(1).

-The first pass is sent using a standard package format.

Pass_1P: STD+EDC

Pass_1S: STD-EDC

-for pass_2 to pass_N

If (number_of_ts) > y,(z, k).

" code with p2sp,(p2dp, p4) package format."

Elseif "code with p2dp, (p4), package format"

Else "re-initialize the algorithm"

end

the transmission process ends when the number of time slots that have been used to send the image reaches the limit imposed by the operational constraint.

SECTION VI: PRELIMINARLY SIMULATION RESULTS

In this preliminary study we have implemented a more simplified algorithm (algorithm#2) to allocate time slots to the several parts of the EZWA coded image.

There are three control tables each of them with different packing structure assignments.

Table_1: Maximum Anti-jamming capability

Pass1 .. Pass3 ➔std with EDC

Pass.. Pass6➔ p2sp (dominant pass: EDC, subordinate pass: No_EDC)

Pass7..Pass10 ➔ p4 (dominant pass: EDC, subordinate pass: No_EDC)

Table_2: Medium Anti-jamming capability

Pass1 & Pass2 ➔std with EDC

Pass3 ..Pass6➔p2dp (dominant pass: EDC, subordinate pass: No_EDC)

Pass7..Pass10 ➔ p4 (dominant pass: EDC, subordinate pass: No_EDC)

Table_3: Minimum Anti-jamming capability

Pass1..Pass2 ➔std with EDC

Pass3 ➔ p2sp (dominant pass: EDC, subordinate pass: No_EDC)

Pass4..Pass10 ➔ p4 (dominant pass: EDC, subordinate pass: No_EDC)

As it was mentioned before, the communications performance capability for a link 16 system depends on the message package format types.

The simulation results are based on the following assumptions

-Image_size= 512 X 512 pixels/8 bits pixel

-Wavelet Decomposition: Seven (7) levels of multirresolution decomposition

-Compression Rate➔ Depends on the number of EZW iterations

-Number of EZW algorithm iterations: from 1 to 10

-Initial EZW threshold : 4096

-Time constraint: 12 sec

The figures #1 shows the different time slot allocations for each EZWA pass parameterized by

-Percentage of Link 16 timeslots allocated to images.

-fixed control table (no adaptive)➔ different antijamming features.

-Compression rate.

We have also modelled the Data Link channel in order to assess the possibility to improve our allocation algorithm.

To simulate the link16 channel the first step is to obtain the bit error rate (B.E.R) for both data channels: with EDC and without EDC. The noise of the channel is simulated by mean of the following expressions:

Channel with EDC features: Error_edc=rand (1,word_length) < B.E.R._edc

Channel without EDC features: Error_no_edc=rand(1, word_lengh) < B.E.R._no_edc

The next step is to obtain the data package from the EZW encoder. We have used an optimized free software package developed by Universidad de Alcalá de Henares, Escuela Politécnica (Spain).

The received data is obtained via XOR logical operator between the "transmitted data package" and the "simulated channel noise". This data is the input for the EZW decoder and finally the received image is displayed.

The application allows to check the number of errors for each EZW iteration. Based on this information the user can simulate the performance of the channel in order to establish a threshold to abort (halt) the images coding process if the errors in the dominant pass are not acceptable.

SECTION VII. CONCLUSION AND AREAS OF FURTHER RESEARCH

We have presented a new approach to transmit tactical images over a tactical data link 16 channel. The embedded Zerotree wavelet algorithm can be used in systems that have the possibility of transmitting data at multiple bit rates. We present a Link 16 messages type allocation algorithm that matches with the multirresolution representation of images obtained by the EZW algorithm. Using an embedded code, an encoder can finish the encoding process when a target rate is reached and this is in line with the layered protection scheme we propose.

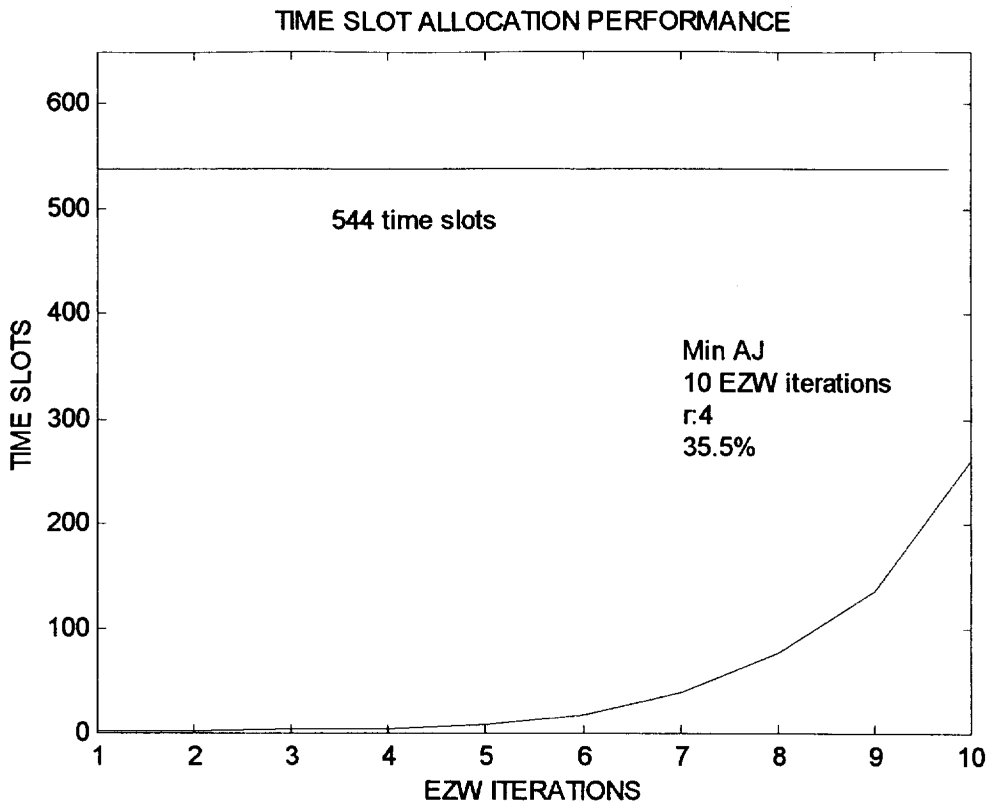
There are several areas of further research

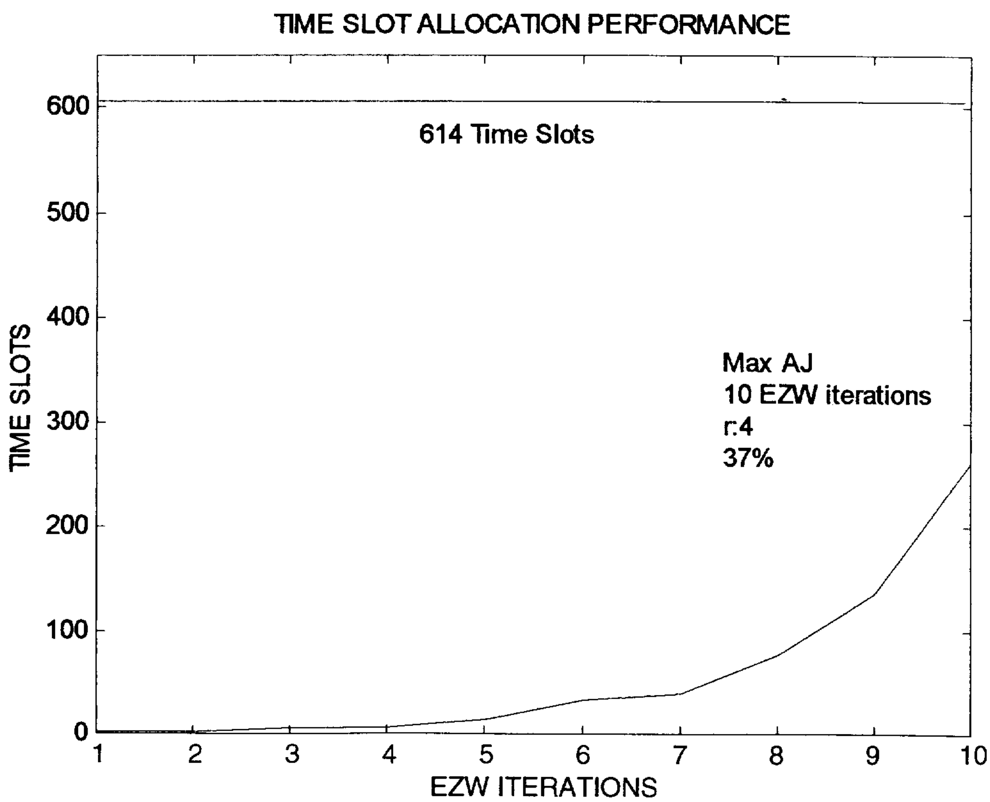
1.-To optimize the allocation algorithm matched to each specific image type

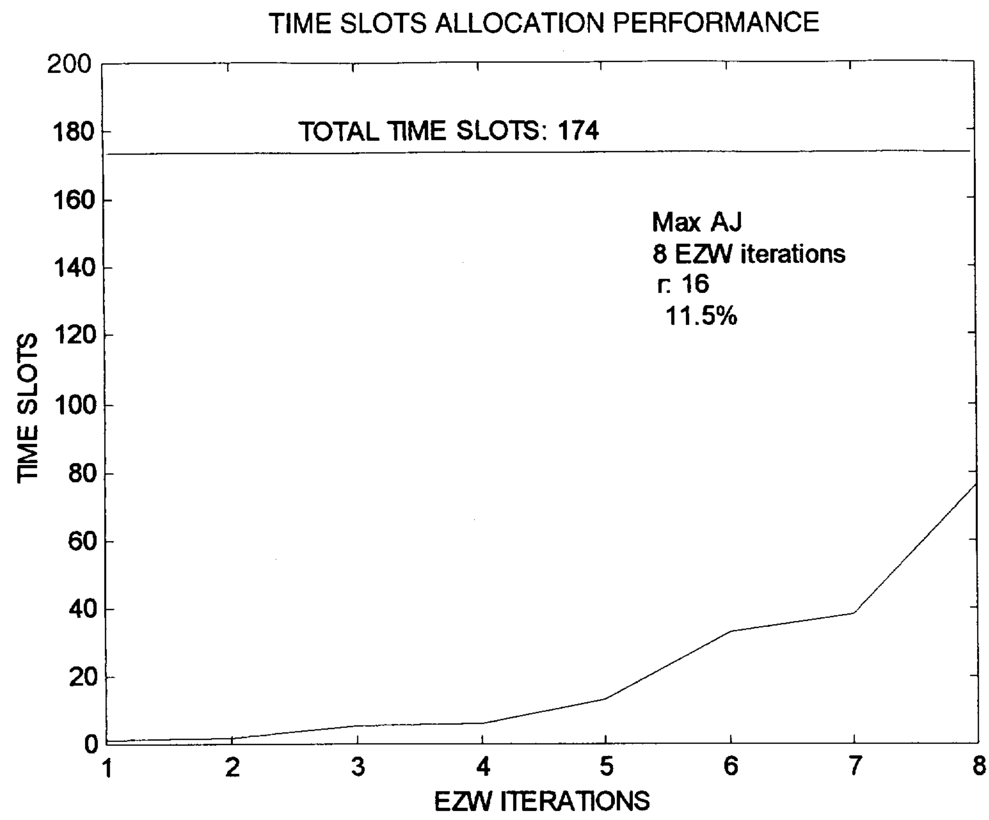
2.-Zoom-in capability on selected areas. This implies a back-channel for voice or messages.

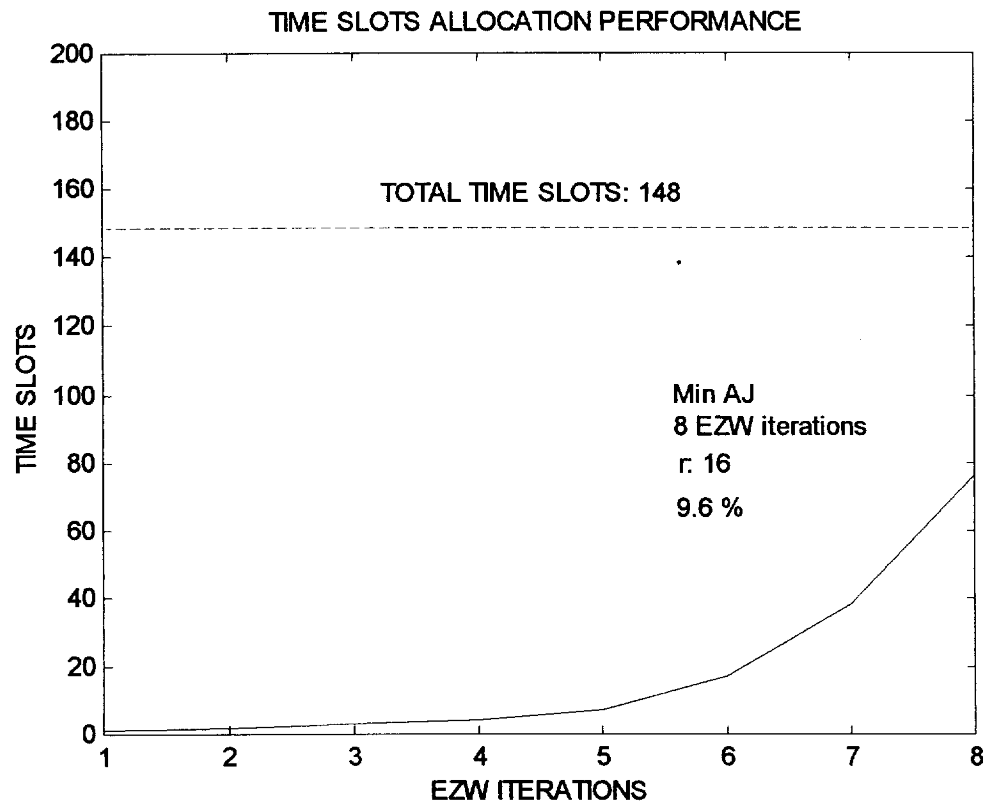
3.-Comparative analysis of an EZW based algorithm or a SPIHT based algorithm.

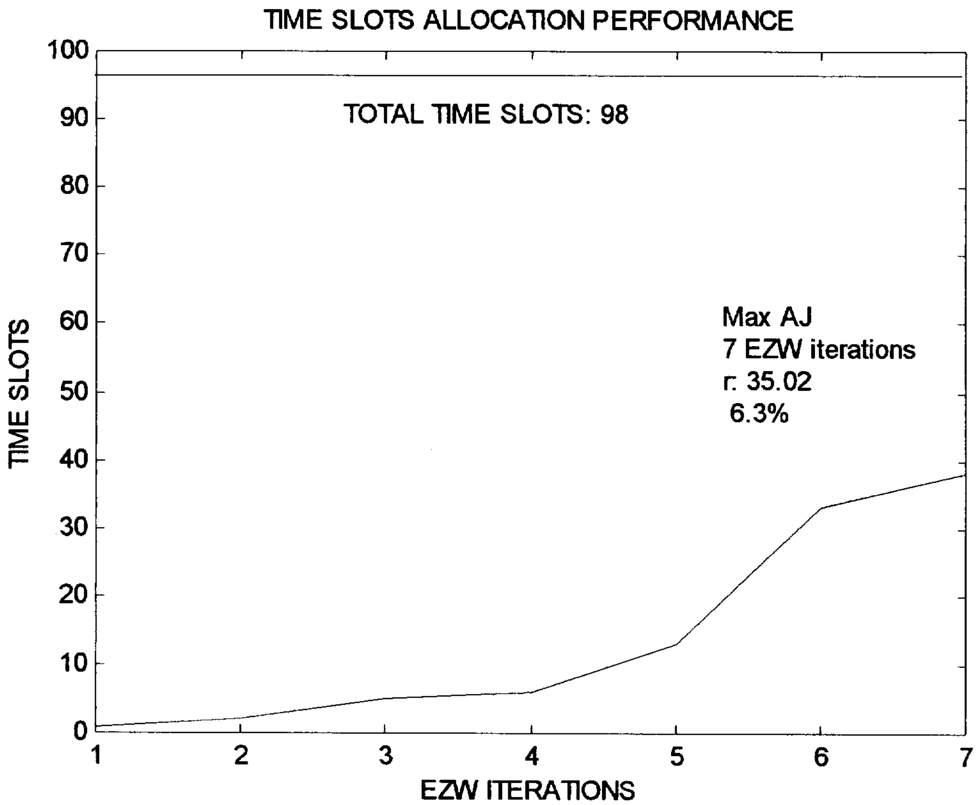
FIGURE: 1 TIME SLOT ALLOCATION PERFORMANCE:

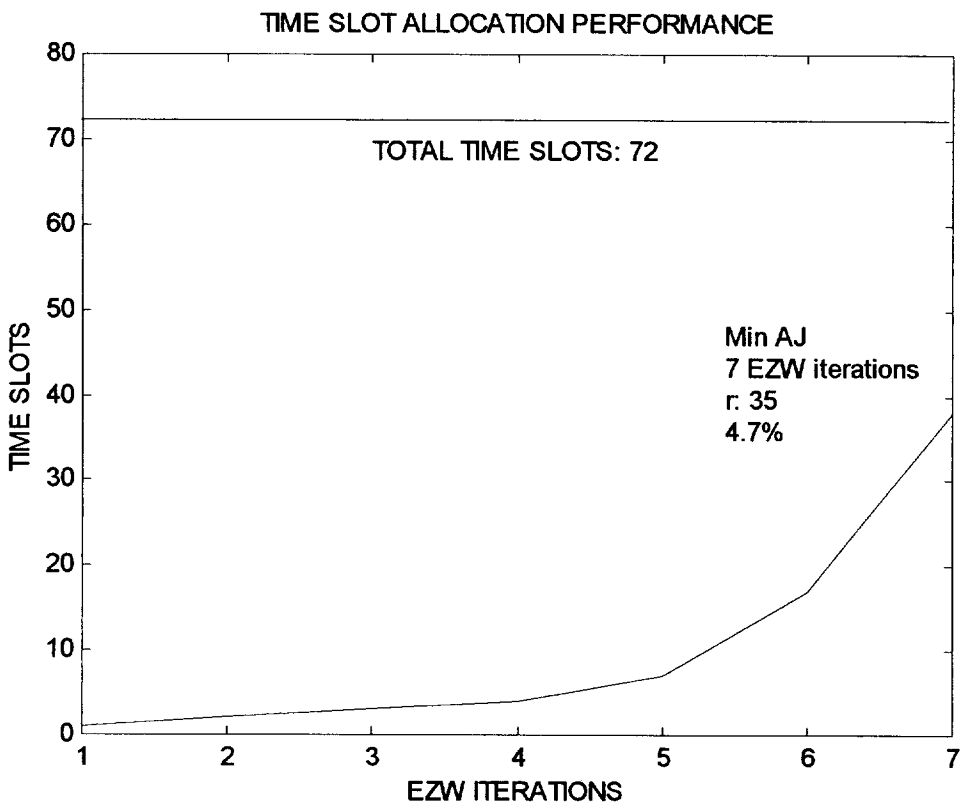





TIME SLOTS ALLOCATION PERFORMANCE
TOTAL TIME SLOTS: 174
Max AJ
8 EZW iterations
r: 16
11.5%
TIME SLOTS
EZW ITERATIONS
200
180
160
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8


TIME SLOTS ALLOCATION PERFORMANCE
TOTAL TIME SLOTS: 148
Min AJ
8 EZW iterations
r: 16
9.6 %
TIME SLOTS
EZW ITERATIONS
200
180
160
140
120
100
80
60
40
20
0
1
2
3
4
5
6
7
8


TIME SLOTS ALLOCATION PERFORMANCE
TOTAL TIME SLOTS: 98
Max AJ
7 EZW iterations
r: 35.02
6.3%
TIME SLOTS
EZW ITERATIONS
0
10
20
30
40
50
60
70
80
90
100
1
2
3
4
5
6
7


TIME SLOT ALLOCATION PERFORMANCE
TOTAL TIME SLOTS: 72
Min AJ
7 EZW iterations
r: 35
4.7%
TIME SLOTS
EZW ITERATIONS
0
10
20
30
40
50
60
70
80
1
2
3
4
5
6
7

REFERENCES

[1] "Embedded Image Coding Using Zerotress of Wavelet Coefficients" ;Jerome M. Shapiro

IEEE TRANS ON SIGNAL PROCESSING VOL 41 No 12 DECEMBER 1993

[2] NATO STANAG 5516 e.d. 2

[3] "A Wavelet Tour of Signal Processing" ; Stéphane MALLAT

ACADEMIC PRESS (1998)

HF Serial-Tone Waveform Design

M.B. Jorgenson and K.W. Moreland
Communications Research Centre
3701 Carling Ave.
Ottawa, Canada, K2H 8S2

Summary

The intent of this paper is to provide the reader with an understanding of the fundamental aspects of HF serial-tone waveform design. Factors which affect the design of a waveform are discussed. Some factors are regulatory, for example, the availability of bandwidth. Others are a consequence of nature, such as the characteristics of the HF channel. The technology used in the detection of serial-tone signals plays a key role in the design of the waveform. Dach of the component elements of a serial-tone modem is briefly described. Empirical waveform performance rules are presented and the design process is illustrated with an example based on the development of a 9600 bps HF waveform standard.

1.0 Introduction

With the increased demand for high data rate HF services, HF waveform design is enjoying a renaissance of sorts. New HF waveforms have been developed for NILE/Link-22 and STANAG 4444, and are being developed for STANAGs 4538, and 4539. A serial-tone standard for data rates up to 9600 bps was provided in STANAG 5066 Annex G as a non-mandatory component of the standard. The Annex G waveform was a commercial success in that it was implemented by vendors, even before adoption. When Annex G was removed from STANAG 5066, the same waveform, with minor modifications to the preamble, interleaver, and the addition of another intermediate data rate, found its way into draft Mil-Std 188-110B.

This paper will provide a tutorial guide to designing an HF serial-tone waveform and is intended to give the reader sufficient background to make judgements regarding the suitability of various proposed waveforms. We will begin by describing the components which make up a waveform and the key factors to consider in waveform design. For reference purposes, a brief description of the types of modulation commonly used at HF will be provided before focusing on the serial tone waveform. The essential elements of a modern serial-tone modem will be described, as will their impact on the design of a waveform. Empirical relationships for predicting the bit-error-ratio (BER) performance of serial-tone modems in a Doppler spreading environment are provided. Finally, the paper will conclude with a waveform design example, based on the authors experience in specifying Annex G of STANAG 5066.

2.0 What is a waveform?

A waveform specification is a description of the on-air signalling used to transmit a digital data signal over a radio channel. It includes a complete specification of the modulation to be used and prescribes the known symbols, commonly referred to as initial training (IT) or preamble, that are sent to establish synchronization as well as any other known symbols which may be inserted to aid in the demodulation process. Many standards, particularly more recent efforts, include forward error correction coding and data interleaving as an integral part of the definition.

As a rule, waveform standards only specify the details of the waveform to be transmitted. It is up to vendors to determine how best to demodulate the signal, although some standards, notably the US Mil-Stds, specify minimum performance figures which must be obtained in order for a vendor's equipment to be considered compliant with the various standards.

3.0 Factors affecting waveform design

A number of factors affect the design of a waveform. One of biggest constraints for HF waveform design, particularly in recent years as the emphasis has shifted to higher data rates, is the channel bandwidth. For military uses, HF spectrum is typically allocated with a 3 kHz channelization, although

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

there are exceptions: some naval broadcasts operate in a 1.24 kHz bandwidth and there are some assignments of 6 kHz for use with independent sideband radios (typically for Link-11). 3 kHz is a respectable bandwidth for sending a 75 bps data signal, typified by the naval broadcasts run by many NATO nations, but it becomes a very narrow piece of spectrum for users who are attempting to transmit at data rates of up to 9600 bps.

The unique characteristics of the HF channel itself offers significant challenges. The ionospheric refraction which allows HF radio signals to propagate over long distances is not without its shortcomings. The received sky-wave signal may suffer distortion in the form of temporal dispersion (delay spread) as well as fluctuation in the signal's amplitude and phase (Doppler spreading). Recent high latitude DAMSON measurements have observed multipath signals of more than 10 ms duration and other signals have shown evidence of Doppler spreading greater than 50 Hz [1,2]. More typical mid-latitude sky-wave channels might show delay spreads of 1 - 4 ms with Doppler spreads of 1 Hz or less. In addition to the sky-wave channel, the HF surface wave channel offers interesting features and challenges. Over sea-water, the HF surface wave propagates far beyond line-of-sight, offering intriguing capabilities for Naval forces. As the surface wave begins to weaken at the periphery of the surface wave coverage region, a Rician channel is observed, with the non-fading component from the surface wave, and another, fading component, arising from a sky-wave path. The noise environment in the HF channel is also somewhat unique. CCIR Recommendation 322 provides a model for the HF noise environment. In general, it is much more impulsive than additive white Gaussian noise, with a much higher peak to mean ratio and tends to introduce burst error events.

The available bandwidth and the channel characteristics serve to limit the data rates which are achievable over HF. Until very recently most naval broadcasts were run at 75 bps, and were often unreliable even at these rates because of the lack of forward error correction (FEC) coding. With the adoption of the modern serial-tone modem, naval broadcasts are moving to a 300 bps data rate and other services are being provided at rates up to 2400 bps. The demand for increased data rates imposed by modern networking protocols has led to a determined effort to push achievable data rates upward and has resulted in new and developing standards for waveforms offering rates of 9600 bps and greater. The context in which these systems are being developed is also changing. At one time, EMCON considerations were considered paramount and most HF transmissions were one way, with no acknowledgment to give away information on the recipient's position. Increasingly, military forces are considering options which involve radiating detectable emissions for substantial portions of the time during operations. This evolving change in philosophy, coupled with the advent of modern automatic repeat request (ARQ) systems, has led to the transmission of data types which require error free reception (executable computer programs, for example).

4.0 Modulation Types

One way of characterizing modulations is to describe them as either bandwidth efficient or power efficient. Examples of bandwidth efficient modulations are phase-shift-keying (PSK), quadrature amplitude modulation (QAM) and multi-carrier modulations. Power efficient modulation types include direct sequence spread spectrum and M-ary frequency shift keying (MFSK). It is noteworthy that all of the above listed modulations have been used at HF. Older naval fleet broadcasts, some of which are still in service today, used an uncoded 75 bps 2-FSK modulation. Both STANAG 4415 and Mil-Std 188-110A specify a 75 bps in-band direct sequence spread spectrum modulation. The Mil-Std 188-141A Automatic Link Establishment (ALE) standard uses an 8-FSK signalling format. Multi-carrier modulations have been used for both Link-11, where 16 tones are employed, and AN/DVT, where 39 tones are used. In the past decade and a half, PSK serial tone modulations have become the modulation of choice for high performance HF systems, with both Mil-Std 188-110A and STANAG 4285 employing PSK modulations for user data rates up to 2400 bps. QAM modulations have been specified in STANAG 5066 Annex G and in draft Mil-Std 188-110B for data rates beyond 2400 bps.

QAM constellations are chosen over PSK constellations for higher rates because of their better signal space distance properties. The QAM constellations fill the entire space, while the PSK constellations are confined to the periphery. The result is, for a given average transmitted power, the QAM modulation offers a much better signal space distance.The advantage to the QAM constellation becomes more pronounced as the number of bits per symbol is increased. Another factor which must be considered at HF is the impact of the peak power limited amplifier. This means that it is not sufficient to consider the performance of a system solely in terms of average power. Peak power performance must also be considered. As well, when considering the design of a constellation, improving the minimum distance is not the primary design criteria. Improving the overall BER is the aim. Simulation studies have shown that a constellation with good minimum distance properties and a good Gray code often outperform other constellations with superior minimum distance properties, but poor Gray coding structures for use with convolutional codes. This is a function of the FEC code employed and if, for example, a Reed-Solomon code were used, the signal space distance would again become paramount.

5.0 Synchronization and Frequency Offset Removal

Synchronization is the process at the receiver of identifying that a transmission is present and determining its timing with sufficient accuracy to permit demodulation. Again, HF offers some unique challenges in this area.

The extreme fading experienced over HF circuits means that it is possible that a fade could encompass the entire duration of the preamble, making detection difficult or impossible, even in channels where the average signal level is sufficient to permit fairly high rate communications. Designers have tried to mitigate this in two ways. STANAG 4285, for example, specifies an 80 symbol preamble which is reinserted every 256 symbols. This ensures that when the signal level rises to levels which will support communications, it can be detected and synchronized to. The disadvantage with this approach is that the ratio of data to known symbols is decreased, with the result that for the same data rate, the strength of the FEC code which can be used is decreased. The alternative to this is to use an initial long preamble to ensure synchronization, and then only include known symbols where they are directly required to assist in demodulation. Long preambles, with durations of up to 4.8 s have been used. It is very unlikely that a channel which would support communications at a rate of 75 bps or more would have a fade of that length with sufficient depth to preclude detection of the signal. This is the approach which was taken in Mil-Std 188-110A where the length of the preamble has been tied to the interleaver used; when short or no interleaving is selected, a 0.6 s preamble is sent while when long interleaving is specified, a 4.8 s preamble is used.

The degree of synchronization required depends upon the algorithm used to demodulate the data. Early techniques required synchronization which was accurate to the symbol. More modern algorithms operate effectively with synchronization which is accurate to within several symbols. This distinction can be critical at HF, where multi-path fading can result in a continually changing synchronization point.

The other use for the synchronization preamble is frequency offset removal. The known symbols in the preamble are used to estimate and remove any frequency offset in the received signal.

From a waveform design perspective, the trade-offs which must be considered are the delay and reduced data rate resulting from adding symbols dedicated to synchronization versus the probability of missing a signal which could have been successfully demodulated if an insufficient number of symbols is used for synchronization.

6.0 Equalization

In the absence of channel induced distortions, band-limited signals can be received with no intersymbol interference if the Nyquist criteria is met. The raised cosine response is a typical example of the kind of filtering that is used eliminate intersymbol interference by utilizing pulse shapes which have zero crossings at T spaced intervals.

Equalization provides compensation for channel induced distortions. At HF, the equalization must be adaptive in nature, changing as the channel itself changes with time. Equalization is required for serial tone modulations where the symbol duration, typically 0.4167 ms, is small relative to the expected time dispersion, which is often as severe as several milliseconds. Multi-carrier and M-FSK modulations, on the other hand, do not, as a rule, require equalization since their symbol spacing, typically between 8 and 13 ms, is sufficiently large as to mitigate the effect of multipath delay spread for most channels.

The simplest form of equalizer is based on a linear tapped-delay-line with the coefficients adapted directly based on some error criteria, usually a minimum mean squared error norm. An improvement on this is the decision feedback equalizer. This is composed of a feed-forward filter and a feed-back filter which operates on previous decisions. Again, the coefficients are adapted directly, based on the estimated error signal. However, in this case the characteristics of the data signal are considered. Further gains can be obtained by separating the equalization problem into two distinct tasks. The first is to identify the channel impulse response, the second is to compute the optimal weights for the decision feedback equalizer based on the estimated channel impulse response. Alternatively, when the channel impulse response is known or can be estimated a block equalization process can be used.

Most modern HF modems use equalization which requires estimation of the channel impulse response. As a consequence, the waveform designer must provide sufficient opportunity to make channel estimations and to maintain and update them as required. The main purpose of the known data segments in the framing structure of HF serial tone waveforms is to provide for isolated blocks of unknown data for block oriented detection and to facilitate the maintenance of an accurate channel estimate for both block and conventional equalizers. If the span of the known data segments is twice the expected delay spread, channel estimates can be made directly using only the known data. Otherwise, an LMS update procedure based on known data and past decisions made on unknown data is used to track and adjust the channel estimate.

7.0 Error Correction Coding

A critical feature of any modern HF data communication system is the forward error correction coding. There are a number of criteria which must be considered in selecting an FEC code. Performance, complexity, compatibility and proprietary rights issues are all significant factors in the choice of a code.

The relative performance of various coding schemes varies with code rate, modulation and the acceptable error thresholds.

7.1 Convolutional codes

Convolutional codes are soft decision, bit-error correcting codes which are usually decoded with a near maximum likelihood detection process known as Viterbi decoding. They are asymptotically optimal in an additive white Gaussian noise environment and, when combined with adequate interleaving, provide good performance in the fading channels found at HF. Convolutional codes perform poorly in burst error environments, which makes it critical to achieve sufficient interleaving to break up fades.

The rate 1/2, constraint length 7 convolutional code is commonly used for HF serial tone data transmission. Both Mil-Std 188-110A and STANAG 4285 Annex E call for this code. When rates greater than 1/2 are required, they can be achieved by puncturing the code, an exercise where selected output bits are not transmitted and a puncturing mask is applied in the decoding process. Rates lower than 1/2 are achieved by repeating the bits output by the encoder. Repeating output bits, to achieve a rate 1/4 code, for example, incurs only a small performance loss relative to what could have been obtained with a true rate 1/4 code. However, the advantage of the repetition strategy is that it is much simpler than developing alternate codecs for each data rate.

7.2 Reed-Solomon codes

Reed-Solomon codes are a class of symbol error correcting codes which provide good burst error performance, particularly when erasures are used. Reed Solomon codes with large symbol sizes, (typically 6 bits or more) usually eliminate the need for a cyclic redundancy check (CRC) for validating data fidelity. The code itself provides an indication of error when it is not possible to correctly decode the received data. This feature can be very valuable in packet data systems. Relative to other coding schemes, RS codes work best at high rates or when the acceptable BER thresholds are particularly stringent.

The major disadvantages associated with RS codes are their relatively poor performance in AWGN and in the difficulty in incorporating soft decision information into the decoding process in a form more sophisticated than simple erasures.

Reed-Solomon codes are used in STANAG 4444.

7.3 Concatenated codes

Concatenated codes attempt to use multiple encodings to overcome the shortcomings of some codes. Powerful concatenated codes have be formed by using convolutional inner codes with RS outer codes. These codes are a good match for one another when used in this way. Convolutional codes are sensitive to burst errors and, when they fail to decode properly, often produce extended error bursts. RS codes, on the other hand, work well with burst errors, so the bursty errors produced by the inner convolutional code can often be corrected by the outer RS codec.

The main disadvantage of concatenated codes is that they require two interleavers to be effective. This limits the amount of interleaving which can be applied to the inner code, with the result that for the error rates usually considered adequate at HF, i.e., in the 10^{-3} to 10^{-5} range, concatenated codes generally do not perform as well as convolutional codes by themselves. However, if a very stringent BER criteria is required, they will perform very well.

7.4 Trellis Coded Modulation (TCM)

This class of codes exploits the improved minimum distance properties which can be obtained by combining coding and modulation. TCM has proven very effective in wireline modems, where it is used extensively. The large gains seen in the wireline environment have not translated into comparable performance in fading channels, although research is ongoing to improve the performance of TCM in fading channel environments. HF in particular represents a difficult environment for TCM because of the interaction which takes place between the coding and the equalizer.

7.5 Iterative codes

Turbo codes are the best known example of iterative codes. These codes are composed of two or more component codes and iterate back and forth between the two codes to achieve performance which is very close to the Shannon capacity bound. These codes achieve results which were believed to be unobtainable with reasonable complexity until very recently. There are three drawbacks associated with these codes. To obtain the impressive performance that they offer, substantial interleaving is required. The computational complexity associated with these codes is substantially greater than convolutional codes, although significant strides have recently been made in reducing the computational complexity. Most, if not all, of these codes are protected by patents and, as such, are subject to proprietary rights which makes their adoption for use at HF very problematic.

7.6 Interleaving

Depending on the kind of code employed, the interleaver may interleave symbols or bits. In the case of Reed-Solomon codes, in order to preserve the burst error capabilities of the code, Reed-Solomon symbols are interleaved. With convolutional and other bit error correcting codes, it is the bits which are interleaved.

Both block and convolutional interleavers are used for HF data communications. The block interleaver has the advantage that if the data packets are sized to fit within an interleaver block, no flush is required. The drawback to the block interleaver is that it is only possible to synchronize at interleaver block boundaries. With a convolutional interleaver, on the hand, synchronization is possible once every cycle through the interleaver and, for the same end-to-end delay, better performance is achieved. The major disadvantage to the convolutional interleaver is that it requires a flush to clear out the interleaver at the end of the transmission.

8.0 Predicting Performance

The performance of serial tone modems as a function of delay spread is such that there is little variation or dependence of BER on delay spread up to a certain critical value. For delay spread values exceeding this value, failure is usually catastrophic. The span of this delay spread operating window is dictated by the span of the known training segments, (also called the probe or mini-probe). Within the delay spread region where the modem is functioning correctly, it is possible to develop empirical relationships between the uncoded BER as a function of SNR and Doppler spread. The example channels that are presented here are two path Rayleigh fading channels. We emphasize that the results are uncoded BERs and a further level of refinement would be necessary to make this applicable to the output of a modem modem which includes FEC.

In the high signal to noise ratio case, with delay spreads in the operating range, it has been found that the following simple relationship results in a very good fit to observed simulation results for the probability of uncoded bit error achieved with an advanced block equalizer.

$$P_b = k_m \left(\frac{\Phi_D}{R_{cs}}\right)^4 \qquad (1)$$

where k_m is a modulation dependent constant, Φ_D is Doppler spread in Hz, and R_{cs} is the channel sampling rate in Hz. Thus, for high SNR, the B.E.R. is proportional to the ratio of the channel sampling rate divided by the Doppler spread, raised to the fourth power, where the channel sampling rate is simply the reciprocal of the time between successive probe sequences. Empirically, the proportionality constant k_m has been determined for BPSK, QPSK, 8PSK and 16 QAM. The values obtained are:

$$\begin{aligned} k_{bpsk} &\sim 0.2 \\ k_{qpsk} &\sim 0.6 \\ k_{8psk} &\sim 3 \\ k_{16qam} &\sim 10 \end{aligned} \qquad (2)$$

An example of the fit obtained using these parameters, in a two path Rayleigh fading channel with 1 ms of delay spread, for a waveform using a repeated framing structure of 19 data symbols followed by 19 training symbols (denoted as 19-19), with BPSK, QPSK, 8PSK and 16QAM modulations is shown in Figure 1. Further analysis of the simulation results would allow refinement in the accuracy of these numbers.

It is possible to use this result to derive some design equations which would be useful in the development and specification of waveforms expected to use advanced block equalization techniques. The bit rate throughput of the waveform, R_b, can be expressed in terms of the symbol rate, R_s, the channel sampling rate, the number of training symbols per probe segment, T, and the number of bits per symbol for a given modulation, n_m.

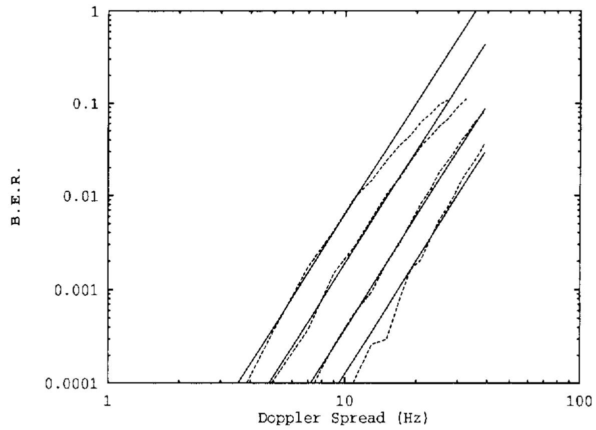


Figure 1 - Fourth power relationship of B.E.R. and Doppler spread for BPSK, QPSK, 8PSK and 16QAM modulations. Simulation (dashed), empirical relationship (solid).

$$R_b = (R_s - R_{cs}T)n_m \qquad (3)$$

Substituting (1) into (3) results in

$$R_b = \left(R_s - \left(\frac{k_m}{P_b}\right)^{1/4} \Phi_D T\right) n_m \qquad (4)$$

which provides the highest bit rate achievable for specified symbol rate, modulation, bit error rate, Doppler spread and probe sequence length. Note that because the probe sequence length determines the maximum delay spread handling capability, this equation provides a very good analytical tool for determining what sort of throughput goals are achievable for a specified environment. Simple algebraic manipulations of this equation lead to formulations in terms of bit error probability,

$$P_b = k_m \left(\frac{\Phi_D T}{R_s - \dfrac{R_b}{n_m}}\right)^4 \qquad (5)$$

and Doppler spread

$$\Phi_D = \left(\frac{R_s - \dfrac{R_b}{n_m}}{T}\right)\left(\frac{P_b}{k_m}\right)^{1/4} \qquad (6)$$

The previous relationships are valid only at high SNR. It would be desirable to establish similar relationships which are valid over a large range of SNR's. Such a relationship would take the form of

$$P_b = P_{b0}(N_0) + f(\Phi_D, N_0) + k_m\left(\frac{\Phi_D}{R_{cs}}\right)^4 \qquad (7)$$

where, in addition to the fourth power term above, there is a constant term dependent upon SNR (and probably other factors such as length of data and training segments), as well as a term which might be dependent upon both Doppler spread and SNR. Here $P_{b0}(N_0)$ is the bit error rate for 0 Hz Doppler spread and is a function of the SNR. For convenience, this dependence will not be shown in the following, but is implicit. Empirical study has shown that a very good approximation to the bit error rate is provided by the following:

$$P_b = \left(\sqrt{P_{b0}} + \sqrt{k_m}\left(\frac{\Phi_D}{R_{cs}}\right)^2\right)^2 \qquad (8)$$

A typical example of the accuracy of this fit is shown in Figure 2 below. The solid lines are the empirical relationship of (8), while the dashed lines represent simulation results. Each bit error rate point in this figure represents the simulation of 5000 modem frames, each containing 3 signalling blocks of 27 data and 27 training (27-27). For the 27 dB curve, this still results in a large degree of uncertainty in the B.E.R. for low Doppler spread values. Thus, an additional run of 50000 modem frames were run at 0 Hz Doppler spread and the P_{b0} values obtained from that run are used computing the curves shown in the figure.

The agreement between the empirically derived relationship and the simulation results in the above figure is quite remarkable. Figure 3 shows the fit of the relationship to the results obtained for the 27-27 waveform with the various modulations at Eb/No = 27dB. Again, agreement is quite good. Figure 4 shows the fit to a waveform with 75% efficiency (81 data, 27 training).

It should be noted that the agreement with the relationships of (5) and (8) is not quite as good over the ensemble of results for 8PSK and 16QAM as it is for BPSK and QPSK. As well, waveforms with very high efficiencies, particularly for the high delay

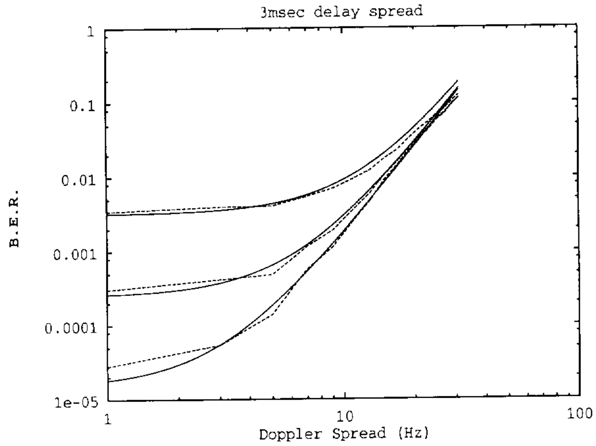


Figure 2 - Fit of empirical B.E.R. vs. Doppler spread relationship (solid) to simulated results (dashed) for qpsk signalling using a 50% efficient waveform (27-27) with Eb/No of 15, 21, and 27 dB.

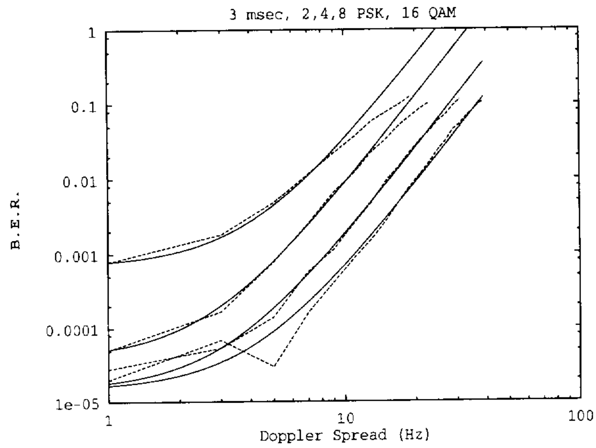


Figure 3 - Fit of empirical B.E.R. vs. Doppler spread relationship (solid) to simulated results (dashed) for various modulation techniques using a 50% efficient waveform (27-27) with Eb/No of 27 dB.

spread cases, show greater departures from (5) and (8) than do those with efficiencies near 50%. In any case, the agreement of simulation with (8) is sufficiently good that a set of design equations is proposed using (8) as a basis:

$$P_b = \left(\sqrt{P_{b0}} + \sqrt{k_m}\left(\frac{\Phi_D T}{R_s - \frac{R_b}{n_m}}\right)^2\right)^2 \qquad (9)$$

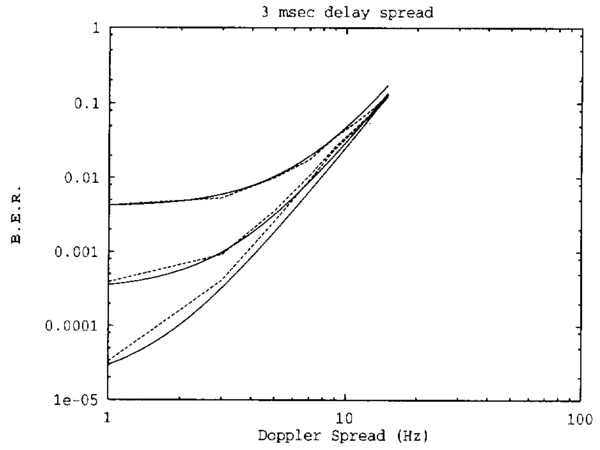


Figure 4 - Fit of empirical B.E.R. vs. Doppler spread relationship (solid) to simulated results (dashed) for qpsk signalling using a 75% efficient waveform (81-27) with Eb/No of 15, 21, and 27 dB.

$$R_b = n_m\left(R_s - \left(\sqrt{\frac{\sqrt{k_m}}{\sqrt{P_b} - \sqrt{P_{b0}}}}\right)\Phi_D T\right) \qquad (10)$$

$$\Phi_D = \left(\frac{R_s - \dfrac{R_b}{n_m}}{T}\right)\sqrt{\frac{\sqrt{P_b} - \sqrt{P_{b0}}}{\sqrt{k_m}}} \qquad (11)$$

These equations have proven to be a useful tool in the initial waveform design stage.

9.0 Waveform Design Example

STANAG 5066 is an HF subnet protocol developed by NATO in support of the BRASS program. During the development of the protocol, a requirement for waveforms providing data rates beyond 2400 bps was identified and the Communications Research Centre (CRC), who had published results of diversity combining trials with 9600 bps waveforms [3], were asked to submit a draft waveform to satisfy the requirement.

In this case, increased data rate was the clear driving factor in the design. This limited the choice of modulation to the bandwidth efficient modulations. The serial-tone approach was taken because of its better peak-to-average performance relative to the multi-carrier approaches.

The user data rate can be expressed as

$$r_d = r_b \eta_{wf} n_b r_c \qquad (12)$$

where r_d is the data rate, r_b is the baud rate or symbol rate, η_{wf} is the waveform efficiency or ratio of data symbols to total symbols sent, n_b is the number of bits per symbol, and r_c is the code rate. Increased data rates can be achieved with increased baud rate, improved waveform efficiency, increased modulation complexity or reduced code rate.

Some initial experimentation was carried out on the effects of increasing the baud rate above 2400 symbols per second. The result was that for typical commercial HF radios with 2.7-2.8 kHz passbands, the increased coding which was available improved BERs for the AWGN channel, but no improvement was seen for a CCIR poor channel where there was 2 ms of multipath delay spread. Performance degraded markedly as delay spread was increased when the signal was passed through real radios. This result has since been confirmed by Nieto, who provides strong evidence of the dangers of comparing modem implementations without adequately factoring in radio effects [4].

Improving the waveform efficiency is a relatively straightforward exercise. Neither STANAG 4285 or Mil-Std 188-110A, the key HF waveform standards, were designed with these high data rates in mind. As a consequence, STANAG 4285 has a waveform efficiency of 50% while the Mil-Std does slightly better with a waveform efficiency of 66.7% for the 2400 bps mode. Combinations of waveform efficiency, number of bits per symbol and standard coding rates which achieved user data rate of 9600 bps were considered.

A decision was made to maintain the waveform structure regardless of the selected user data rate and to achieve the various rates by changing the modulation. This also allowed for auto-baud to be implemented as a modulation recognition. As much as possible, compatibility with STANAG 4285 was to be maintained for ease of implementation.

The specification of the known data is the first major task. The preamble must be long enough so that it extends significantly beyond any radio induced transients and is designed to have good correlation properties to aid in signal detection and acquisition. At the end of the preamble, a sequence optimized for channel estimation was appended to ensure that a good channel estimate was available to the receiver at the beginning of the data.

The required delay spread capability is what drives the length of the training segment. With conventional implementations where a channel estimate is made on the preamble and only updated, not re-estimated, on subsequent known data segments, the length of the known data segments is just slightly more than the expected delay spread. The higher modulation complexities contemplated in the Annex G waveform require better channel estimates than previous implementations. For the Annex G waveform, where it was envisioned that the user would make a direct estimation of the channel estimate with each known data segment, the training segment duration needed to be more than twice the expected delay spread. The incentive for making the training segment short is that lengthening the training segment forces the data segment to be lengthened by a proportionate amount in order to maintain the waveform efficiency. The bigger the data segments get, the more susceptible the implementation is to Doppler spreading and frequency offset errors. On the other hand, longer known data segments allow for better channel estimates.

In the end, after simulation of several candidates, the frame structure shown in Figure 5, along with STANAG 4285 and Mil-Std 188-110A shown for comparison, was adopted. Another feature of the Annex G waveform design was the inclusion of a regularly reinserted preamble which could be used for late acquisition or reacquisition.

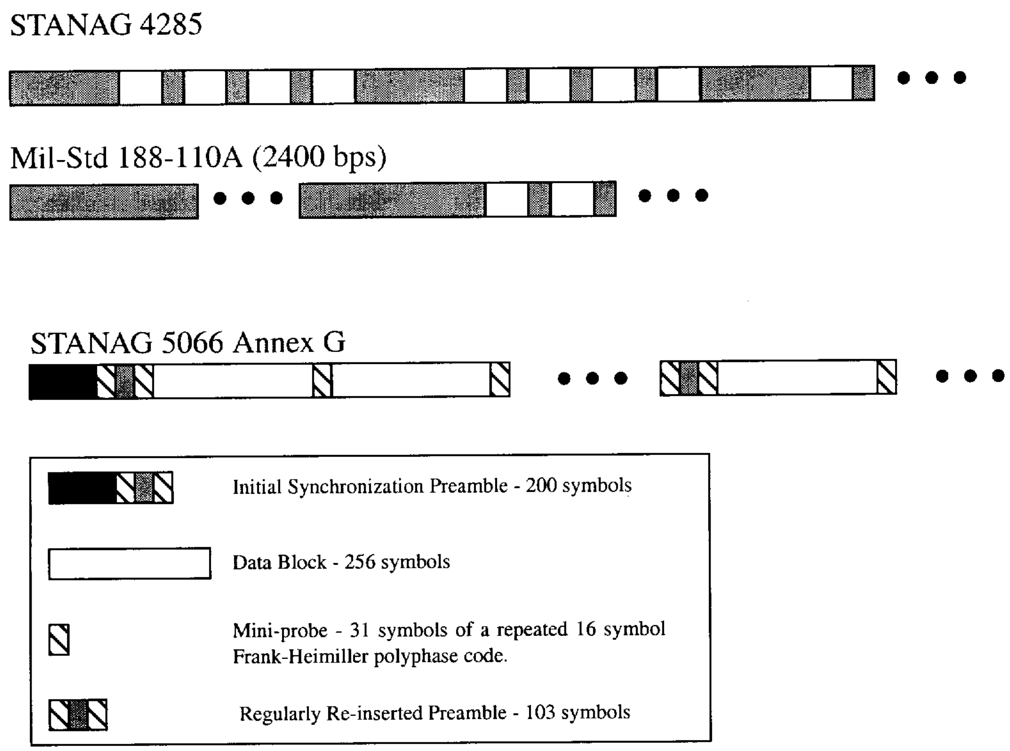


Figure 5 - Waveform structure.

A constellation optimization was undertaken for the QAM modulations used in the standard. The resultant constellations are shown in Figures 6 and 7. These clearly improve signal space distance properties relative to the conventional square constellations, but what is not so readily apparent is that they also preserve the excellent Gray-coding properties of the square constellations.

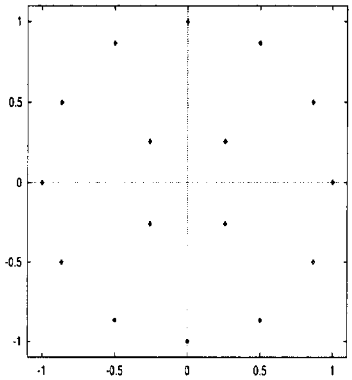


Figure 6 - 16QAM constellation used to achieve 6400 bps.

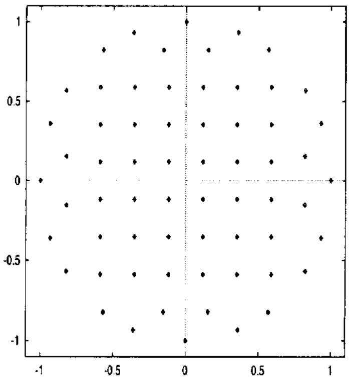


Figure 7 - 64QAM constellation used to achieve 9600 bps.

The FEC was based on the rate 1/2 constraint length 7 convolutional code specified in both STANAG 4285 and Mil-Std 188-110A, although in this case it is punctured to rate 3/4. A convolutional interleaver, much like the one in STANAG 4285, but using a nominal 48 row structure, was specified. This allows the interleaver to synchronize immediately after any preamble and every data block thereafter.

A summary of the Annex G waveform, with the STANAG 4285 2400 bps Annex E coding option shown for reference is provided in tabular form below.

<table>
<tr><th>Waveform (data-training)</th><th>Modulation</th><th>Code Rate</th><th>User Data Rate (bps)</th></tr>
<tr><td rowspan="4">256-31</td><td>64-QAM</td><td>3/4</td><td>9600</td></tr>
<tr><td>16-QAM</td><td>3/4</td><td>6400</td></tr>
<tr><td>8-PSK</td><td>3/4</td><td>4800</td></tr>
<tr><td>QPSK</td><td>3/4</td><td>3200</td></tr>
</table>

For comparison, STANAG 4285 2400 bps mode:

Waveform (data-training)	Modulation	Code Rate	User Data Rate (bps)
32-16	8-PSK	2/3	2400

In comparing the two waveforms, it should be noted that they are designed for different purposes and services. The Annex G 4800 bps mode requires just slightly more SNR than the STANAG 4285 waveform in a benign channel. This is not surprising as both use the same modulation, 8PSK, and the code rates are not all that different, rate 3/4 vs. rate 2/3. On the other hand, the Annex G waveform will fall apart completely in the presence of large Doppler spreads, while the STANAG 4285 waveform can cope with substantial Doppler spreading.

10.0 Summary

This paper has provided a tutorial overview of the design of HF serial tone waveforms.

11.0 Acknowledgments

This work was partially funded by the Department of National Defence, Canada.

12.0 References

[1] T.J. Willink et. al., "Robust HF data communications at high latitudes", submitted to *IEE Proceedings - Microwaves, Antennas and Propagation*, 1999.

[2] M.J. Angling et. al., "Measurements of Doppler and multipath spread on oblique high latitude HF paths and their use in characterising data modem performance", *Radio Sci.*, Jan-Feb, 33(1), 1998.

[3] M.B. Jorgenson et. al., "Polarization diversity for HF data transmission", *7th International Conference on HF Radio Systems and Techniques, Nottingham*, UK, July 1997.

[4] John Nieto, "An implementation of STANAG 5066 Annex G Waveforms", *Nordic Shortwave Conference*, Faro, Sweden, August 1998.

Beyond 9600 bps at HF

M.B. Jorgenson, Robert Johnson and K.W. Moreland
Communications Research Centre
3701 Carling Ave.
Ottawa, Canada, K2H 8S2

Summary

This paper focusses on the emergence of high data rate HF data communications capabilities providing data rates of 9600 bps and beyond. The use of independent sideband equipment and allocations is suggested as a feasible means of providing data rates of up to 32 kbps, particularly in a naval environment where surface wave propagation extends well beyond line of sight. Results obtained with a modem implementation and HF channel simulator are presented to substantiate claims that the performance is acceptable for many naval scenarios.

1.0 Introduction

In this paper we look at the emergence of high data rate HF data communications and consider the potential for future increases in data rate beyond 9600 bps.

It is only in recent times that achieving user data rates in excess of 2400 bps over HF channels has been seriously considered. Indeed, the era of the 75 bps uncoded FSK radio-teletype signal is only now drawing to a close with the modernization of NATO naval broadcasts. At the same time, military forces of all kinds are coping with the digitization of the battlefield. Networking services, real-time video feeds, high resolution imagery and other applications and services are creating demands for ever higher data rates. This evolution is requiring data rates which have traditionally been unsupportable at HF and, consequently, may lead to the marginalization of HF services unless means can be found to increase the rates attainable over HF circuits.

The authors' development of modems and signalling strategies capable of transmitting 9600 bps using standard 3 kHz bandwidth HF channels culminated in a demonstration of this capability over a 1800 km skywave link [1]. The results were sufficiently encouraging that the signalling strategy was incorporated in draft NATO STANAG 5066 (v1.02) as a non-mandatory Annex (G). This standard was adopted with minor modifications to provide the high data rate component of draft Mil-Std 188-110B [2].

It is unlikely in the extreme that data rates in the neighbourhood of 64 kbps, which are quickly becoming minimum standards for the digital battlefield, can be attained over 3 kHz HF channels. In order to obtain rates of this sort, it will be necessary to revise frequency allocations to permit wider bandwidth HF channels [3]. This is likely to be a time-consuming process, however, and it would seem prudent to determine what can be accomplished within the present regulatory framework.

In previous work on high data rate transmission over 3 kHz HF channels, rates of up to 16 kbps have been demonstrated [4]. In addition to the standard 3 kHz channels, there are currently some independent sideband (ISB) channel assignments available with (effectively) 6 kHz bandwidth. These are currently allocated to Link-11 circuits, where suitable radios can make use of both sidebands to produce a diversity transmission of the tactical data link signal. These ISB assignments are particularly interesting when considering what data rates might be possible with current frequency assignments and radio equipments.

In the following section, we will consider the kinds of HF channels which could be expected to support high data rates with conventional SSB and ISB transmitters. We will then examine the signal flow model of a high data rate system to identify which aspects may be problematic for high data rate transmissions. This is followed by results obtained in real-time simulation tests. An Annex G implementation (up to 9600 bps) is used to provide a baseline for comparison.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

2.0 The HF Channel

The HF channel offers particularly intriguing challenges. We will briefly discuss three HF channels of interest: the skywave channel, the surface wave channel, and the Rician channel, composed of a non-fading surface wave component and a fading skywave component.

The skywave channel provides support for long-haul beyond line of sight (BLOS) communications without the need for relays. However, the fading which is experienced on these circuits can be severe, and only a small percentage of skywave channels will support communications at data rates of up to 9600 bps. In our view, 9600 bps, which requires a 64QAM constellation, is the practical limit for communications over 3 kHz HF skywave channels. A much larger percentage will support a 6400 bps waveform using a 16QAM constellation.

HF surface wave propagation provides beyond line of sight communications as the wave propagates over the curved surface of the earth. How far beyond line of sight depends strongly on the composition of the surface. The HF surface wave propagates for distances of several hundred kilometers over highly conductive surfaces such as seawater, much farther than over fresh water or land. This phenomenon makes HF surface wave communications particularly significant for naval applications.

The noise environment for all HF channels is somewhat more impulsive than the additive white gaussian noise (AWGN) that is commonly used to test HF modems. CCIR Rec 322 provides a recommended noise generator for use in simulating HF channels [5]. Interestingly, for many modern HF data communications systems, CCIR 322 noise is actually easier to cope with than AWGN. This result is a consequence of the coding scheme employed. Although the CCIR 322 noise is more likely to generate large error events than AWGN of the same power, the error events tend to be more spread out, and hence easier for the FEC coding to cope with, as long as interleaving is employed.

2.1 Naval Application of HF Surface Wave Communications

In general, the relative strengths of groundwave and skywave signals as a function of range behave as illustrated in Figure 1. At short range, groundwave will dominate and at long range, skywave will be the primary propagation mechanism (Rayleigh channel). At intermediate ranges, there is a transition region where the Rician channel model is appropriate. An assessment was undertaken to obtain an indication of:

- the received signal levels (RSL's) and signal-to-noise ratios (SNR's) that could be expected by naval platforms at various ranges, and;

- the boundaries of the transition region between groundwave and skywave domination.

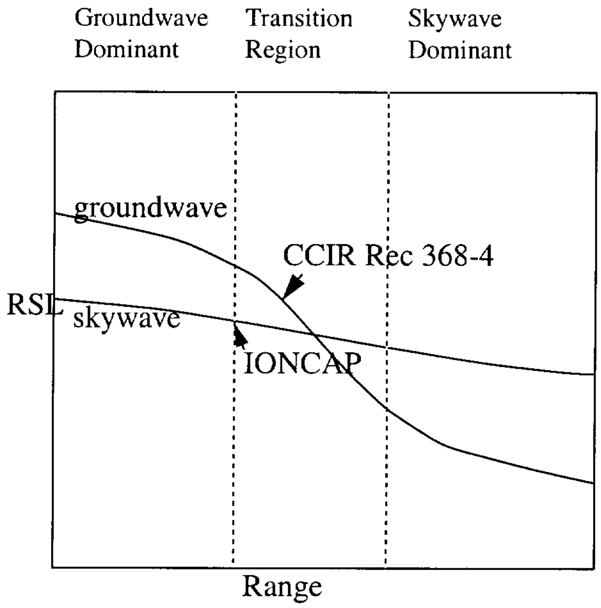


Figure 1 - Channel model determination based on comparing groundwave and skywave received signal levels as a function of range.

The IONCAP [6] propagation model was used to define HF skywave conditions for ship-to-ship communications in the North Atlantic. One ship was located at (50 deg N, 30 deg W) and the other was positioned due north of it at the range being tested. The co-ordinates are important because geographic location influences the level of atmospheric noise and solar effects. A sunspot number of 100 was used, as this represents an average value.

The antenna considered was a wideband monopole, of relatively short length ($l \ll \lambda/4$). These antennas have essentially a cosine-squared elevation angle power gain pattern, which discriminates against high angle skywave signals. A 500 W transmitter was considered.

IONCAP noise levels were used after making the adjustment for 3 kHz bandwidth. IONCAP uses the CCIR Rep 258-4 environmental (man-made) noise model with the addition of natural (atmospheric) noise. For this investigation, the "Rural" environment model was selected. This has a power of -150 dBw in a 1 Hz bandwidth (-115.2 dBw in 3 kHz) at 3 MHz, with a decrease with frequency of 27.7 dB per decade. It must be noted that the IONCAP noise levels represent median values.

	Where Groundwave Domination Ends		Crossover Point Occurs At		Skywave Domination Starts At	
Freq. (MHz)	Range (km)	~ SNR (dB)	Range (km)	~ SNR (dB)	Range (km)	~ SNR (dB)
2	380	42.6	600	39.0	>600	?
4	400	37.5	500	40.5	600	30.5
7.5	385	32.0	460	36.5	550	26.5
10	395	24.7	450	31.0	520	20.0
20	295	11.4	350	0.0	>600	N/A

TABLE 1. Range and signal levels of skywave and groundwave signals for various frequencies.

In order to impart some idea of the noise model impact, selecting the worst case environmental category in the CCIR noise model (industrial) results in ~10 dB increase in noise level.

The anticipated signal levels are shown in Table 1. The extent of surface wave propagation implies that HF signals requiring SNR's in the 30 dB range could be used to support operations within a naval battlegroup, which might be spread over several hundred kilometers of ocean. The alternatives to HF in this scenario involve aerial relay platforms or satellite communications. Both of these alternatives are costly and may be impractical for many NATO navies.

3.0 System Model

The system used to characterize the performance of the high data rate waveforms considered in this paper is shown in Figure 2. Digital data is presented to the modem from a data source. It is then encoded, interleaved and modulated by the modem and presented to the transmitter (exciter and power amplifier) as an audio signal centered at 1800 Hz. The exciter automatic level control (ALC) circuit operates to maintain the transmitted signal within a specified power range and the signal is upconverted and further filter by the exciter's transmit filters. The signal is amplified by an HF power amplifier and applied to an antenna.

The channel models considered in this study are: the simple AWGN channel; a Rician channel where there is a single non-fading path and second fading path with 6 dB lower average power; and the CCIR poor channel which has two independently fading paths, each with a Doppler spread of 1 Hz, separated by 2 ms.

After transmission through the channel, the signal is filtered and down-converted by the receiver. An automatic gain control (AGC) circuit operates to maintain the received signal level within an acceptable range.

The audio signal from the receiver is passed to the receive modem. Initially, a modem synchronization function searches for the preamble which, when detected, is used to remove any frequency offset between transmitter and receiver and to establish the timing for the frame structure. The last segment of the preamble is used in the calculation of a channel estimate for use in the equalization process. Soft decisions out of the equalizer are fed to the de-interleaver and then to the decoder. Finally, data out of the modem is passed to an external device where, in this case, it is compared to the data which was transmitted and errors are counted.

When an ISB configuration is selected, the data stream splits into two paths following the interleaver. From that point forward, there are effectively two parallel signal paths, one for the upper sideband and another for the lower sideband. Bits out of the interleaver are passed alternately to each path. On the receive side, the two signal paths remain separated until the soft decisions are to be de-interleaved. The only exception to this separation of USB and LSB is that some advantage is taken of the knowledge that both are present in the synchronization algorithm employed.

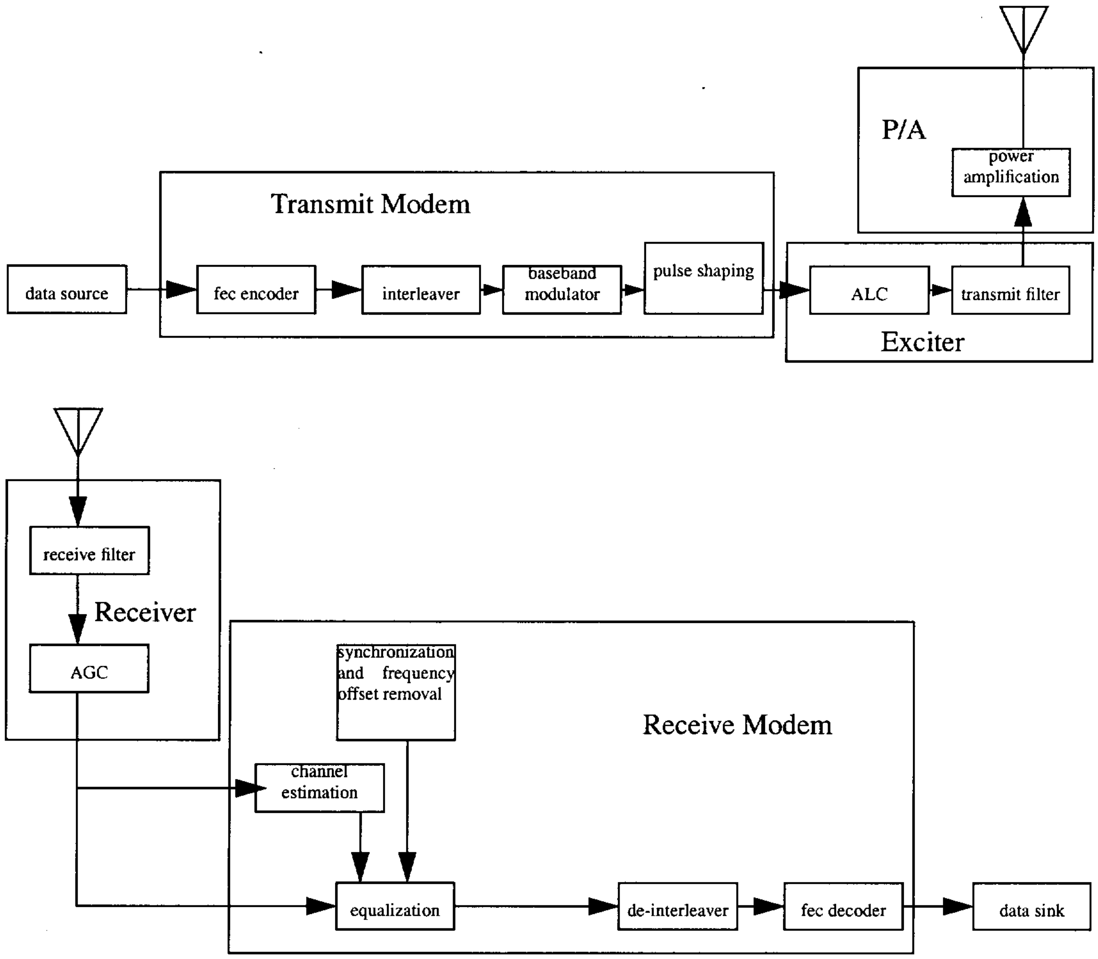


Figure 2 - Signal flow diagram for experimental apparatus.

Surprisingly it was found that the exciter, rather than the receiver, has presented the greatest difficulty with these new waveforms. In the exciter which is employed in our testing, it was very difficult to completely turn off the Automatic Level Control (ALC). The front panel setting ALC-off only partially disabled the ALC and, as the ALC was designed for voice applications, it did not tolerate 256QAM constellations well. In fact, even 64QAM constellations cause difficulty with radio ALC circuits when the ALC cannot be disabled.

Typical receiver automatic gain control implementations have not proven to be a problem. Certainly, in a skywave channel with the attendant fading, this would be such a problem. However, for the kinds of channels which these higher rate waveforms would operate in, namely surface wave and Rician channels where the dominant path is steady, the AGC attacks at the beginning of the transmission and remains steady thereafter. In the radios we employed, no difficulty was encountered with the radio responding to the signal level variability due to the QAM constellations themselves.

The AGC and ALC circuits of legacy HF radios are designed for voice applications. They attack quickly and, particularly in the case of receiver AGC, have selectable decay times. While this is good for voice and acceptable for near constant envelope data modulations, it creates problems with multi-level modulations (like the QAM constellations we use). With newer digital radios, the effect of AGC can be incorporated in the output signal, creating a signal which appears linear throughout the dynamic range of the radio. This will solve the ALC/AGC problems currently being experienced.

Another concern which has proven to be groundless, at least for the equipment which we have used, was that the power amplifier might not be sufficiently linear to support the 256QAM signal without some distortion. Although it is difficult to rule out entirely, primarily because of problems observed with the ALC, it does appear that the P/A is sufficiently linear to avoid distortion in the data signal.

When large data block sizes are used, as they are in all high data rate HF applications, frequency offset removal becomes a critical component of the modem. This results because of the ambiguity which is introduced if residual offsets of more than 1/(2 * data block duration) are present.

4.0 Performance Results

4.1 Baseline: STANAG 5066 Annex G

The performance of an implementation of the Annex G waveforms in an AWGN channel is shown in Figure 3. The three modulation schemes used for, 4800 bps to 9600 bps, are 8PSK, 16QAM and 64QAM. Note that increasing the data rate by a factor of 2 from 4800 bps to 9600 bps costs more than 7.5 dB in required SNR. With the 64QAM waveform, we still remain well within the predicted SNR values for a naval force which is spread widely over the ocean.

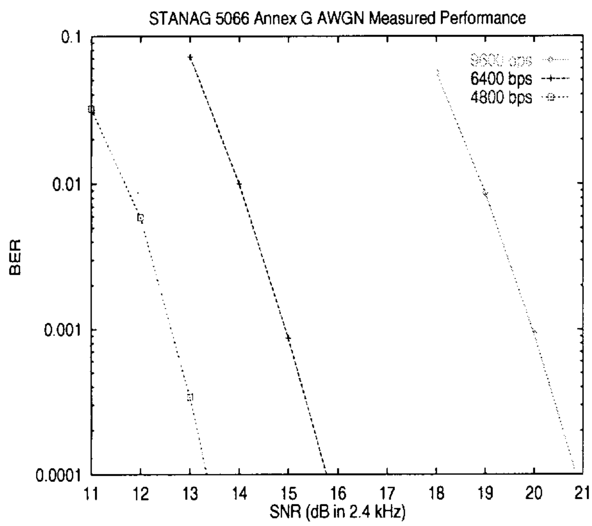


Figure 3 - Annex G waveforms in AWGN.

Waveform (data-training)	**Modulation**	**Code Rate**	**User Data Rate (bps)**
256-31	64-QAM	3/4	9600
	16-QAM	3/4	6400
	8-PSK	3/4	4800
	QPSK	3/4	3200

TABLE 2 - Annex G waveform parameters

Figure 4 shows the performance of the Annex G waveforms in a Rician channel. This channel is of particular interest when operating at the edges of the surface wave coverage region. It is also of interest when antennas which do not discriminate against high-angle radiation are used. One relevant instance of this in the naval environment would be communication to a helicopter. Performance in the Rician channel is only degraded by approximately 2 dB from the performance obtained with the AWGN channel. The use of a good equalizer is the key reason why this degradation is so small.

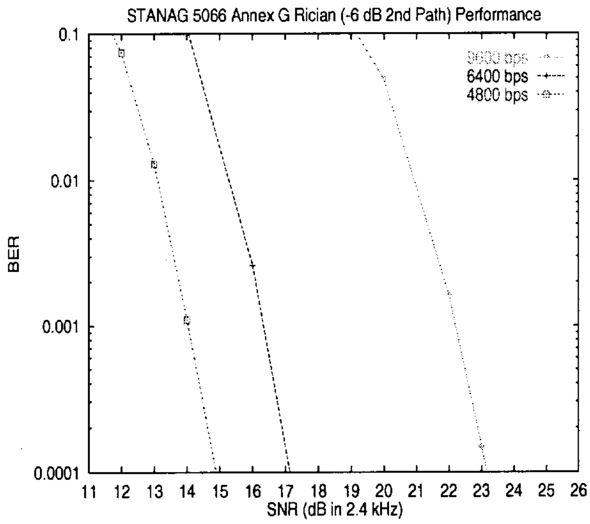


Figure 4 - Annex G waveforms in a Rician channel with second path fading, 2 ms delay spread, 1 Hz Doppler spread.

Figure 5 shows the performance of the Annex G waveforms in a CCIR Poor channel. The significant increase in the required SNR is due to the fading nature of this channel. The 64QAM, 9600 bps signal requires approximately 30 dB SNR for reception. This level of SNR is only rarely achieved over sky-wave links with the usual sorts of fitted HF equipment. As we believe that the 9600 bps waveform at 30 dB SNR represents the practical limit for sky-wave transmission, we will not consider the CCIR Poor channel further. Rather, we will limit our discussion to the surface wave channel and the Rician channels which we expect to see in the naval battle-group scenario.

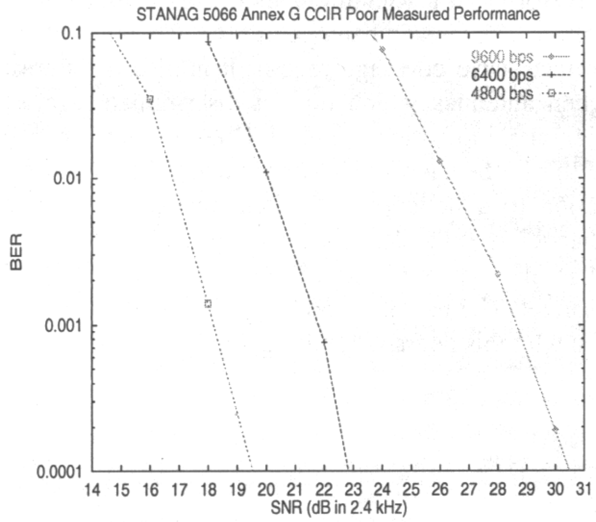


Figure 5 - Annex G waveforms in CCIR Poor channel, defined as 2 fading paths, 2 ms delay spread, 2 Hz Doppler spread.

4.2 Data Rates Beyond 9600 bps

The Annex G waveform has a waveform efficiency of 8/9; that is, on average, for every 8 symbols of on-air data which are sent, a single known symbol is added for equalizer training etc. The convolutional code used is a rate 3/4 code. At 9600 bps, the Annex G waveform uses a 64QAM modulation which transmits 6 bits for every symbol. At 2400 baud, this means that the absolute upper limit, if no training symbols and no coding was employed, would be 14.4 kbps.

We could increase the waveform efficiency, although at 8/9 there is not much room left for reducing the percentage of training symbols. In point of fact, the waveform structure of the high rate waveforms considered in this paper vary only slightly from the Annex G parameters.

Another alternative is to increase the code rate. To do this substantially with a convolutional code is difficult and incurs significant performance costs. To avoid these problems, we have used an iterated code, developed at CRC, called a Hyper-code [7]. Like turbo-codes, this class of codes works by iterating two codes against each other to obtain performance which is within a few tenths of a decibel of the Shannon capacity. Hyper-codes obtain performance essentially equivalent to turbo-codes for high rates, but with some advantages in terms of implementation complexity and block sizes. With these codes it is quite feasible to look at code rates of 0.9 or higher.

The next parameter which can be changed is the modulation. The highest modulation used in Annex G is 64QAM. We believe that many HF surface wave channels are capable of supporting a 256QAM modulation and this is the modulation which we have chosen. The 256QAM modulation used was a simple 16x16 square constellation. If past experience can be used as a guide, it should be possible to improve on the performances reported here by 2 or more dB by optimizing the constellation.

Finally, the most significant change is the use of an independent sideband radio, which offers two 3 kHz channels. In our implementation, the output of the codec is interleaved and then successive bits are passed alternately to the modulator for the upper and lower sideband. On the receive side, the sidebands are demodulated independently and then the soft decisions output from two independent equalizers, one for each sideband, are input to the de-interleaver/codec.

Waveform (data-training)	Modulation	Code Rate	User Data Rate (kbps)
330-30	256QAM	0.913	32
288-31	256QAM	0.837	28.8
250-31	64QAM	0.758	19.2
288-31	16QAM	0.837	14.4

TABLE 3 - ISB Waveform Parameters

The consequences of these design choices are evident in Figure 6 below. With an ISB radio it is possible to send 14.4 kbps in an AWGN channel, using a 16QAM constellation, with less than 16 dB SNR (measured in 2.4 kHz). Contrast this with the 21 dB SNR required to send 9.6 kbps using the 64QAM modulation specified in Annex G. The benefits of the improved coding can be seen when comparing the 19.2 kbps ISB implementation to the Annex G implementation. These would be expected to be comparable if the coding used was the same as both employ 64QAM modulations. The Hyper-code has improved the performance of the 19.2 kbps implementation by approximately 2 dB. It is worth noting that the 28.8 kbps and 32 kbps waveforms, both of which use a 256QAM constellation, are operating with acceptable error rates in SNRs below 30 dB. Essentially no difference is seen in the BER curves when various interleaving depths are used.

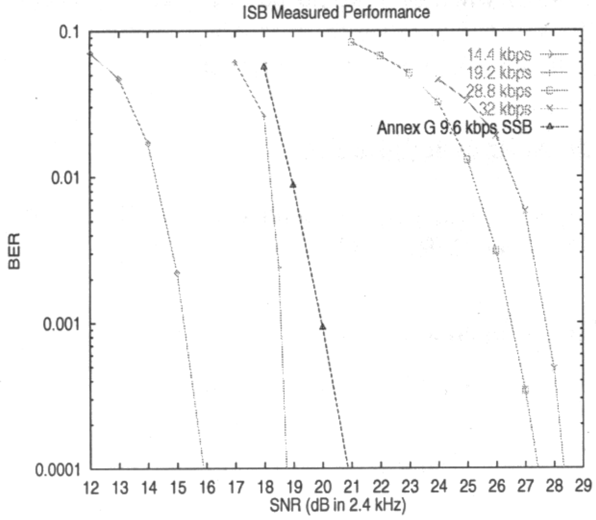


Figure 6 - ISB waveforms in AWGN channel with Annex G 9600 bps waveform shown for reference.

In the Rician channel results, shown in Figure 7, we again see a shift of approximately 2 dB, relative to the AWGN channel for the lower order QAM constellations. However, notice in Figure 7 that the 256QAM constellations have shifted by 3-4 dB, reflecting the greater sensitivity of these constellations to distortion of any kind. Again, note that even for the highest data rates, the SNR required is still in the vicinity of 30 dB. This, in conjunction with the SNR projections in section 2, suggests that these data rates should be achievable in a naval scenario between platforms spread over several hundred kilometers of ocean.

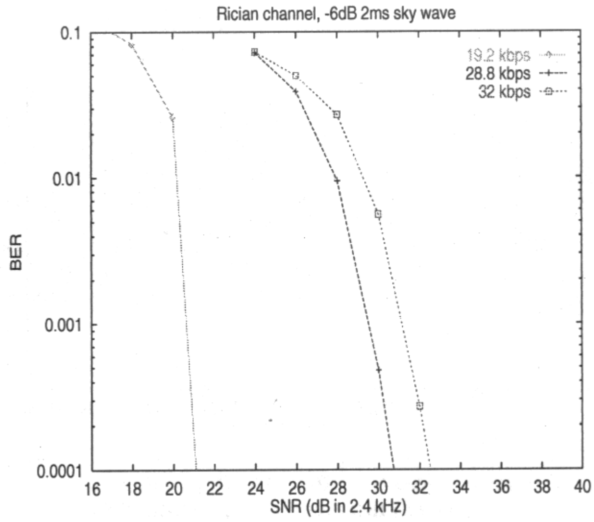


Figure 7 - ISB waveforms using long interleaver (10 s) in Rician channel.

Figure 8 shows the impact of interleaving on BER performance for the high data rate waveforms. In this case, the incremental change in data rate from 28.8 kbps to 32 kbps is seen to cost substantially in terms of the irreducible error rate when no interleaving is used. The performance of the 32 kbps waveform with short interleaving is actually comparable to that of the 28.8 kbps waveform with no interleaving. The 28.8 kbps waveform performs fairly well as long as some degree of interleaving (i.e. short ~ 1s) is used.

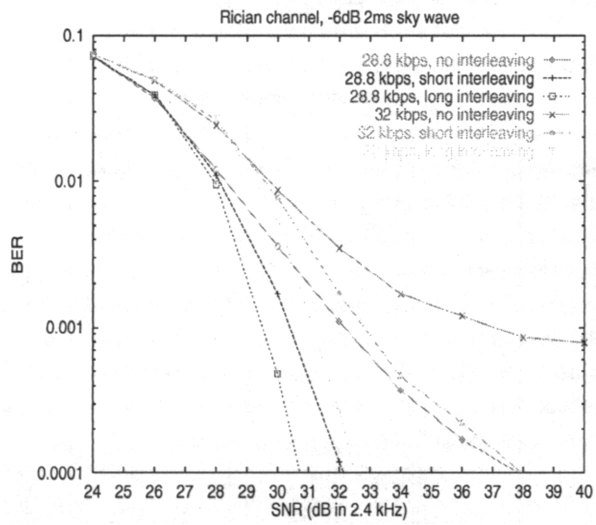


Figure 8 - ISB waveforms with 0, short (1 s) and long (10 s) interleaving in Rician channel.

5.0 Recommendations

In developing the communications architecture for future naval deployments, consideration should be given to using high data rate HF ISB to provide coverage within a battlegroup or surface action group. In a networked application, connectivity to the group would likely be via satellite to one node in the group, with distribution to the rest of the group being passed via HF. This would allow navies not having indigenous satellite capabilities to keep costs down and could even prove beneficial to those navies with satellite assets by reducing the burden placed on them. If 64 kbps is a minimum required data rate, it could be obtained by mult-channelling two or more HF ISB circuits together.

One of the most difficult aspects of using these new higher data rate modems to existing radio equipment is the unwanted modulation of the data signal by the ALC and AGC. One possible addition to HF radios which could make this much easier would be to incorporate an ALC/AGC trigger line under modem control. If this were done, the modem could trigger the ALC/AGC circuit to attack at known times with the result that the change in signal level could be anticipated and corrected internally by the modem.

Use of independent sideband (ISB) 3 kHz channel assignments is the last way available to increase bandwidth while remaining within conventional frequency assignments and using conventional radios. To go beyond this requires wider bandwidth radios and a change in frequency assignment philosophy. A quick look at the Shannon capacity curve reveals that with data rates of 9600 bps or greater in 3 kHz we are now operating well into the bandwidth efficient region of the curve. Further increases in data rate will come with exorbitant increases in transmit power levels. Alternatively, if additional bandwidth can be made available, the same transmit power levels can be used to transmit much higher data rates. As we ponder the future of HF radio in the face of increasing competition from little-LEO satellite communication systems, it behooves us to consider carefully whether it would be in the best interest of HF users to liberalize the frequency allocation rules to allow for higher data rate services.

6.0 Acknowledgments

This work was partially funded by the Department of National Defence, Canada.

7.0 References

[1] M.B. Jorgenson et. al., "Polarization diversity for HF data transmission", *Seventh International Conference on HF Radio Systems and Techniques*, Nottingham, U.K., July, 1997.

[2] Private communication, Bill Furman, Harris Corp. and Ron LeFever, General Atronic Corp., March 1999.

[3] Stephen Elvy, "High data rate communications over HF channels", *Nordic Shortwave Conference*, Faro, Sweden, August 1998.

[4] N.C. Davies et. al., "A 16,000 bits per second high data rate HF modem", *Nordic Shortwave Conference*, Faro, Sweden, August 1998.

[5] International Radio Consultative Committee (C.C.I.R.) *Report 322-3. Characteristics and Applications of Atmospheric Radio Noise Data.* I.T.U. Geneva, 1988.

[6] L.R. Teters, et. al., *NTIA Report 83-127*, US Dept. of Commerce,1983.

[7] A. Hunt, S. Crozier and D. Falconer, "Hypercodes: High performance low-complexity error correcting codes", *Proceedings of the 19th Biennial Symposium on Communications*, Kingston, Ontario, Canada, June 1998.

THE HIGH LATITUDE PERFORMANCE AND AVAILABILITY OF STANAG 4415 USING MULTIPLE FREQUENCIES

Torgeir Bergsvik and Vivianne Jodalen
Norwegian Defense Research Establishment (FFI)
P.O. Box 25, N-2027 Kjeller
bet@ffi.no vjo@ffi.no

Paul S. Cannon and Paul Arthur
DERA Malvern
Malvern, Worcs.
WR143PS, UK
pcannon@dera.gov.uk pcarthur@dera.gov.uk

SUMMARY

The high latitude HF channel has been measured and characterised in terms of Doppler spread, delay spread and signal-to-noise ratio. The performance of data modems compliant with STANAG 4285 (2400 bps mode, non-robust) and STANAG 4415 (75 bps, robust) has been determined over a comprehensive range of simulated channel conditions. A comparison is made between the channel measurements and the modem characterisations, and modem availabilities during the measured channel conditions are determined. The robust modem shows 50-70% higher availability than the non-robust modem on the measured paths. The most important factor contributing to modem failure is low signal-to-noise ratio.

The paper also addresses the number of frequencies required for a HF circuit to achieve maximum communications availability. On a 1500 km path it is shown that the STANAG 4285 modem requires 6-8 frequencies whereas the STANAG 4415 modem reaches maximum availability with 2-3 frequencies.

1 INTRODUCTION

High speed modems such as STANAG 4285 (NATO (1)) (2400 bps) are unlikely to perform well on high latitude paths that experience severe multipath, rapid fluctuations and absorption. Under these conditions the data rate has to be reduced, and robust coding and modulation must be utilized in order to reduce the error rate. The minimum signal to noise ratio (SNR), the maximum multipath spread (MS) and Doppler spread (DS) that a modem can tolerate whilst meeting a required bit error performance can be determined by measurements under simulated channel conditions (Arthur and Maundrell (2)).

The Doppler And Multipath SOunding Network (DAMSON) has, for many years, collected data on multipath spread, Doppler spread and signal to noise ratio on different paths in northern Scandinavia. In this paper we use DAMSON data and modem performance characterisations to infer the percentage availability of modems if operated on the DAMSON paths. STANAG 4285 (mode 2400 bps, long interleaver) and STANAG 4415 (NATO (3)) (75 bps, mode long interleaver) have been selected for analysis being a non-robust and a robust waveform, respectively. When the modems fail, we have tried to determine which of the channel parameters (SNR, MS or DS) causes the modem to fail.

Traditionally, HF users have been assigned a very limited number of frequencies, maybe only one day-frequency and one night-frequency. The communications availability of such a system is then often low, and sometimes there is no communications at all. By increasing the number of assigned frequencies the communications availability can increase, but not necessarily in proportion to the number of frequencies assigned. In the second part of this paper we determine the modem availability gain which can be derived from a pool of assigned frequencies. This information can be used for frequency allocation and in the automatic channel selection process (ACS) of automated HF systems where a number of frequencies are sounded before the best frequency is selected.

2 DAMSON MEASUREMENTS

The DAMSON measuring equipment is described in Davies and Cannon (4) and previous analysis is reported in Angling et al (5). This channel sounder uses pulse compression sounding within a bandwidth of 3 kHz, and after real-time processing at the receiver the scattering function of the channel is stored for later analysis. For the DAMSON data collected and used in this analysis the delay resolution of the sounding equipment was 0.6 ms, the delay range 12.5 ms, the frequency resolution 0.65 Hz and the frequency range ±40 Hz.

From received scattering functions, channel parameters such as Doppler spread, multipath spread and signal-to-noise ratio are extracted. The SNR is measured as the signal energy over all modes relative

PARTS OF MATERIAL ORIGINALLY PUBLISHED AT THE NORDIC HF CONFERENCE 1998

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

to the background noise in the 3 kHz channel. The multipath spread for this analysis is a composite delay spread, ie the delay between the leading edge of the first occurring mode at the receiver and the trailing edge of the last occurring mode. The edges are defined as the limits containing 80% of the power. In order to determine Doppler spread, a weighted sum of the modes is calculated and the 80% power limits determined.

The equipment runs automatically and measurements have been taken 24 hours a day. The measurement schedule cycles through 10 frequencies from 2.8 to 21.9 MHz in 10 minutes storing one scattering function for each frequency. If signals have not propagated, or have been disrupted by interference, a scattering function is nevertheless stored, but the instant is identified in the analysis software as containing no signals. For each frequency the maximum number of scattering functions is 144 in 24 hours.

In this work a DAMSON channel measurement is considered to be a point in three dimensional space that describes the channel conditions at the time of the measurement. The three dimensions are SNR, DS and MS.

3 PERFORMANCE CHARACTERISATION OF MODEMS

A procedure for deriving comprehensive modem performance characterisations is described in Arthur and Maundrell (2). Modems are assessed through an automated process which employs an HF simulator based on the standard Watterson model. A series of tests is performed, with each test using a specific combination of Doppler spread and multipath spread, and the SNR being adjusted until the modem bit-error rate (BER) falls within a specified range. This value of SNR is identified as the 'threshold SNR' in order to give acceptable performance of the modem with the specified channel conditions. The characterisations for this analysis employed a BER range of $5 \cdot 10^{-3}$ to $0.2 \cdot 10^{-3}$, Doppler spreads ranging from 0 to 40 Hz in steps of 1 to 4 Hz and multipath spreads ranging from 0 to 40 ms in steps of 1 to 4 ms. If the Doppler spread and multipath spread were such that the measured BER was greater than $5 \cdot 10^{-3}$ when the SNR was set to 60 dB, then in our analysis we call these channel conditions 'saturated'. The 'threshold SNR' for various Doppler and multipath conditions are characterised in three dimensional space constituting a 'performance surface' for the modem. In general low DS/MS values will require a low SNR whereas large DS/MS values will require a large SNR or will be saturated.

4 AVAILABILITY OF MODEMS ON THE DAMSON PATHS

This work combines the measured channel conditions (SNR', DS', MS') with the simulated performance surfaces of the selected modems in order to determine whether the modems would have worked during the measured channel conditions. The method used is illustrated in Figure 1. The DAMSON point (SNR', DS', MS') is referenced to the closest point on the surface (SNR, DS, MS). If SNR'>SNR, the DAMSON point is above the simulated surface and we say that the modem will work during the measured channel condition.

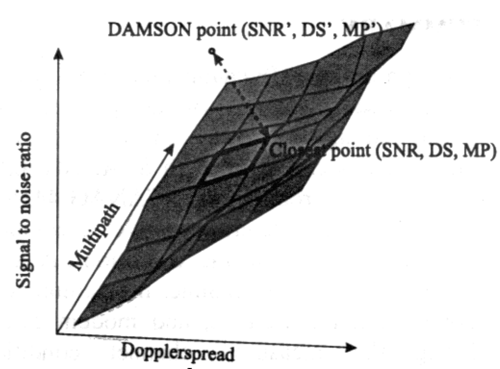


Figure 1 Method of comparison between DAMSON measurements and simulated performance of modems

We have arbitrarily selected time periods of between 7 and 40 days of DAMSON data from each season, and we have determined the percentage availability of the modem when signals have propagated. Instances of no DAMSON propagation are not counted and so the result is not an 'overall' availability. Each frequency has been compared separately.

RESULTS ON THE ISFJORD-LYCKSÄLE PATH

Availabilites have been calculated for day time and night time, different seasons, different DAMSON paths (auroral paths/sub-auroral paths), etc. We show here a few results, but more comprehensive results can be found in Bergsvik (6).

The availability of STANAG 4285 (2400 bps) during the period 3-13 Sept 1996, hours 19-01 UT on the 1500 km north-south Isfjord (78°) – Lycksäle (64°) path is shown in Figure 2. Likewise, the availability of STANAG 4415 (75 bps) is shown in Figure 3. For this period the mean SNR for frequencies above 3 MHz varies between 5 and 13 dB, the maximum mean DS on a frequency is 4.1 Hz and the maximum mean multipath spread is 1.5 ms. The standard deviation of the measured channel parameters is small. So channel conditions in this period are thus fairly benign.

On this path for the September period, we see that across the frequencies the robust low data rate communication did achieve a 60-80% higher availability than the non-robust high data rate communication. A similar trend is observed for day time conditions, except that the highest and lowest frequencies show lower availability for the robust waveform.

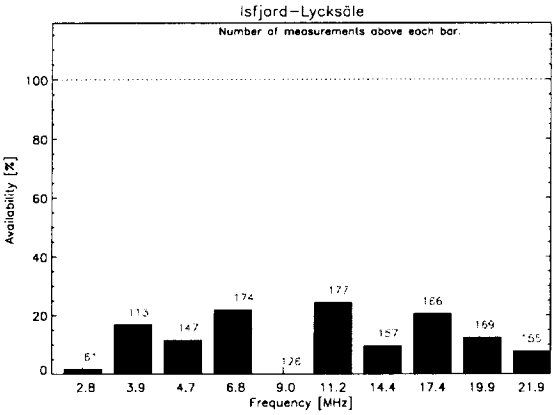


Figure 2 *Modem availability on propagating channels for STANAG 4285 (2400 bps, long interleaver) for period 3-13 Sept 1996, hours 19-01 UT*

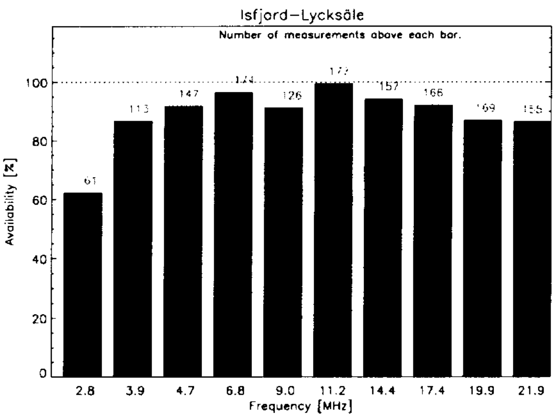


Figure 3 *Modem availability on propagating channels for STANAG 4415 (75 bps, long interleaver) for period 3-13 Sept 1996, hours 19-01 UT*

5 CAUSE OF MODEM FAILURES

When DAMSON channel measurements have been identified as lying below the characterised surface, the modem is judged not to work. We have then determined which channel parameters cause the failure of the modem by comparing the distance from the DAMSON point to the surface in each of the three dimensions. If the distance below the surface in one particular dimension is greater than a certain threshold, we count the channel parameter in that dimension as being a cause of the failure. All three parameters can be counted as causes of failure for the same DAMSON point. The thresholds provide a guard band which allows for the resolution of the measurements and characterisation. The thresholds are set to 2 dB for SNR, 3 Hz for Doppler spread and 1 ms for multipath spread. If the location of a DAMSON point is such that no distance to the surface can be found in a given dimension, then the respective channel parameter is deemed not to be the cause for the failure. Further, if the DAMSON point is located in the saturated region of the characterisation, the SNR will generally not be counted as being the cause of the failure. This is the general method of determining causes of modem failures. For the robust modem (STANAG 4415) however, saturation is never reached within the Doppler and multipath ranges tested, and the general method can be simplified.

RESULTS FOR THE ISFJORD-LYCKSÄLE PATH

Statistics on the cause of modem failures have been determined for the same data as shown in Figure 2 for the non-robust modem. In Figure 2 we saw that there were more DAMSON channel measurements lying below the characterisation surface than above. (A 50% availability would indicate half the points above the surface and half below). From Figure 4 we see that on all frequencies more than 70% of the data points lying below the surface are more than 2 dB below it, and we thus say that low SNR is the most important factor contributing to the failure of the modem in this period and on this path. Large Doppler spreads and multipath spreads cause failures to a lesser extent, but as stated previously, the selected period was reasonably benign.

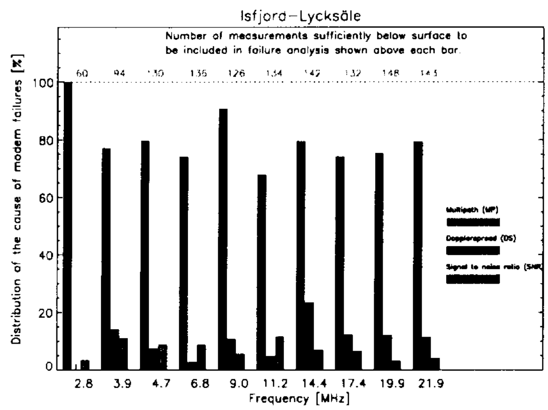


Figure 4 *Distribution of the cause of modem failures for STANAG 4285 (2400 bps, long interleaver) for period 3-13 Sept 1996, hours 19-01 UT*

For the robust modem there were very few data points lying below the surface and none of the modem failures could be attributed to Doppler spread or multipath spread, but purely to low SNR. There were no DAMSON datapoints lying outside the simulated ranges of Doppler spread and delay spread. The trends observed for each modem were the same for the day period. Experience indicates that for a geomagnetically disturbed period the Doppler spread and multipath spread would be a more important cause of failure for the non-robust modem. For the ranges of Doppler and delay spread measured by DAMSON under disturbed conditions, we believe that the robust modem would still give acceptable performance. Further data analysis is required to confirm this.

6 ASSESSMENT OF INCREASED AVAILABILITY OF MODEMS BY HAVING MULTIPLE FREQUENCIES AVAILABLE

If an HF user has only one assigned frequency, that frequency may not propagate at certain times of day, or it may be disturbed by interference so that the availability of communications is low. If the user is allowed to use an additional frequency, then the availability of communications may increase considerably. Further frequencies might be expected to provide further improvements, but it is likely that there will be diminishing gains. Based on signal propagation on the ten allocated DAMSON frequencies and the simulated performance of the two selected modems, we assess here the optimum number of frequencies that a user should have available.

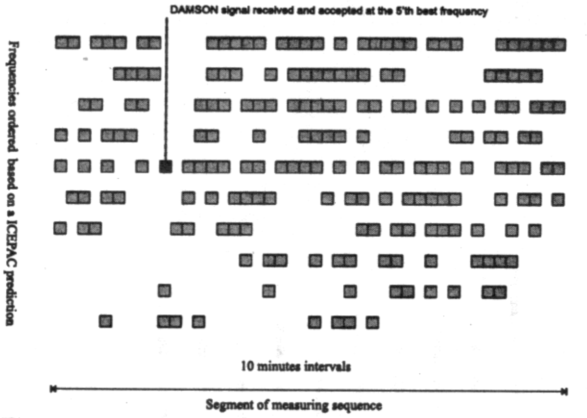


Figure 5 *Method of calculating 'overall' availability of modems when a set of frequencies is available. Boxes indicate that signals have been received and the modem test has been met*

First, predicted reliabilites from ICEPAC (Stewart and Hand (7)) of all the DAMSON frequencies give an individual ranking of the frequencies. The predicted best frequency is then selected, and we assume that this is the only assigned frequency. For a given modem we then count the number of times this DAMSON frequency has been propagating <u>and</u> at the same time the data point lies above the simulated performance surface. This percentage availability represents an 'overall' availability, which means that instances of no propagation have also been considered. We then select the <u>two</u> best frequencies based on the predictions, and count the number of instances where <u>one</u> or <u>both</u> of the frequencies have been propagating and at the same time channel conditions are sufficiently good that the modem would have worked. This analysis is continued for three, four, up to ten DAMSON frequencies. The method is illustrated in Figure 5.

RESULTS ON THE ISFJORD-LYCKSÄLE PATH

The increasing 'overall' modem availability derived from having multiple frequencies available is shown in Figure 6 for the non-robust modem and in Figure 7 for the robust modem. The time period is the same as for the previous analysis. For the non-robust modem the availability continously increases to a maximum of 71% until nine frequencies are added to the frequency set. However, the largest gains are achieved by adding the first two frequencies to the set. If the frequency ranking from prediction is not in agreement with the modem performance on the different frequencies, this will be shown in the figures as a lower increase in availability from one frequency to the next than for subsequently added frequencies.

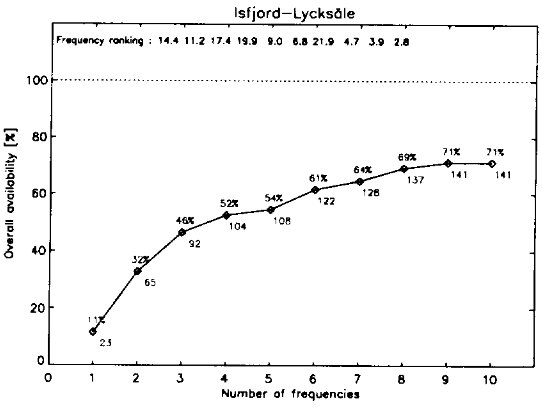


Figure 6 *'Overall' availability of STANAG 4285 (2400 bps, long interleaver) when the frequency set consists of 1,2,...10 frequencies. Period 3-13 Sept 1996, hours 19-01 UT. Maximum possible number of measurements for period: 198*

For the robust modem a maximum true availability of 90% is reached by just having two frequencies in the frequency set. Better availability is not achieved by adding more frequencies.

For daytime conditions (hours 09-13 UT) the situation is similar. For the non-robust modem seven frequencies are needed in order to read a maximum overall availability of 52%, whereas for the robust modem 77% overall availability is achieved by three frequencies and 80% is achieved by six frequencies.

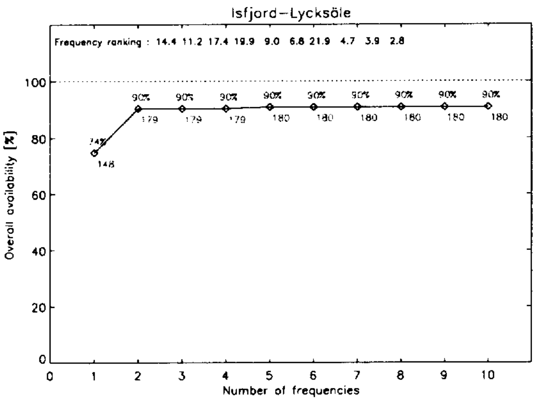


Figure 7 *'Overall' availability of STANAG 4415 (75 bps, long interleaver) when the frequency set consists of 1,2,...10 frequencies. Period 3-13 Sept 1996, hours 19-01 UT. Maximum possible number of measurements for period: 198*

The number of necessary frequencies found here based on the DAMSON data may be slightly lower than for a traditional communications system since DAMSON uses pulse compression sounding which suppresses interferers.

7 THE HARSTAD-KIRUNA PATH

This is a short, 190 km east-west path at latitudes 67° -68°N. This DAMSON path shows excessive multipath at all times and Doppler spreads in excess of 80 Hz when the ionosphere is disturbed. Even if the path is short, propagation exists on all DAMSON frequencies because of off-great-circle propagation. Seven days of data from a quiet period in April/May 96 have been analyzed.

Modem availability on propagating channels for the non-robust modem is, for this path, extremely low; only 20% for a few low frequencies. In general, the modem fails because of low SNR, particularly at the higher frequencies. During daytime, large multipath is another cause of failure for up to 70% of the DAMSON datapoints. During the night, Doppler spread is the next most important factor that causes the modem to fail (up to 50 % of data points).

For the robust modem the availability is 90-100% for the lower frequencies, decreasing to approx 10% for 21 MHz. Low SNR is the only reason for this modem to fail on this path.

There is a large gap between the maximum overall availability that can be reached for non-robust and robust modems on this path. However, the number of frequencies necessary in order to achieve the maximum availability is the same for both modems; two frequencies are necessary during day and three frequencies at night. So if frequency assignments are made according to ICEPAC predictions, the data analyzed here show that just three frequency assignments are enough for this path.

8 COMPARISONS

Results from performance analysis depend on factors such as antenna type and transmitting power. Due to these reasons, availability results from the modem analysis will change whenever some of these factors are modified. But, if availability results for several modems are compared and the data used is taken from the same measurement campaign, the comparisons will show the tested modems mutual performance.

Despite the increased use of datatransmission on HF, voice and morse are still important methods for establishing communication. Performance characterisation is assessed for voice and morse, and represented as a BER surface with the method described by Arthur and Maundrell (2). In order to show the improved availability using the robust modem, we have compared the robust modem with the non-robust modem, and with voice and morse. The mutual performance is shown in Figure 8.

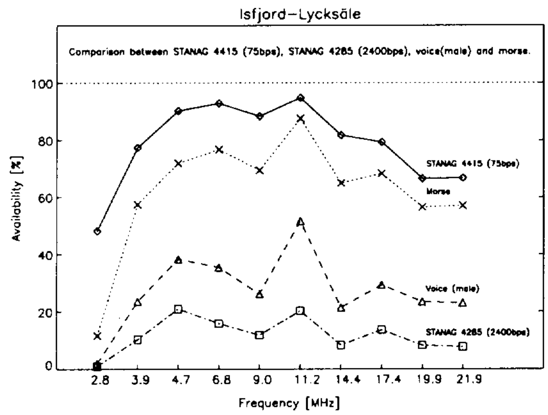


Figure 8 *Comparison between STANAG 4415 (75bps, long interleaver), STANAG 4285 (2400bps, long interleaver), voice (male) and morse for the period 3-13 Sept 1996, hours 00-24 UT.*

Availability for the robust modem is exceeding 65% on all frequencies above 3 MHz, with morse lying in general 10 to 20% lower. Both voice and the non-robust modem show a significantly lower availability on all DAMSON frequencies. All these communication methods are vulnerable for a decrease in received SNR while an increase of Doppler and delay spread would primarily reduce the availability of the non-robust modem due to its low tolerance to both Doppler and delay spread. The robust modem, voice and morse would not experience a significantly reduction of availability due to their tolerance to high Doppler and delay spread.

9 CONCLUSIONS

One period on each DAMSON path has been analyzed so far. The periods have been arbitrarily chosen based on data availability, and they are from different seasons. General conclusions that can be drawn based on all the paths are as follows:

The availability of the robust modem (STANAG 4415, 75 bps) when signals are propagating is 50 %-70 % higher than for the non-robust modem (STANAG 4285, 2400 bps).

When both modems fail, even if signals are propagating, the most important reason is low SNR of the received signal. Doppler spreads and delay spreads are not seen to be the reason for failure of the robust modem, but for the non-robust modem these two channel parameters can significantly contribute to the failure. On the short Harstad-Kiruna path, Doppler spread is an important factor at night, whereas multipath spread is more important during day.

The ranking of frequencies based on ICEPAC predicted reliability shows good agreement with the experimental data.

The non-robust modem needs more assigned frequencies (6-8) to reach a combined maximum 'overall' availability than the robust modem (3-5). However, on the short Harstad-Kiruna path, 2-3 frequencies are enough for both modems.

The number of necessary frequencies is not different for night and day, except on the short path where 2 frequencies are necessary during the day and 3 during the night for both modems.

Comparing a low data rate modem with a high data rate modem may seem odd since they are intended for different purposes. This study points out that data rate in certains areas must be sacrificed for availability and a limited frequency resource.

REFERENCES

(1) STANAG 4285 (1990): Characteristics of 1200/2400/3600 bit per second modulators /demodulators for HF radio links, NATO, Brussels

(2) Arthur P C, M J Maundrell (1997): Multi-dimensional HF modem performance characterization, *Conference Publication 441*, IEE, 7th Int Conf on HF Radio Systems and Techniques, Nottingham, UK

(3) STANAG 4415 (1998): Characteristics of a robust non-hopping serial tone modulator /demodulator for severly degraded HF radio links, NATO, Brussels

(4) Davies N C, P S Cannon (1993): DAMSON- A system to measure multipath dispersion, Doppler spread and Doppler shift on multi-mechanism communications channels, *CP-543*, AGARD Symposium on communication channels, Rotterdam, Netherlands

(5) Angling M J, P S Cannon, N C Davies, T J Willink, V Jodalen, B Lundborg (1998): Measurements of Doppler and multipath spread on oblique high-latitude HF paths and their use in characterizing data modem performance, *Radio Science, Vol 33, no 1, pp 97-107*

(6) Bergsvik T (1998): Evaluering av HF-modemer og flerfrekvens systemer paa hoye breddegrader basert paa kanalmaalinger, *FFI/Rapport-98/04983*

(7) Stewart F, G Hand (1990): ICEPAC, private communication, ITS, U.S. Dept. of Commerce, Boulder, Colorado.

Mobile Internet/Intranet Access

Davras Yavuz, Paul Vollebregt
NATO C3 Agency, Communications Systems Division, Radio Branch
PO Box 174, 2501 CD Den Haag, The Netherlands

Tel: 31 70 314 2422 (DY) - 31 70 314 2407 (PV) Fax: 31 70 314 2176
yavuz@nc3a.nato.int **vollebregt@nc3a.nato.int**

Abstract

The phenomenal expansion of Internet/Intranet services/networks has been driven by six technologies:
1. Semiconductor computer/processor and memory developments
2. Photonic/fiber-optic transmission revolution
3. Signal processing and related software developments
4. Packet switching, protocol and related developments
5. WWW/HTML standards/developments
6. Magnetic/optical storage developments

These advances have of course also impacted many other areas but the emergence of Inter/Intranets and the previously unimaginable services provided by them is definitely the most significant.

These developments now make mobile access to Inter/Intranets desirable and in some cases essential in many applications. TCP/IP (IP V4) which is the backbone of present Inter/Intranets, was not intended to support mobile users but available/ongoing mobile IP and IP V6 developments will soon provide some of the additional required functionality. In the meantime useful mobile access can be provided with suitable gateways as has been demonstrated by various implementations by STC/NC3A in the past three years. When considering mobility, distinction must be made of line-of-sight[1] (LOS) and non/beyond LOS access. With LOS access capacity limitations are much less severe. With non-LOS access capacity available depends whether an active relay (e.g. SATCOM) is available or not. The type of satellite (GEO, MEO, LEO) will determine the size, characteristics of the mobile transceiver. Mobile non-LOS access relying on ionospheric or other natural phenomena has severe capacity limitations and as a result, services/applications intended for fixed networks are not generally suitable. However, such services as email, browsing web sites with minimal graphics can be highly practical. We will focus on non-LOS access as this has important tactical applications in mountainous terrain and for long distance access in terrain virgin from a communications infrastructure viewpoint.

STC/NC3A has implemented a non-LOS HF access system using the STC packet radio protocol developed in 1991 with a single-tone modem. The system was used in 1997 during joint trials with Norway on a coast-guard/icebreaker ship in the Norwegian polar seas from locations as far north as 83^0. Internet gateway/access was through the STC/NC3A remote site at Staelduijnen. Since then this gateway has been available for various trials and continues to be made available on request. It has been estimated that six or seven such gateways suitably located in NATO member nations and with appropriate frequency management could provide Internet access essentially from any location. Details of the system used and some of the results of these trials will be presented. Some information on the mobile IP activities will also be given.

1. Introduction

Internet revolution has its origins in the fundamental work done for ARPA/DARPA in the sixties; USAF/Rand Corp "Distributed Communications" edited by Dr. P. Baran, published in 11 volumes in 1964 addressing means for highly reliable communications in adverse conditions. The word "packet" not having been thought of yet, the expression "distributed adaptive message block network" is used instead. The word "packet" replaced this expression in the late sixties. The aim of that study was survivable communications under highly disrupted environments (natural & man made), through richly connected redundant networks through which "packets" of data could be routed obtaining high probability of messages delivery. In Dec 1969 first flow of "packet" data took place in ARPANET comprising of four nodal processors spread out in the USA. By 1975, more than 100 university computers were inter-connected. Although, the military community, in this case USAF, was the primary sponsor of this activity, the commercial world was far more quicker in realising the potential of these techniques and soon exploited them for banking services/ATMs, travel/ticketing services, ... etc. These have subsequently lead to the present TCP/IP (V4)

[1] We imply radio line-of-sight (LOS) which is in general slightly more than visual LOS due to atmospheric ref-ractive/bending effects which are function of frequency.

Paper presented at the RTO IST Symposium on "Tactical Mobile Communications", held in Lillehammer, Norway, 14-16 June 1999, and published in RTO MP-26.

protocol stack and the Internet "revolution" that began in the late eighties.

The phenomenal expansion of Internet/Intranet services/ networks has also been driven by the convergence of other technologies given below which are simply mentioned for completeness and will not be discussed:
- Extraordinary developments in semiconductor computer/processor and memory devices
- Photonic/fiber-optic transmission revolution providing essentially $\propto$ capacity for fixed links
- Signal processing and related software developments
- WWW/HTML standards/developments
- Magnetic/optical storage developments

2. Mobile Access

With the proliferation of Internets (and closed networks or "Intranets" using the same basic technology), requirements for mobile access are also increasing. However, the practically limitless capacities available for fixed links/networks due to the photonic/fiber-optic revolution has no counterpart for mobile links/networks [we define a "mobile link" as any link with at least one end mobile]. As shown in **Figure 1** the divergence between capacities available from fixed networks to those from mobile ones is enormous and is still increasing. Mobile links must primarily still rely on electromagnetic (radio) propagation with some short distance, line-of sight (LOS) possibilities with infrared and optical propagation. From the military perspective LOS requirement is highly restrictive as operations in mountainous terrain and also long distance links require non-LOS or beyond LOS mobile communications in many cases. SATCOM systems are essentially two (or more) LOS links with an active relay(s) that can be geosynchronous (GEO), medium (MEO) or low earth orbit (LEO) and provide such communications. However, at present, the cost is high, infrastructure required is complex and as a result availability is limited in most mobile/tactical scenarios of interest to NATO. Cellular systems and public mobile radio systems can also be used for mobile access as we shall cover briefly below.

There has been some improvement in the capacities achievable with radio systems. These have been primarily with LOS systems where, moving to higher frequencies, signal processing, signal constellation optimization combined with developments in coding theory have resulted in increases. However, these improvements are nowhere near the increases due to the photonic revolution where capacities available to fixed links/networks have multiplied exponentially. There are some futuristic ideas in research laboratories involving laser beams following mobile platforms, soldiers in the field, ... etc., but these are beyond our scope. The possible propagation modes for non-LOS radio communications are; ionospheric (HF-sky wave), ground/surface wave (low-HF and below), meteor scatter (low VHF). Troposcatter is another possibility but is not really suitable for mobile operations because of the large powers required and the increasingly unwanted spectral pollution properties. We will concentrate on systems based on ionospheric and surface propagation (HF sky and surface waves).

As result of the above elaborated divergence, the mobile user will, for the foreseeable future, always have significantly smaller capacities available. This limitation will be even more drastic in non-LOS environments where some form of natural phenomena (ionospheric reflection, refraction, ...) or SATCOM systems capable of supporting small terminals will need to be utilized. Applications designed for the capacity rich environment of fixed networks will clearly not perform well in the much lower capacity environment of mobile networks. It is therefore essential that, applications must be specifically designed/tailored for mobile networks. The unpredicted speedy proliferation of the Internet has, in a sense, caught us off-guard in this respect and much work is now being initiated and will need to be done. STC/NC3A activities described below have focused on military applications and particularly non-LOS scenarios.

3. Related STC/NC3A Activities

The STC HF packet radio ARQ protocol that was developed in 1991 was later used as the basis for the Maritime Gateway system, HF component of the CSNI multinational trial network and the PC-NET system a Partnership-for-Peace (PfP) applications. The Maritime Gateway is a demonstration system initially running on SCO UNIX and now ported to the Windows NT platform as a HF Mail terminal. The system can also be configured to provide exchange of normal e-mail messages, files to/from the Internet. A version for NAEW applications has also been developed and is now in use both on the ground and in the NAEW aircraft as expanded below. Various other "flavors" of the same basic product for HF data communications have been developed and trialled.

The aim of PC-NET is to provide e-mail like communication capabilities between PfP and/or NATO member ships as was identified during the first peacekeeping field exercises held in the autumn of 1994. Similar systems with certified security devices have been developed and implemented for NAEW and a number of land mobile applications.

PC-NET provides a rapid deployable integrated tactical message capability for ships at sea, especially within a Navy Battle Group, to communicate targeting, weather, operational tasking, and imagery information. PC-NET is a packet-based, digital ARQ communication protocol, supporting Navy messaging between PfP and NATO ships and/or shore via different communications media that include HF, VHF as well as UHF. The system is offered as a COTS package, the package includes special software and hardware for the Microsoft Windows PC based platform. The system can work with any standard SSB HF radio as was demonstrated during various exercises including; Co-operative Partners in July 1997 in Varna Bulgaria, Sea Breeze in August 1997 in Odessa Ukraine and Strong Resolve in 1998 in Denmark, Norway and Italy, and many ongoing exercises.

All these communication systems make use of a basic automatic-repeat-request (ARQ) error-control scheme. This method of error control is simple and provides high system reliability. With an appropriate Cyclic-Redundancy-Check (CRC) polynomial for error detection, virtually error-free data transmission can be attained. Basic ARQ systems offer the best performance/complexity ratio compared with other error-control schemes, Hybrid ARQ schemes are much more complex to implement and only offer advantages with high bit-error-rates (BER). Basic ARQ schemes will give a much better performance on channels with a fair amount of errors compared with Go-back-N error–control scheme's used in AX25, an amateur radio protocol derived from X25. In ARQ systems the transmitted data is segmented in small segments (packets). During transmission a number of these packets are transmitted whereby every packet has its own CRC check and reference number to ensure proper error detection. The receiving node will continuously check for CRC errors and mark packets received in error, immediately following the transmission interval the receiver will request for a retransmission by the transmitter of those packets that were received in error by using the reference number.

Hybrid ARQ protocols use a different and more complex method of error-detection and correction scheme than basic ARQ protocols. Hybrid I ARQ schemes transmit Forward-Error-Correction code (FEC) along with the packet, this FEC data is only used when errors have been detected in the packet by means of CRC check. First the redundant FEC information transmitted together with the packet will be used to correct errors, the normal ARQ retransmission scheme will be employed when there are still errors after FEC correction has been applied. The advantage of this scheme is the reduction of retransmissions and thereby increasing the throughput, however the maximum throughput by low BER will be lower due to the FEC overhead on every packet. Hybrid II ARQ schemes transmits FEC code only when the receiving node requests a retransmission thereby reducing the overhead in low BER conditions. When BER increases the node will start retransmitting more and more FEC packets, these will be used to correct errors by applying code combining error correction techniques. The disadvantage of this method of error correction is that often more FEC is transmitted than what was absolutely necessary and thereby increasing overhead and lowering the overall throughput. Hybrid III ARQ schemes are trying to overcome these problems by only transmitting those parts of FEC data that are needed to correct the segments in the packets received in error. Thereby reducing the FEC overhead to a minimum resulting in a much-improved throughput even under poor channel conditions.

STC/NC3A HF data systems make use of a basic ARQ scheme in combination with other techniques to increase overall throughput under varying channel conditions, one of the techniques is adaptive modem speed whereby the effective modem speed is changed to increase effective user throughput (**Figure 2**). The other technique reduces the number of packets transmitted during a transmission interval and thereby reducing the end-to-end delay of the link. During studies conducted prior to the development phase of the STC HF protocol in 1991 it was concluded that dynamic packet length would only give a marginal throughput improvement under varying channel conditions. One of the conclusions of the development phase was that dynamic transmission speed would give much better performance improvements compared to dynamic packet length resulting in a higher effective user throughput. The dynamic speed decision algorithm is constantly analyzing the link quality and based on that is making decisions whether to select a higher or lower modem speed (**Figure 2**). Link quality measurements can be based on channel SNR and/or BER numbers, both PC-NET and the Maritime Gateway make use of BER numbers, only since information obtained from the modem regarding SNR are not reliable enough and standardized.

A number of systems have been operational based on above developments at STC/NC3A, the foremost of which is the Data Terminal(Air) [DT(A), and ground/deployable versions DT(G)/DT(D)] in use with the NAEW aircraft since about two years. The system provides long-distance/non-LOS secure data communications to/from the NAEW aircraft using any one of the HF transceivers in the aircraft; error free transmission of binary and/or ASCII files with certified security devices, trading off channel conditions with message delivery time. Packet radio with ARQ, as described above, is essential for reliable connectivity over difficult to predict channels and HF sky-wave/ionospheric channel, particularly to/from a mobile user, is certainly in that category.

Inter/Intra net access can also be implemented with cellular systems such as GSM, DCS, ... and conventional VHF/UHF private mobile radio (PMR) systems. An emerging open standard PMR system defined by ETSI[2] is TETRA[3] providing data communications for security, safety, healthcare and other emergency services which is also suitable for certain classes of multinational military applications, such as crises management and peace-keeping. All these systems require extensive infrastructure in the form of base-stations, switches and high capacity interconnections between them (TETRA does have a direct mode that can provide LOS mobile to mobile capability without any infrastructure). SATCOM systems such as INMARSAT, IRIDIUM, ..., have distinct advantages from the coverage viewpoint but also require costly and complex infrastructure, especially for lower orbit constellations. These systems are all essentially LOS requiring intermediate active relays in the form of base-stations or satellites but can provide excellent service when available. STC/NC3A has performed various tests/trials for possible operational applications of such systems but these can not be covered in this brief presentation (some information may be found at our web site www.nc3a.nato.int).

4. Mobile IP

The tremendous proliferation of Internet during the last decade has made TCP/IP (V4) by far the most dominant and cost effective protocol stack for data communications applications. The evolving mobile IP and next generation IP(V6) activities/standards aim to facilitate mobile access to the ever expanding Internet Infrastructure. This technology will also apply directly to closed Intranets. A number of initial implementations are already available freely on the Internet and also commercially in parallel with much ongoing activity.

Since Internet and the present generation of protocols (IP v4) was not developed with mobility in mind some new elements (software agents) are required to support mobile operation. The basic essential features are that the Internet structure should not need to be changed; TCP connection should be maintained as mobile node moves without any changes to the correspondent nodes with which the mobile wishes to set up connections. This requires a "home agent", "foreign agents" and extra functionality (software) in the mobile node.

Briefly mobile IP operation is as follows :

1) Mobile node is always addressed by it home IP address (i.e. messages from the correspondent node always go via the home network)
2) When mobile is away messages are captured by the "home agent"
3) When mobile node is away it receives a "care of address" from the "foreign agent" of the foreign network
4) Mobile node send its "care of address" to its "home agent"
5) "Home agent" forwards messages to the mobile node

There are security some security issues (malicious home & foreign agents and possibility of messages to home agent changing care of addresses, etc.) that are presently being explored. Possible solutions offering some level of protection can be implemented but are outside our scope in this presentation.

5. Conclusions

Mobility support to Internet users is being increasingly demanded and will need to be provided in some manner. The large divergence between capacities available to fixed networks (result of essentially limitless capacities from photonic/fiber-optic networks) and mobile networks which must still rely on electromagnetic propagation is a significant problem and will continue to be so.

STC/NC3A has developed a number HF data systems building upon the open HF data protocol work initiated in 1989. This activity resulted in a number of versions of the protocol which was provided to many NATO national institutes and companies freely from 1991 onwards, to catalyze open data systems on HF. Subsequently, various versions of HF data systems based on the same basic concept comprising of:

1) HF packet radio protocol
2) Single tone modem supporting adaptive speed
3) A standard PC possibly with the addition of a standard processor card

Maritime, air/NAEW, land and PfP versions of the HF data terminal have been implemented and some are now in operational use.

[2] European Telecommunications Standards Institute

[3] GSM: Global System for Mobile cell-phone communications TETRA: TErrestrial Trunk Radio Access system for security, health, disaster and related commercial PMR (Private Mobile Radio) services. Both are ETSI standards, GSM is implemented in more than 80 nations with >160 service providers, TETRA systems are now becoming available.

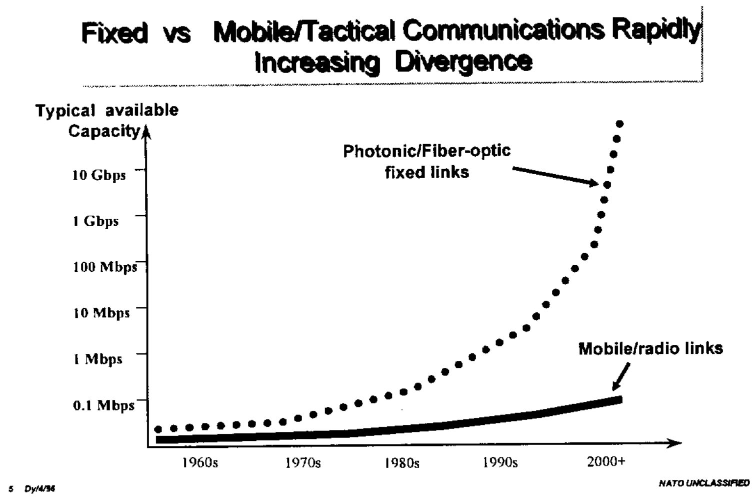


Figure 1. The enormous divergence between capacities available to fixed networks/links through photonic transmission (fiber-optic) techniques and the gradual increase in the capacity available to mobile networks/links. Note that the vertical scale is logarithmic.

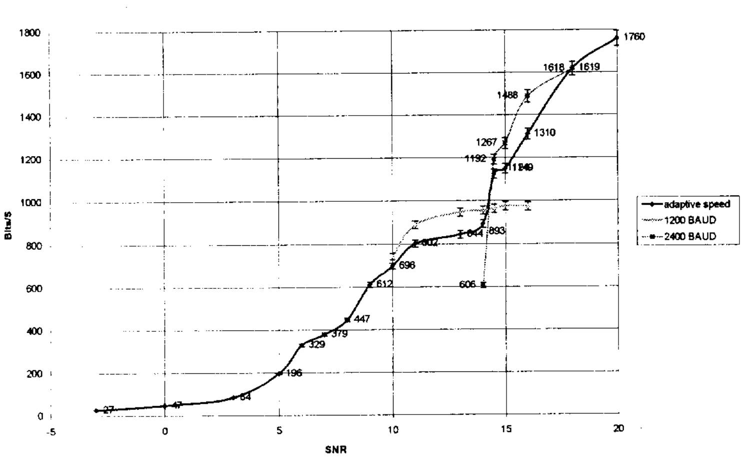


Figure 2. Comparison of fixed modem rate and adaptive modem rate operation for a single-tone modem measured with Harris 5710 modems in the PC-Net implementation, with an HF channel simulator for which the conditions are given below. Half duplex single cahnnel with Z-modem application providing a continuous bit stream (Further details are available in STC TN-506, TM-937 additional background in Clark & Eken paper in HF York Conference Proceedings 1994 and more recent developments in NATO STANAG 5066).

	Path 1	Path 2
Rel Pow (dB)	0	0
Delay (ms)	0	1
Doppler sp (Hz)	2	2
Diff freq offset	0	0

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO-MP-26 AC/323(IST)TP/2	ISBN 92-837-1022-3	UNCLASSIFIED/ UNLIMITED

5. Originator Research and Technology Organization
North Atlantic Treaty Organization
BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Tactical Mobile Communications

7. Presented at/sponsored by
the Information Systems Technology Panel (IST) Symposium
held in Lillehammer, Norway, 14-16 June 1999.

8. Author(s)/Editor(s)	9. Date
Multiple	November 1999

10. Author's/Editor's Address	11. Pages
Multiple	324

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Tactical communications
Mobile equipment
Secure communication
Military communication
Command control communications
PCS (Personal Communication Systems)
COTS (Commercial Off The Shelf)
Protocols
Communications networks
Wave propagation
Speech recognition
Signal processing
High frequencies
LAN (Local Area Networks)
WAN (Wide Area Networks)

14. Abstract

This volume contains the Technical Evaluation Report and 35 unclassified papers, presented at the Information Systems Technology Panel Symposium held in Lillehammer, Norway from 14th to 16th June 1999.

The papers were presented under the following headings:
- Personal Communications and COTS
- Protocols and Networks
- Propagation
- Speech and Signal Processing
- High Frequency

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
Directeur - Recherche et développement - Communications et gestion de l'information - DRDCGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

FRANCE
O.N.E.R.A. (ISP)
29, Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

HONGRIE
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Centro documentazione tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

PAYS-BAS
NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

POLOGNE
Chief of International Cooperation Division
Research & Development Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

REPUBLIQUE TCHEQUE
VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko obranného výzkumu (NISČR)
Mladoboleslavská ul., 197 06 Praha 9

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Telefax 0(1)55.61.22.99 • E-mail mailbox@rta.nato.int

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone.**

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evère, B-1140 Bruxelles

CANADA
Director Research & Development Communications & Information Management - DRDCIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

CZECH REPUBLIC
VTÚL a PVO Praha /
Air Force Research Institute Prague
Národní informační středisko obranného výzkumu (NISČR)
Mladoboleslavská ul., 197 06 Praha 9

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1, P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (ISP)
29 Avenue de la Division Leclerc
BP 72, 92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

HUNGARY
Department for Scientific Analysis
Institute of Military Technology
Ministry of Defence
H-1525 Budapest P O Box 26

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Centro documentazione tecnico-scientifica della Difesa
Via Marsala 104
00185 Roma

LUXEMBOURG
See Belgium

NETHERLANDS
NDRCC
DGM/DWOO
P.O. Box 20701
2500 ES Den Haag

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
NO-2007 Kjeller

POLAND
Chief of International Cooperation Division
Research & Development Department
218 Niepodleglosci Av.
00-911 Warsaw

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076-1320
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-837-1022-3