Evolution of Networking: Megatrends, Current Problems, and Future Directions

Dr. Pekka Nikander

Ericsson Research Nomadic Lab
Hirsalantie 11
FI-02420 JORVAS, Finland

pekka.nikander@nomadiclab.com

Helsinki Institute for Information Technology
Metsänneidonkuja 4, P.O.BOX 9800
FI-02015 TKK, Finland

pekka.nikander@hiit.fi

ABSTRACT

Networking, as a piece of technology, seems to be on the brink of its second fundamental revolution. The first revolution was the move away from wires to packets; a phenomenon that started in late 1950's, gained speed in the 1980's and especially 1990's, with the wide adoption of the Internet and related technologies, and seems to be coming to its conclusion during the next decade. The second revolution, moving from packets to information, is only starting now. It has been most prominently discussed by Van Jacobson, one of the grand old men in Internet technology, in his February 2006 Stanford Clean Slate seminar and again in August 2006 Google Tech Talk.

Partly referring to the insights of Dr. Jacobson, this presentation starts with the megatrends: the evolution of networking from interconnecting wires through interconnecting nodes to a world of interconnected information. To emphasise the need for a long-term fundamental change, a number of primary reasons behind many of the most pressing contemporary problems in the Internet are discussed, including loss of trust, surge of unwanted traffic, choking of the routing system, poor support for mobility and multi-homing, and lack of privacy and accountability. From there, a number of currently prevalent approaches for alleviating these problems are considered, including proliferation of middle-boxes, moving connectivity to overlay networks, and virtualisation. The presentation concludes with a brief look at a few of the most interesting research directions going on at the time, illustrating some aspects of the envisioned information-centric networking paradigm.

1.0 INTRODUCTION

Communication is a basic human need, necessary for running complex organisations and societies. Various inventions allowing remote communications, from drum and smoke signals through the 18th century French and Swedish optical semaphore systems to the contemporary worldwide digital networks, have each shaped the society in different ways, reflecting the new communication capabilities. From the highest abstraction point of view, new technical inventions tend to change the related social structures in new and unpredictable ways, gradually leading to other innovations.

In the networking world, so far the most important, still-ongoing change has been the shift from circuits-based to packet-based communication. Especially in the wireline world, the over-a-century-old, connection-oriented telephony network is fast being replaced with a packet-based network, utilising Internet technologies. Analogous development is also visible in most wireless networks, though in a slower pace due to a number of technical and market reasons.

Nikander, P. (2007) Evolution of Networking: Megatrends, Current Problems, and Future Directions. In *Emerging Wireless Technologies* (pp. 1-1 – 1-18). Educational Notes RTO-EN-IST-070, Paper 1. Neuilly-sur-Seine, France: RTO. Available from: http://www.rto.nato.int.

At the same time with the packet-based networks taking over the last remains of circuit switching, a new major change seems to be emerging. The focus in networking research is shifting from end-to-end, node-oriented view towards an information-centric view. It will no longer be important where a piece of information physically resides, i.e., which node is hosting it. Instead, the focus is on making the information available wherever it is needed, as independently as possibly from mobility, security, connectivity, and other real-life complications that the nodes in the network may undergo. Technically, achieving such new kind of networking will be immensely complex. A major challenge is to figure out how to design information replication mechanisms that work reliably in a network where the operators and users are typically both co-operating and competing at the same time, and may occasionally be even mutually hostile, while still maintaining some kind of (perhaps indirect) connectivity.

In this presentation, we will first explore the revolutionary trends themselves. Starting from the early days of networking, based on interconnecting wires, we will carry on to the present interconnection of nodes, and finally try to outline what the world of interconnected information might be. After that, we will dive deeper into the root reasons for a number of present problems, including unwanted traffic, problems in the routing system, difficulties in supporting mobility and multi-homing, problems with dealing with resources and congestion, privacy and accountability, and trust and reputation. In Section 4.0, we will briefly discuss some of the current, short-to-medium-term attempts to alleviate some of the problems, including some of the shortcomings of the alleviation attempts themselves. Finally, in Section 5.0, we will briefly outline some longer term research directions, including the information-oriented approaches to the networking. Section 6.0 concludes the presentation.

2.0 MEGATRENDS: WIRES — NODES — INFORMATION

As pointed out by Van Jacobson in his seminal talks in February [1] and August 2006 [2], the networking technology appears to be on the brink of its second revolution. The first revolution was the about four decades long shift from interconnecting wires to interconnecting hosts, or devices. It has its roots in the 50's and is coming to its end only now, with the slow disappearance of the last traditional telephone backbones. The second revolution will be the shift from focusing on devices to the information itself. With the advances in technology and concurrently changing lifestyles will make it gradually impartial which devices we use; there will be intelligent, networked devices all around of use, to the extend that the typical user can't keep track of them all. Hence, the important thing will be to get the right information at the right place at the right time, instead of merely allowing a user (or an automated agent) to access it remotely.

In a way, we can characterise the situation by noting that the first revolution freed the devices from the slavery of wires, the second revolution will free the information from their bonds to the devices.

2.1 Early Networks: Connecting Wires

In the early days of communication, there were a large number of competing systems. However, over the first half of century or so of electronic communication, the telephone system became clearly the dominant communication system. From the beginning, the telephone network was conceptually based on interconnecting wires. There was a physically separate pair of wires from the central office to the subscriber's equipment; at the central office, the lines were manually connected to create the required interconnection for the phone call.

This basic premise has been carried through all incarnations of telephony, and is still visible, to a degree, in the SIP-based Internet telephony systems. The base goal of the systems is to create a two-way real-time connection between the two end devices, originally for carrying analogue voice signal but today focusing on carrying information in any of a number of pre-defined encoding formats[1].

For the purposes of the present discussion, the important point is in the focus of inter-connecting end-user devices. In a way, we can make a distinction between two levels of concern. On the highest level, i.e., closest to the user, the issue at stake is getting the real-time information representing the media stream to the intended recipient(s) sufficiently fast and error free. At a lower level, closer to the machinery, the means is to create a real, hard-wired or virtual connection between the user devices. In either case, the approach implicitly means allocating resources, i.e., physical wires or digital bandwidth, for the sole use of the particular connection. So, from that point of view, the early era of networking consisted of creating networks of interconnected "wires", either physical or virtual.

2.2 The Internet: Connecting Nodes

With the advent of the packet switched networking, the foci shifted. Closer to the user, the main issue was access to the virtual resources provided by remote computer nodes, i.e., hosts. At the networking side, focus no longer remained on using inter-connecting wires to get the information through as fast as possible but on more effectively multiplexing the resources through packetising the data and handling each packet individually. Using the early era technology, when the available data rates were relatively low, these two approaches seemed completely incompatible; it was simply impossible to acquire much-enough of reliable-enough capacity for real-time voice traffic without explicitly reserving the resources for it.

The different starting point led to a radically different network architecture, the main differences being visible in naming, addressing, and routing. In the early wireline networks, there was only one type of addresses (the telephone numbers), which directly mapped to the individual wires at each central office. Hence, there was almost no real routing; all the necessary information was encoded to the address itself, which directed the rotary switches to inter-connect the right wires. (Of course, later on that all changed, but the underlying attitude didn't change that fast.)

In the packet-based networks, the focus was from early on on utilising path diversity to provide resilience. Each packet was considered free to pick any path among the alternatives, thereby providing inherent robustness against the failures of individual routers and links. However, this different approach brought up the necessity of having an explicit routing and forwarding algorithm, picking up the best next hop for each packet. In order to be able to do that, the algorithm needed to have more information about the state of the network than was necessary before, leading to the evolution of the routing protocols such as OSPF and BGP.

In today's networks, the resource situation is completely different from the early times. Most links have a sufficient capacity to from a few to thousands of simultaneous real-time video streams. For most purposes, including real-time voice and video, there is no longer any inherent need to reserve resources just in order to ensure enough of capacity, with the potential exception of radio-based access networks where the capacity still remains a prime factor. However, from the business point of view, whenever there is persistent congestion at one point in the network, there is a incentive to eventually increase the capacity. Hence, one can claim that currently the only longer-term source of congestion is regulation, dividing the availability of the radio spectrum unequally between the different uses. Of course, at the same time short-

[1] From the technical point of view, there is little if any reason why the communications in SIP are primarily limited to pre-defined formats; presumably, the devices could take care of negotiating the desirable data formats end-to-end, without any help from the SIP-based networking machinery. However, there are strong business reasons for including the encoding formats into the SIP itself so that the operators can exert more control over the communications. Similar business reasons, as associated legacy thinking, are visible in many aspects of the current SIP-based networks.

time-scale congested conditions are a matter of fact, resulting from flash crowds and temporary equipment and link failures.

For our purposes, the main point is the way how the new focus radically changed the architecture; i.e., how the nature of addresses changed and the underlying change necessitated the emerge of new mechanisms, the routing protocols. Another interesting points are more efficient use of resources, resulting from more aggressive multiplexing, and how the nature of the congestion situations primarily changed from blocked calls to degraded quality of communication. There are still other important lessons to learn in the associated socio-economic structures, but I will defer their discussion until Section 3.4.

2.3 The Future: Connecting Information

While the packet-switched networking technology has remained basically the same since the early 1980's. or at least since the introduction of NAT in mid 1990's, the advent of the World Wide Web and the vast increase in bandwidth have made a large number of new uses of the network possible. The network is no longer used just to convey written or otherwise recorded messages or pass real-time voice or video, but for entertainment, e.g., interactive gaming, and, more generally creating a new kind of social consciousness. The younger generations consider it normal to be "on-line" all the time; i.e., being all the time reachable by their friends independent of their location on the globe. More recently, the usage patterns have started towards sharing presence and experiences; i.e., using the network to roughly keep track of the whereabouts of the friends and other acquaintances, allowing new kinds of quite dense social structures to emerge.

At the same time, the proliferation of intelligent, networked devices and search services is gradually making it impractical to identify information by the device hosting them. The user is no longer interested in finding a file on a certain directory hosted by a certain physical computer. Instead, the user is interested in finding pieces of information related to some issue, concept, or idea. In many cases it is impartial if the information is found in a given format or another, or where within the global Internet the information can be found from. As can be seen from the usage statistics of peer-to-peer networks, there are strong incentives for both acquiring data and making it available even in the case where the legal property holders of that information would like to inhibit such practises.

Even in the case of authoring new content, e.g., writing these words or editing a video, the actual storage location of the content doesn't matter, as long as it will be available at the device (or devices) where the actual authoring or editing takes place. Hence, one can claim that with the new usage patterns and device proliferation the very concept of a host, in the meaning of a computer or network node hosting data or other resources, is becoming obsolete. In any case, most users no longer care about it.

From the networking point of view, this new viewpoint imposes a huge challenge. Once again, the whole architecture needs to be rethought. Once again, the concepts of naming, addressing, and routing will need to be completely re-scrutinised, and most probably some completely new concepts and technologies will emerge, similar to the emergence of routing protocols as a result of the first revolution.

3.0 FUNDAMENTAL PROBLEMS

Receding a little bit from the megatrends closer to the present day situation, we can see that the current Internet is plagued by a host of practical problems, many of them resulting from the very architectural choices made, sometimes consciously but more often blindly, during the original design process in the 70's and 80's. In this section, we have a closer look at a number of them.

3.1 Unwanted Traffic

The various forms of unwanted traffic, including spam, distributed denial of service, and phishing, is arguably the biggest problem in the current Internet. Most of us receive our daily dosage of spam messages; the more lucky of us just a few of them, the more unlucky a few hundreds each day[2]. Distributed denial of service attacks are an everyday problem to the large ISPs, with each major web site or content provider getting their share. And, as we all know, phishing is getting increasingly common and cunningly sophisticated.

The root reasons to unwanted traffic seems to be best characterised with economics. However, before we can get there, we have to take a little detour into architectural thinking.

The current Internet architecture can be considered as a distributed extension of the well-known message passing inter-process communication paradigm[3]. That is, within a single centralised computer, an operating system typically provides resources for and isolates a number of concurrent, parallel processes, each running a separate program or performing a specific task. In any modern operating system, the communication between the processes is strictly controlled by the operating system, in order to provide a level of sanity, thereby preventing bugs or malicious code in one program to cause (much) harm to other programs.

According to the established scientific consensus, there are three basic forms of inter-process communication (IPC): message passing, shared memory, and tuple spaces. In the message passing paradigm, a sending process creates a message and sends it to a specific other process, by naming the recipient with some sort of a name. If shared memory is used, two or more processes have all access to the same piece of physical memory, allowing them all to read and write it. In the tuple space paradigm, such as in the Linda system [3], each process can add (anonymous) messages to a public message board, and any other processes can read them from there. To facilitate practical communication, the messages are usually organised as tuples, allowing each potential recipient to indicate their interest by registering for certain values or value ranges at specific tuple positions.

Now, the current packet-based networking, as we know it from the Internet, can clearly be considered as an extension of the message passing IPC paradigm. In both of them, there is a sender (whether a process or a node) that creates a message, names a recipient, and asks for the underlying machinery to pass the message to the recipient. However, what is noteworthy here is that current Internet implementation gives all power to the sender. Initially, when the sender creates a message and dispatches it, the network has no idea of whether the recipient will be interested in the message in the first place. Indeed, the network makes not whatsoever effort to consult the recipient. Only when the message arrives at the named receiving node (or its proxy, such as a firewall), the recipient's consent is consulted. Only then are unwanted messages dropped.

Hence, from an economic point of view, we can characterise the current Internet as a global, distributed message passing IPC system where the main cost of unwanted communication is paid by the recipient. This is a direct (though certainly unintentional) consequence of the network architecture. By explicitly and directly naming all the potential recipients, we create a system where the senders can easily express their desire to send data to any recipient in the network. Given that under the typical contracts the marginal cost of sending additional packets is very close to zero (up to some capacity limit), there are few or no incentives for refraining from sending unwanted traffic; sending some packets just for fun costs so little that it doesn't matter. At the same time, even a marginal response rate creates a strong incentive for

[2] My personal record is some 500 spams a day. However, the recent trend has been slightly declining, and typically I get only 100-200 spams daily, sometimes as few as 50.

[3] The idea of characterising networking architectures in terms of IPC extensions has been borrowed from a forthcoming book by John Day.

sending unsolicited advertisements, and even a small success rate creates a strong incentive for DDoS-based extortion.

To summarise, our claim is that the current unwanted traffic problem is a compound result from the following factors:

- An architectural approach where each recipient has an explicit name and where each potential sender can send packets to any recipient without the recipient's consent.
- A business structure where the marginal cost of sending some more packets (up to some usually quite high limit) is very close to zero.
- The lack of laws, international treaties, and enforcement structures that would allow effective punishment of those engaging in illegal activity in the Internet.
- The basic profit-seeking human nature, driving some people to unethical behaviour in the hopes for easy profits.

Of course, there is nothing that we can do with the last cofactor, other than to accept it. While the third one is more regulatory in nature and therefore falls beyond the scope of this presentation, there appears to be quite a lot of what we can do to address the first two, more technical aspects. We will return to them in Sections 4.2 and 5.0.

3.2 Choking of the Routing System

Another major problem in the current Internet, the relative choking of the routing system, is perhaps less visible to the average user than unwanted traffic. However, from the overall networking point of view it is at least as important phenomenon as the latter; perhaps it is even a more important one, since if the routing system collapses, a large fraction of the traffic currently carried by the Internet will become practically impossible or at least prohibitively expensive.

The current Internet routing system relies solely on the Border Gateway Protocol (BGP); a protocol that has received some facelifts but internally has remained the same for the last decade or so. At the same time, the business environment where the Internet Service Providers (ISPs) compete has become immensely more complex and competitive. The early-internet attitude of primarily co-operating in order to bring the Internet to the users has all but disappeared and replaced by hard, competitive practises that sometimes flirt with the limits of unethical and illegal [4].

From the operational point of view, the main technical term characterising the current routing complications is *traffic engineering* [5]. While there may be no single exact definition for the term, it is commonly used the desire of the ISPs to control which way other ISPs forward traffic that is destined to them. A basic fact is that the current routing system, i.e., primarily BGP, does not offer any good facilities for it — almost all of the various ways that the ISPs attempt to perform traffic engineering can be considered as some sorts of protocol violations or hacks relying on obscure side effects of the routing mechanisms [6].

From a more technical point of view, there are indications that the routing system itself may be growing close to its limits in terms of scalability [7]. The current core routers routinely handle from 200 000 to 300 000 separate forwarding table entries (each corresponding to a separate IPv4 prefix), and the estimate is that within five years the number will grow to anything between 500 000 and one million. As such, that may not be a large problem; according to some unpublished sources the vendors are well capable of building routers that can handle a few million routing prefixes in an effective and cost-efficient manner [8].

However, a potentially bigger problem appears to be routing table converge. Already now it takes in excess of an hour for a core router to build it routing and forwarding tables after a reboot. Furthermore, there are indications that the routing system may fluctuate days or even weeks after major events affecting the links, such as a recent undersea earthquake near Taiwan that cut a handful of undersea communication cables [9].

As a result of these effects, the result of the complexity of the implementations and the number and diversity of tweaks starts to be astonishing. As one specific example, many BGP routers use the congestion control mechanisms in the TCP protocol to throttle the updates from peers to a rate they can handle. One reason this is done is to prevent local input queues from taking up too much memory. Meanwhile, the peers that would like to send those BGP updates, perform "output queue compression"; i.e., they look into the queue of updates which are waiting to be sent and remove any messages which have been rendered incorrect by events which transpired since that message was put in the queue [10].

Altogether, these issues should be taken as early warnings, indicating that our current routing system may be near its inherent capacity limits. With further growth, the ISPs may no longer be able to perform effective traffic engineering, probably leading to some market consolidation and loss of competition, and the routing system may start to experience global-scale instabilities making large fractions of the Internet unavailable for excessive time periods.

3.3 Mobility and Multi-homing

Effective mobility support requires a level of indirection [11], something that the current Internet architecture is gravely missing[4]. The indirection is needed to map the mobile entity's stable name to its dynamic, changing location. Effective multi-homing (or multi-access/multi-presence) support requires a similar kind of indirection, allowing the unique name of a multi-accessible entity to be mapped to the multitude of locations at which it is reachable.

Within the Internet community, the classical approach to address these problems has been to consider mobility and multi-homing as separate, technical *problems,* something that just needs to be solved through engineering. The main result of this attitude are the Mobile IP protocols, which are architecturally based on re-using a single name space, the IP address space, for both stable names and dynamic locators. White the approach certainly works, it creates two major drawbacks. Firstly, it binds the communication sessions (TCP connections and application state) to the stable home addresses. This, in turn, when combined with the only known scalable solutions to a number of related security problems, creates an undesirable dependency on a constant reachability of the home address. In other words, the Mobile IP architecture is intrinsically bound to the availability of the home addresses.

Secondly, approaches that use items from a single name space for multiple purposes create a number of potential semantic problems. The so called *alias* problem may be the easiest of those to understand. In practical terms, when Mobile IP is used, there are no easy way to tell if two IP addresses actually point to a single host (e.g., due to one being its home address and another one its care-of address) or not. That, in turn, may lead to very confusing problems for quite a large number of applications.

Another solution, quickly maturing from research to practical engineering, is the one proposed in the Host Identity Protocol (HIP) architecture [12]. The HIP approach adds a new name space and a new layer of

[4] The details of why the required piece of indirection is missing are long and contentious, and not repeated here. However, the main point is that while the IP routing infrastructure uses the IP addresses as locators, or names of topological locations, for historical reasons the transport layer, the socket API, and many applications use them as stable, invariant node names. Consequently, if a node changes its location or is available at several locations at the same, the semantic expectations that the upper layers make get violated, requiring a solution within the IP layer, i.e., with only one name space and therefore without the possibility of genuine, fully architected indirection.

indirection, roughly between the IP and the transport layers[5]. The HIP approach is discussed in length in a companion presentation [13].

3.4 Compensation, Resource consumption, and Congestion

As briefly discussed already in Section 2.2 above, the current Internet architecture is, in general, not well equipped to deal with short-term resource shortages. That is, while the TCP and other related well-behaving protocols courteously back up and reduce their resource consumption in the face of congestion, there is nothing in the architecture itself that enforces such behaviour. Consequently, a selfish implementation might, instead of reducing resource consumption, start deliberately sending multiple copies of each packet in order to get a larger fraction of the data through on the first try [14]. The net result would be that the congestion would worsen, eventually leading all well-having implementations to crawl.

At the same time, there has been numerous attempts by the IETF to introduce various mechanisms to allow resource consumption, including the RSVP protocol [15] and the differentiated services packet markings [16]. None of them have received any larger success, at least in the wider Internet.

Again, as in the case of unwanted traffic, we strongly suspect that the core reasons lie in the economic domain. The Internet was fundamentally built upon the idea of at-least minimally co-operating agents. As we already discussed in Section 3.1, in the very core of the Internet architecture lies the assumption that if a host does not want to receive traffic, the sending nodes will cease sending. Similarly, almost as close to the core, lies another strong assumption: the hosts and routers will co-operate in getting the maximal amount of traffic through for the maximum number of hosts. These two assumptions are engraved deep into the architecture and the implementations.

Looking from an economic point of view, and assuming selfish rather than co-operating agents, both of the core assumptions start to appear silly. If an agent has an incentive to send traffic (such as extortion), why should it stop just because someone doesn't want to receive it? Similarly, if an agent benefits from sending more traffic, why should it cease sending it under congestion? For selfish agents, basically nothing. We have arrived to the classic domain of economics: controlling access to limited resources. Hence, my basic claim, shared by some other authors [17], is that the lack of resource and congestion control in the Internet is inherently related to the lack of compensation methods. If the senders are not forced to somehow compensate for the resources they use, or at least overuse, there are no way of creating incentives for stopping them.

Looking at the two problems (lack of resource control and lack of congestion control) more closely, we can see that they are actually the same problem, only working on different time scales. The resource control mechanisms attempt to make sure that there are sufficient resources at all times. The congestion control mechanisms attempt to make sure that the available resources are allocated according to some "fairness" or "relative utility" principle in those cases where the demand exceeds the supply.

As we already discussed, there actually are economic mechanisms for dealing with longer-term resource shortages: the operators will eventually add capacity to anywhere where there are constantly or often congestion, since it makes sense to them. Assumedly, that will also take care of the base resource allocation problem, over longer time periods. Hence, the main remaining problems appears to be congestion control, or devising mechanisms that create a real economic incentive for the senders to cease sending or reduce their sending rate. One potential approach in that space might be the so called Re-feedback scheme by Bob Briscoe et al [17].

[5] Technically speaking, the new HIP layer is co-located with the IPsec functionality within the IP layer, at the boundary between hop-by-hop and end-to-end functions of IP.

3.5 Privacy and Accountability

The final two problems that remain to be discussed are somewhat different than the ones above. While the problems above are more related to the functionality provided by the architecture, and thereby to what can be done with the network and what cannot, especially the privacy problem is more related to what should not be doable. In general, we are interested in preventing bad things from happening, in one hand by imposing restrictions on information flow, and in the other hand by creating explicit incentives for trustworthy behaviour.

The privacy problem is a complex one, with at least three different starting points. From the Orwellian point of view, the question is pretty much about freedom of speech and governmental control. A sufficient privacy system ensures that we can express our opinions and think freely, even when our opinions are socially unacceptable or hostile towards the governing regime, within reasonable bounds (like not committing plainly criminal acts.) The Kafkaesque aspect of privacy focuses on citizen's ability to retain their autonomy without fear of unfounded litigation or other harassing legal/other action [18]. Thirdly, the economic aspect of privacy relates to the fine balance between socially beneficial differentiated pricing vs. socially harmful price discrimination [19]. From these three different points of view, it seems a necessity to provide a reasonable base-level of privacy as a built-in feature in future networks.

The flip side of privacy is accountability. Unbounded privacy encourages unwanted side effects, such as rampant advertising (spam). To counter these, increased privacy requires increased accountability. A key to understanding this apparent paradox is to consider the different dimensions of communication. At the baseline level, we can make a difference between four dimensions: *content* of communication, the *parties* communicating, their *locations*, and finally the very fact that a piece of communication took place. If the system is able to provide strong "insulation" between these so that each party gets only the relevant information, a high level of privacy can be preserved. At the same time, if these components can be combined, post hoc, in the face of criminal activity, it becomes possible to provide accountability. For example, if the party seeing the content doesn't know the real-world identity of the peer and definitely not his or her location, the party seeing the identities doesn't know the contents of the communication nor their locations, and the party knowing the locations has no clue about contents nor identities, the system can be engineered to be highly privacy protecting. If, at the same time, it is possible to combine the knowledge, e.g. through manual actions and relatively pricey cryptographic operations, accountability becomes possible in a way that makes privacy violations hard and costly.

3.6 Trust and Reputation

The final problem we consider is the lack of trust and reputation. The original Internet architecture was build with a fairly homogenous, mutually-trusting community in mind. The minimal assumption was that the other users can, at least, be trusted to honour recipients consent. That is, if a recipient does not want to receive some traffic, the sender can be trusted not to send it. Clearly, the reality today is far from those assumptions. The user community is very large and diverse. The hosts cannot be trusted to respect protocol specifications any more, due to prevalence of botnets and other malware.

Essentially the same problem, in a different disguise, can be also observed closer to the applications and end users. Malware has reduced the amount of trust one can place on their own devices. Phishing and other scams are eroding trust in web services. In general, the surge of criminal activity without suitable mechanisms for control and litigation or other forms of retaliation has created a situation where many people feel genuinely unsafe to conduct their business in the network.

More generally, we can state that the current network and applications suffer from the lack of standardised, wide spread mechanisms for asserting trust and reputation. The few mechanisms that are out there, SSL certificates, S/MIME, and PGP, either contain known fundamental problems (like the SSL

certificate's dependency on DNS names and the lack of suitable hooks in the user interfaces) or are not that widely used, often due to deployment and usability problems. Furthermore, these mechanisms only assert some level of trust and correctness of operation. They do not provide any means for asserting reputation, thereby lacking the most basic economic incentives for continued improvement.

4.0 ALLEVIATION ATTEMPTS

The present short-to-medium term attempts to solve the current problems are typically relatively isolated; i.e, each of them only tries to address one of the problems, or at most a few. In general, they can be divided into two general classes, with a somewhat blurred line in between: Middle-box-based approaches and Overlay-networks-based ones. We now look at these, each in turn.

4.1 Middle boxes and their Problems

The term middle box is commonly used to denote devices, other than bridges and routers, that lie on the communication path between two end points. By definition, middle boxes act on a layer at or higher than the inter-networking layer, basically breaking one or more of the basic end-to-end characteristics of the original IP, IP addresses being non-mutable, non-mobile, reversible, and omniscient [12]. Examples of middle boxes include Network Address Translators (NATs), firewalls, various kinds of application protocol proxies, and TCP accelerators. Typically, middle boxes acts as some sort of proxies at some protocol level, even when they are not called so.

In general, a middle box tries to solve one (or more) of the above mentioned problems (or some other problem) by moving a piece of functionality from a peer host to another network node, the middle box itself. The factors that typically allow such a change of the communication pattern to alleviate problems include the following:

- The middle box may have more CPU, memory, and/or bandwidth in its disposal than the end host, thereby making it better equipped to deal with malicious or other harmful traffic. They may also be placed at a "better" location than the end-host. Firewalls are a typical example in this category.
- The middle box may act as a stable reference point in the network, for example, when the end host is mobile and may change its IP address dynamically. Mobile IP Home Agents can be seen as such middle boxes.
- The middle box may hide the location and/or identity of the real end point. NATs and many other middle boxes provide this functionality, often as a side effect.
- The middle box may be located within a different administrative domain than the end host, thereby changing the trust, reputation, or compensation aspects of communication. This and the next aspect apply to many types of middle boxes.
- The middle box may be known to be better administered than the end host, thereby reducing the possibilities of the end host in engaging itself in criminal or otherwise harmful activity, such as sending spam or violating congestion control.
- The middle box may implement a piece of functionality, such as accounting for billing purposes, that would be impossible to implement on an end host. This "benefit" is clearly a double-edged sword.

While the middle boxes often have clear benefits, as indicated above, they typically also have their drawbacks, often architectural ones. The potential drawbacks include the following:

- Firstly and fore-mostly, middle boxes break the end-to-end principle. In the typical case, the network implements functionality that in principle could have been implemented in the end host; however,

following that principle typically would require changes to the overall architecture, and therefore may not be economically and practically feasible. While the drawbacks may not be obvious, they can be deduced from the end-to-end principle. That is, in the typical case, the middle box creates a point of failure in the communication path, making the survivability of the end-to-end transaction dependent on the middle box. For example, if the middle box crashes, end-to-end communication cannot continue even if there was available an alternative communication path. Relatedly, each middle box creates a potential performance bottleneck and may make deployment of new applications and services harder.

- Some middle boxes plainly and simply break protocols. While this may be intentional and beneficial to the network owner, it is typically harmful to the end user.
- In general, middle boxes make troubleshooting harder. It is no longer sufficient to inspect the end hosts for application problems and the network for problems, but problems related to the middle boxes must also be considered. However, more often than not the end hosts or users may not know about the existence of the middle boxes, thereby making troubleshooting immensely more complex.
- Middle boxes mix poorly with security. For end-to-end security, either a middle box must allow encrypted traffic to flow essentially unchanged through it, thereby reducing its benefit, or the trust model has to be changed in a fundamental way so that the middle box may know the relevant cryptographic keys. In general, middle boxes change the trust model, creating the question if you trust the middle box for your data, especially if the middle box is located within another domain.

In general, to properly reap any of the middle box benefits without drawing oneself knee deep into the troubles, one has to change the architecture in some way. One approach towards this direction is the *Delegation Oriented Architecture (DOA)* by Balakrishnan et al [20].

4.2 Overlay networks

Depending on the case and view, overlay networks can be considered either as architectural generalisations of middle boxes or ad hoc consortia of end nodes assembled to overcome some architectural deficiency. In any case, an overlay approach creates a new layer, or overlay, on the top of the existing IP infrastructure by explicitly employing a number of end nodes as new kinds of relays on an inherently end-to-end path. In other words, what is an end-to-end session from an overlay point of view is a series of independent packet streams from the IP layer point of view.

From an architectural point of view, the overlay approaches add to the current architecture in one way or another. In their simplest form, they simply try to "fix" the current Internet connectivity shortcomings, created by NATs, firewalls, and policy routing, by creating an additional "routing" layer on the top of the IP layer. For example, the Host Identity Indirection Infrastructure (Hi^3) [21] can be considered as such an overlay. Overlays aiming to provide privacy or added protection, such as SOS by Keromytis et al [22] or k-anonymous overlays by Wang et al [23], work basically in the same way but simultaneously provide additional security services.

More complex overlays, such as many peer-to-peer networks, utilise the relaying end hosts for more complex functions than mere routing. They may want to store large amounts of application-level data on the end hosts, or use a large fraction of their CPU or bandwidth resources. Such overlays create an incentive problem; why should the owners of the end hosts to allow other users to consume a large fraction of their resources. In general, there are both market-based and community-based approaches to this problem [24]. Typically, overlay solutions meant to be used by relatively closed user communities rely on people's general trustworthiness and the community-based feelings. In larger or more anonymous overlay networks it typically becomes necessary to employ market-based solutions, such as artificial micro-currencies or explicit reputation [25] to solve the so called freeloader problem.

An important aspect in overlay networks is that they change the naming structure. In the more primitive cases, they merely replace the current IP addressing structure with another destination-oriented name space. However, at the same time they may make denial of service attacks harder by *dispersing* the interface, i.e., instead of choking a single target host the potential attacker must now flood the whole overlay system, which may consist of thousands or millions of nodes. At the same time, they may increase the overall cost of sending by introducing artificial costs or utilising above mentioned co-operation enhancing mechanisms to force the participating nodes to play by the rules. The more advanced overlays further change the rules by changing the naming focus from nodes and locations to pieces of information.

From a wider point of view, many overlay networks, including some peer-to-peer file sharing networks, are already inching towards the future, data-oriented networking (see Section 2.3). They provide the user with a view to the information content, the user not knowing where the data is actually stored. In the typical cases, the data is replicated and it may even be striped, encrypted and/or integrity protected in order to make the network more robust against external disturbance[6]. This can be considered as a primitive form of information inter-connection, an early sample of what the third wave of communication is likely to look like.

5.0 RESEARCH DIRECTION

Leaving back middle boxes and overlays we now proceed to a quick roundup of the more researchy future directions. Roughly speaking, most of the research projects aiming to study and create pieces of the networking future can be divided into *virtualisation*, *recursive or layerless*, or *data oriented* approaches. However, these categories are by no means exhaustive nor exclusive; there are projects that ponder additional dimensions and there are projects that aim to go forward in all of these directions. However, the division helps to structure the discussion; therefore we now consider each of the named approaches one by one.

5.1 Virtualisation

Virtualisation is both a research direction and network reality today. Most businesses use IP-based virtual private networks, and there are products provide both lower-layer (such as Virtual Private LAN Services) and higher layer (such as virtual hosting) services. In general, the research approaches aim to provide a more comprehensive, better architected support for virtualisation.

The business reasons for virtualisation include *pluralism, isolation, customisation,* and *amortisation* [26]. Pluralism allows a multitude of different architectures and services to be provided over a single physical infrastructure, thereby both allowing to provider to better meet the potential demand and the research community to more easily study new, radical architectures and approaches. Isolation makes sure that mutually competing users and usage patterns do not disturb each other. Virtualisation provides better security and resource control. Customisation refers to a less extreme form of pluralism where an essentially same architecture or service is provided in slightly different forms, perhaps in a differently parametrised form, to different users. Finally, amortisation refers to the flip side of pluralism and customisation. Virtualisation allows a single set of infrastructure components to be used for a multitude of different purposes, thereby allowing a single investment to be amortised over a number of services or usages.

[6] Today, in the case of peer-to-peer file sharing, the disturbance takes place in the form of forensics and bogus data injection by the powerful copyright representative organisations. However, putting the legal issues aside, there is a clear pattern: the users of such an overlay network attempt to protect themselves and their business from outside tampering and interruption.

The highest profile virtualisation research projects today are CABO [27] and VINI [28] at Princeton and Diversified Internet [29] at Washington University in St. Louis. The VINI project provides a XORP-based [30] PlanetLab-like environment, running over the U.S. LambdaRail [31] and Internet2 networks, interconnecting a few dozens of nodes. CABO relies heavily on VINI and aims to design and deploy a shared, wide-area experimental facility to support a wide range of research in networking and distributed systems. Diversified Internet seems to take a more hardware-oriented approach to virtualisation, aiming towards a plurality of diverse network architectures to coexist on a shared physical substrate.

Considering the above mentioned problems, virtualisation may help with many of them. They may reduce incentives to send unwanted traffic by restricting it within virtual networks, thereby allowing some networks to continue their operation even in the face of denial of service attacks taking place in other networks. Scalability pressures on the routing system may get reduced by dividing routing in two layers. At one layer the system provides routes for the virtual networks (at a coarse granularity), at the other layer routing works independently within each virtual network (at a finer granularity). However, care must be taken to prevent unwanted side effects, such as oscillation caused by basically independent routing decisions becoming synchronised between different virtual networks. They may help with compensation, resource consumption, and congestion by creating a coarser granularity system where current compensation methods may be sufficient to directly effect resource allocation and where congestion may be limited. They may help with privacy and trust through the natural reduction of scope.

On the other hand, virtualisation is likely to make mobility harder, create their own security problems, add new challenges to the management systems, and create their own trust and reputation systems within the virtual-service provisioning system.

5.2 Layerless and Recursive Architectures

While seemingly direct opposites, layerless and recursive networking approaches have lots of commonalities. Both take a look at the structure of the host-local networking "stack" instead of focusing on resources (such as in the virtualisation approaches) or content (such as in the data oriented approaches).

Research on layerless networking [32] aims towards higher flexibility and better support for different operational environments by releasing the protocols from the straitjacket of strict layering. As practical examples of this, it might make sense to run TCP directly over the local link layer in an ad hoc WLAN environment, or alloy different name resolution mechanisms to be used depending on the networking environment. In general, these approaches tend to provide some kind of a protocol toolbox, allowing the communication system to be configured dynamically by composing these components in different ways, depending on external conditions and local policies.

The recursive approaches [33][34] take a slightly different approach. Instead of aiming towards extreme flexibility and the ability to structure the local "stack", they still adhere to a form of layering. These approaches are based on the important observation that most "layers" in the current stacks tend to have a highly regular superstructure. Each "larger" layer (sometimes consisting of what is today a single layer, sometimes of what considered as two neighbour layers) has basically two different sets of functions: routing and forwarding, and end-to-end error and congestion control. Additionally, to provide for upper and lower layers, some sort of mapping or resolution mechanisms are also needed, in order to convert between the layer-specific naming systems [35].

While the layerless and recursive approaches do not seem to directly provide any remedy against the contemporary problems, other than perhaps mobility due to the ease of adding the required layer of indirection, they inherently add flexibility to the system. Flexibility, in turn, can be utilised to provide better, more optimised services, allowing faster evolution of solutions. Recursive layering may provide scalability for routing by allowing the routing system to become more structured, with more layers. On

the other hand, the added flexibility may deepen the security problems though adding more options in generating and sending different forms of unwanted traffic. It may also add load to the routing system through allowing the end nodes to have more control over their routing, thereby creating a more dynamic operational environment for the routing protocols.

5.3 Data or Information Oriented Networking

The final category in our repertoire of research approaches is data or information oriented networking, directly aiming towards the third wave of networking, as envisioned by Jacobson. In these projects [32][36][37][38][39][40][41], the aim is either to make information the first class citizen, thereby creating an environment of inter-connected information, or provide support for such approaches. Unfortunately, due to the diversity of the approaches, space does not permit us to explore the differences between the various approaches, forcing us only to consider the common denominators.

Basically all of the projects in this category aim to change the notion of what is the network. Instead of focusing nodes and connections between nodes, the approaches focus on information, trying to make the information available where-ever it is needed. Many of the approaches assume that data is self-certifying, i.e., the name of a piece of data can be used to verify the integrity and/or authenticity of the data. Several focus on intermittent connectivity, aiming to maximal availability even when portions of the network may not be always connected or may reside beyond prohibitory slow links, such as battlefield radio links or interplanetary connections. Common approaches include multicast and data replication and caching, creating the problems of scalable multicast routing and cache consistency.

While these approaches do not attempt to directly address any of the current Internet problems, they actually do so indirectly, by changing the playing field. For example, by making data the only addressable object some of the approaches make many of the current forms of unwanted traffic simply impossible, or at least change their nature so much that we cannot understand how such attacks could be launched in the new architecture. The nature of routing, resource consumption, and congestion change fundamentally, allowing the network to have more control over them than today. The mobility and multi-homing problems get replaced by data availability and cache consistency problems.

On the other hand, the problems of privacy, accountability, trust, and reputation still pertain, though in a slightly different form. For example, since the object of interest is now data instead of connections, the privacy problems will now condense around data creation and consumption instead of connection tracking and eavesdropping. Similarly, while the problems of node trustworthiness and reputation become less obvious (though perhaps no less important), a new set of problems related to data content and its reliability emerge.

6.0 CONCLUSIONS

In this presentation we have outlined, inspired by Van Jacobson, the past and current development trends in communication networking through its early era of interconnected wires to the present, with the Internet as the prevailing communication medium, towards a likely future of interconnected information. Along the way we have shown how technical innovations change the nature of human behaviour, thereby changing the nature of needs and problems. We listed a number of major problems plaguing the current Internet: unwanted traffic, cracks in the routing system, mobility and multi-homing, managing resource consumption and congestion, privacy and accountability, and trust and reputation. Finally, we discussed present attempts to alleviate these problems, including middle boxes, overlays, and virtualisation, and briefly described a number of research project, roughly dividing them to virtualisation, layerless or recursive, and data oriented research approaches.

7.0 REFERENCES

[1] Van Jacobson, "If a Clean Slate is the solution, what was the problem?", presentation at Stanford 'Clean Slate' Seminar, Feb 27 2006, Stanford University.

[2] Van Jacobson, "A New Way to Look at Networking", presentation at Google Tech Talks, Aug 30 2006, Google, Inc. http://video.google.com/videoplay?docid=-6972678839686672840

[3] L. I. Patterson, R. S. Turner, and R. M. Hyatt, "Construction of a fault-tolerant distributed tuple-space," in Proceedings of the 1993 ACM/SIGAPP Symposium on Applied Computing: States of the Art and Practice (Indianapolis, Indiana, United States, February 14 - 16, 1993). E. Deaton, K. M. George, H. Berghel, and G. Hedrick, Eds. SAC '93. ACM Press, New York, NY, 279–285. http://doi.acm.org/10.1145/162754.162899

[4] William B. Norton, "The Art of Peering: The Peering Playbook," On-line publication (several versions), Version 1.2, 15 May 2002, Equinix. http://www.blogg.ch/uploads/peering-playbook.pdf

[5] Bernard Fortz, Jennifer Rexford, Mikkel Thorup, "Traffic engineering with traditional IP routing protocols," IEEE Communications Magazine, 40:10, Oct 2002. IEEE, 118–124. http://www.cs.princeton.edu/~jrex/papers/ieeecomm02.pdf

[6] Feng Wang and Lixin Gao, "On inferring and characterizing internet routing policies," in Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement (Miami Beach, FL, USA, October 27 - 29, 2003). IMC '03. ACM Press, New York, NY, 15-26. DOI= http://doi.acm.org/10.1145/948205.948208

[7] Geoff Huston and Grenville Armitage, "Projecting Future IPv4 Router Requirements from Trends in Dynamic BGP Behaviour," in Proceedings of Australian Telecommunication Networks and Application Conference 2006, Melbourne, Australia, 4–6 December 2006.

[8] John Scudder, "Router Scaling Trends," presentation at APRICOT 2007, Feb 26 2007. http://submission.apricot.net/chatter07/slides/future_of_routing/apia-future-routing-john-scudder.pdf

[9] Alin Popescu, Todd Underwood, Earl Zmijewski, "Quaking Tables: The Taiwan Earthquakes and the Internet Routing Table", presentation at NANOG 39, Feb 2007. http://www.nanog.org/mtg-0702/underwood.html

[10] Geoff Huston, "Damping BGP," The ISP Column, Internet Society, June 2007. http://ispcolumn.isoc.org/2007-06/dampbgp.pdf

[11] J. Noel Chiappa, "Endpoints and Endpoint Names: A Proposed Enhancement to the Internet Architecture," unpublished Internet Draft, 1999. http://ana.lcs.mit.edu/~jnc/tech/endpoints.txt

[12] Robert Moskowitz and Pekka Nikander, "Host Identity Protocol (HIP) Architecture," RFC 4423, Internet Engineering Task Force, May 2006. http://www.ietf.org/rfc/rfc4423.txt

[13] Pekka Nikander, "In-depth Look at the Host Identity Protocol (HIP): Providing Agile Mobility, Multi-homing, and Security," NATO IST-070 Lecture Series on "Emerging Wireless Technologies. Oct 2007.

[14] Honggang Zhang, Don Towsley, and Weibo Gong, "TCP Connection Game: A Study on the Selfish Behavior of TCP Users", in Proceedings of 13th IEEE International Conference on Network Protocols (ICNP 2005), Nov 2005.

[15] Bob Braden, Lixia Zhang, Steve Berson, Shai Herzog, and Sugih Jamin, "Resource ReSerVation Protocol (RSVP) — Version 1 Functional Specification," RFC 2205, Internet Engineering Task Force, Sep 1997.

[16] Steven Blake, David L. Black, Mark A. Carlson, Elwyn Davies, Zheng Wang, and Walter Weiss, "An Architecture for Differentiated Services," RFC 2475, Internet Engineering Task Force, Dec 1998.

[17] Bob Briscoe, Arnaud Jacquet, Carla Di Cairano-Gilfedder, Allessandro Salvatori, Andrea Soppera, and Martin Koyabe, "Policing congestion response in an internetwork using re-feedback," in Proceedings of SIGCOMM '05: 2005 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications (Philadelphia, Pennsylvania, USA, August 22 - 26, 2005). ACM Press, New York, NY, 277-288. DOI= http://doi.acm.org/10.1145/1080091.1080124

[18] Daniel J. Solove, "The Digital Person: Technology and Privacy in the Information Age," 288 pages, New York University Press, Feb 2004.

[19] Andrew Odlyzko, "Privacy, Economics, and Price Discrimination on the Internet," in Proceedings of the Fifth International Conference on Electronic Commerce, ed. N.Sadeh (New York: ACM Press, 2003), 355–66. http://www.dtc.umn.edu/~odlyzko/doc/eworld.html

[20] Hari Balakrishnan, Karthik Lakshminarayanan, Sylvia Ratnasamy, Scott Shenker, Ion Stoica, and Michael Walfish, "A layered naming architecture for the internet," in Proceedings of SIGCOMM '04: 2004 Conference on Applications, Technologies, Architectures, and Protocols For Computer Communications (Portland, Oregon, USA, August 30 - September 03, 2004). ACM Press, New York, NY, 343-352. DOI= http://doi.acm.org/10.1145/1015467.1015505

[21] Pekka Nikander, Jari Arkko, and Börje Ohlman, Host Identity Indirection Infrastructure (Hi3)," in Proceedings of The Second Swedish National Computer Networking Workshop 2004 (SNCNW2004), Karlstad University, Karlstad, Sweden, Nov 23-24, 2004.

[22] Angelos D. Keromytis, Vishal Misra and Dan Rubenstein, "SOS: Secure Overlay Services," SIGCOMM Comput. Commun. Rev. 32, 4 (Oct. 2002), 61-72. DOI= http://doi.acm.org/10.1145/964725.633032

[23] Pan Wang, Peng Ning, and Douglas S. Reeves, "A k-anonymous communication protocol for overlay networks," in Proceedings of the 2nd ACM Symposium on information, Computer and Communications Security (Singapore, March 20 - 22, 2007). R. Deng and P. Samarati, Eds. ASIACCS '07. ACM Press, New York, NY, 45-56. http://doi.acm.org/10.1145/1229285.1229296

[24] Yochai Benkler, "The Wealth of Networks — How Social Production Transforms Markets and Freedom," Yale University Press, Apr 17, 2006. http://yalepress.yale.edu/yupbooks/book.asp?isbn=0300110561

[25] Animesh Nandi, Tsuen-Wan Ngan, Atul Singh, Peter Druschel, and Dan Wallach, "Scrivener: Providing Incentives in Cooperative Content Distribution Systems," ACM/IFIP/USENIX 6th International Middleware Conference (Middleware 2005), Grenoble, France, November 2005. http://www.cs.rice.edu/~dwallach/pub/middleware2005.html

[26] Nick Feamster et al, "Virtualisation Revisited — Introduction/Overview," presentation at National Science Foundation Third NeTS FIND PI Meeting, Arlington, VA, June 27–28, 2007. http://www.nets-find.net/ThirdPIJune7/find-june07-sessionsummary.ppt

[27] Nick Feamster, Lixin Gao, and Jennifer Rexford, "CABO: Concurrent Architectures are Better Than One", NSF NeTS FIND project description, http://www.nets-find.net/Cabo.php

[28] Andy Bavier, Nick Feamster, Mark Huang, Larry Peterson, and Jennifer Rexford, "In VINI Veritas: Realistic and Controlled Network Experimentation," In Proceedings of SIGCOMM'06, Pisa, Italy, September 2006.

[29] Jonathan Turner, Patrick Crowley, Sergey Gorinsky and John Lockwood, "An Architecture for a Diversified Internet," NSF NeTS FIND project description, http://www.nets-find.net/DiversifiedInternet.php

[30] Mark Handley, Eddie Kohler, Atanu Ghosh, Orion Hodson, Pavlin Radoslavov, "Designing extensible IP router software," in Proceedings of the 2nd Conference on Symposium on Networked Systems Design & Implementation - Volume 2 (May 02 - 04, 2005). USENIX Association, Berkeley, CA, 14-14.

[31] John Silvester, "National LambdaRail(NLR) — Packet, Wave and Frame Services," presentation at GLIF 2006, Akihabara, Japan, September 11, 2006. http://www.nlr.net/newsroom/pres/glif06-silvester.pdf

[32] James Scott, Pan Hui, Jon Crowcroft, and Christophe Diot, "Haggle: A Networking Architecture Designed Around Mobile Users," Invited paper at The Third Annual IFIP Conference on Wireless On-demand Network Systems and Services (WONS 2006), Les Menuires, France, January 2006.

[33] Joe Touch, "RNA: A Recursive Network Architecture," NSF NeTS FIND project description, http://www.nets-find.net/Rna.php

[34] George Rouskas, Rudra Dutta, Ilia Baldine, Arnold Bragg, and Dan Stevenson, "The SILO Architecture for Services Integration, Control, and Optimization for the Future Internet," NSF NeTS FIND project description, http://www.nets-find.net/Silo.php

[35] John Day, "Patterns in Network Architecture: Rethinking Network Architecture," 592 pages, Prentice-Hall, ISBN 0-132252422, 2007.

[36] Stephen Farrell, Vinny Cahill, "Delay- and Disruption-Tolerant Networking", Artech House Publishers, September 30, 2006, ISDN 1-596930632.

[37] Van Jacobson, "Content Centric Networking: A Self-Organizing Network That Meets Information Needs", Palo Alto Research Center (PARC) project description, http://www.parc.xerox.com/research/projects/networking/contentcentric/default.html

[38] Roy Yates, Dipankar Raychaudhuri, Sanjoy Paul, and James Kurose, "Postcards from the Edge: A Cache-and-Forward Architecture for the Future Internet," NSF NeTS FIND project description, http://www.nets-find.net/Postcards.php

[39] Tilman Wolf, "Service-Centric End-to-End Abstractions for Network Architecture," NSF NeTS FIND project description, http://www.nets-find.net/ServiceCentric.php

[40] Teemu Koponen, Mohit Chawla, Byung-Gon Chun, Andrey Ermolinskiy, Kye Hyun Kim, Scott Shenker, and Ion Stoica, "A Data-Oriented (and Beyond) Network Architecture", in Proceedings of SIGCOMM'07, Kyoto, Japan, August 27–31.

[41] Pekka Nikander, Mikko Särelä, Sasu Tarkoma, and Dirk Trossen, "Publish-Subscribe Inter-Networking (PSIN): A perfect match in the middle", to appear.