MIL-STD-1553 Device Characterization using Organic Interface Functionality

John M. Willis, Robert F. Mills, Logan O. Mailloux, Scott R. Graham
Air Force Institute of Technology
Wright-Patterson Air Force Base, OH 45433

{john.willis}, {robert.mills}, {logan.mailloux}, {scott.graham}@afit.edu

ABSTRACT

Vehicles, aircraft, and satellites typically use control buses for efficient communication among devices and microcontrollers. Examples include the controller area network (CAN) bus in automobiles, the Aeronautical Radio Incorporated 429 (ARINC 429) standard used in commercial aircraft, and MIL-STD-1553 used in military aircraft. These bus standards were developed and implemented roughly 40 years ago when the inherent air-gapped nature of the systems was considered to provide adequate security. Over time, this assumption has proven false with numerous accounts of breaches in air-gapped systems (e.g., Stuxnet). One concern with vehicles and aircraft might be the possibility of rogue devices being installed (perhaps via maintenance technician) which could, in turn, disrupt the proper operation of the vehicle's data bus. Previous research has investigated the use of device fingerprinting via subtle variations in analog waveforms of the MIL-STD-1553 bus, but additional hardware and software were required to collect and process the transmitted messages. In this paper, we explore the feasibility of using the organic capability of commercial-off-the-shelf avionics data bus interface cards to ascertain whether or not a new device has been installed or is masquerading as a different one. Because equipment already present on the aircraft is being used, this additional security feature is readily achievable without impacting size, weight, or power requirements.

1.0 MOTIVATION

Dedicated control buses were developed in the late 1960s and early 1970s when the growth of electronics and computing was making communication and sharing information between devices inefficient [1]. These early buses moved away from the conventional point-to-point wiring and analog signaling for communication and adopted a shared pathway and digital signaling. In the realm of computer networks, as more devices were inter-connected, the need for security became evident as these systems proved to be vulnerable to hackers. In turn, this evolved into a cyclical hardening process where designers would attempt to secure a system and hackers to compromise it. This practice ultimately led to far more protected and resilient computing and networked systems.

However, the control buses within vehicles, aircraft, satellites, and many other cyber-physical systems (CPSs) followed a different path. These systems did not undergo the same hardening process because they were mostly cut-off from adversarial hackers due to their inherent air-gapped design, limited access to components, and innate complexity. These factors led to an inherent-trust methodology where it is assumed that any device communicating on a bus is trustworthy. Many of these early standards were adopted and widely used across multiple industries despite having no methods for ensuring device authentication or data integrity. Stuxnet [2], [3] and attacks against modern vehicles [4] have proven the impregnability assumptions of these of air-gapped systems to be false and exemplified the credibility issue with CPSs [5].

MIL-STD-1553 is one such standard that was developed and widely adopted by the United Stated Department of Defense [6]. Unfortunately, this standard was written without security in mind and is the backbone of many systems that are still used today. Moreover, replacing the bus with a modern, more secure alternative is cost prohibitive. These limitations mean engineers and designers need to find ways to incorporate security and resiliency into legacy systems without significantly impacting cost, performance, or scheduling.

While there are many examples of such processes being studied for the CAN bus [7]–[10], their application is somewhat limited. One study specifically looks at using the periodic nature of the 1553 message frames to develop an intrusion detection system (IDS) [11]. This method operates at the network layer where it uses message addressing and Markov chains to predict the probability that one message address will follow another. This technique is useful in detecting when a rogue device may be inserting false messages during dead-bus times or conducting a denial-of-service attack. However, the technique would not be beneficial in trying to detect a device that is masquerading as another or injecting false data. For example, if a device on the bus was replaced with one that was compromised and communicated at the expected intervals, it is unlikely the technique would detect any malicious activity.

This paper examines using the Radio Frequency Distinct Native Attribute (RF-DNA) fingerprinting process to discriminate between devices and detect rogue devices at the physical layer of communication [12], [13]. Note, this technique has been applied extensively at the Air Force Institute of Technology (AFIT) on wired systems where it is often termed Wired Signal Distinct Native Attribute (WS-DNA) fingerprinting. Unlike conventional WS-DNA, which collects signals using sophisticated oscilloscopes or other system signal collectors, applying this technique to organic 1553 interface cards like those found within operational systems allows the systems to detect rogue devices whether it was communicating at the specified interval or not.

The remainder of this document is organized as follows: Section II provides a brief overview of MIL-STD-1553 and the Alta Data Technologies interface card used for collection. Section III describes the RF-DNA process and how it was used to collect and process the MIL-STD-1553 waveforms. Section IV describes the experiment layout and the scenarios being tested, and Section V gives the results of the study. Section VI discusses the impact and of the results and how they can be used to build more resilient systems. Then Section VII concludes the paper and provides recommendations for further study.

2.0 BACKGROUND

2.1 MIL-STD-1553 Physical Layer

The layout of a MIL-STD-1553 bus consists of three types of interconnected terminals. The bus operates in a master/slave relationship with a bus controller (BC) issuing commands for multiple remote terminals (RTs) and being monitored by a bus monitor (BM) [14]. There is only one BC, but there can be multiple BMs and one to 31 RTs on a single bus. Fig. 1 shows a simple MIL-STD-1553 configuration with a two RTs. The terminals are connected to a bus network topology using transformer couplers and a twisted-pair shielded cable. The standard specifies that the cable must have a characteristic impedance of 70 to 85 ohms, a capacitance of 30 pF/ft, and a cable attenuation of 1.5 dB per 100 feet [15]. Each end of a cable must be terminated with a 78-ohm resistor which is meant to match the cable's characteristic impedance. These resistors minimize the reflections on the cable by simulating an infinite transmission line [16].

The bus has no direct current voltage component when communicating which allows the terminals to be connected to the bus either through direct coupling or transformer coupling [17]. The two coupling methods differ with the use of an additional coupling transformer. The transformer coupling method is far more prevalent with the direct coupling being used only for early 1553 systems. Direct coupling is more susceptible to impedance mismatch and is also limited to one-foot stubs, where the transformer coupling can have stub lengths up to 20 ft [1]. The two coupling methods can be seen in Fig. 2.

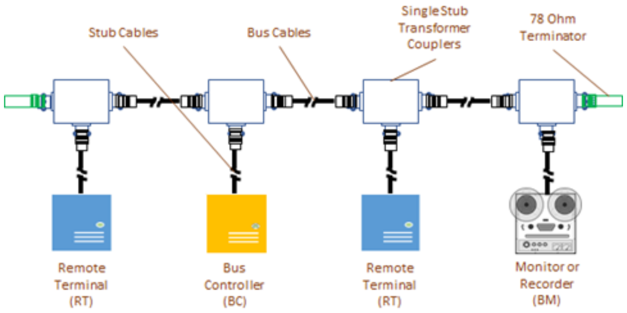


Fig. 1 - Simple Bus Configuration.

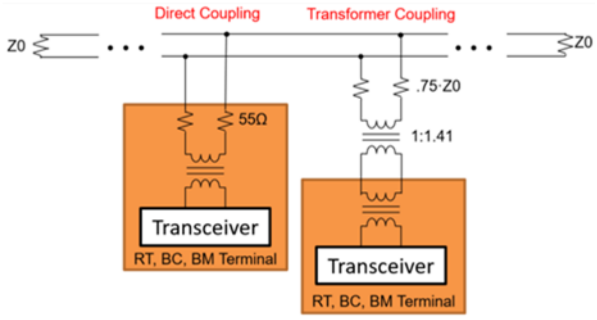


Fig. 2 - Direct and Transformer Coupling.

2.2 MIL-STD-1553 Protocol

MIL-STD-1553 sends bits serially over the bus using pulse code modulation. The protocol encodes bits using a bi-phase Manchester II format, which allows the bus to communicate using high and low voltage levels where the line-to-line or peak-to-peak voltage out of the transmitter is between 18-27 Volts [18]. Manchester encoding represents the bit values (zero and one) by a pulse transition and the polarity of that transition. These transitions occur at the halfway point of the clock cycle for the data bus. The clock operates at 1 MHz for the 1553 standard, so a logical 1 is represented 0.5 µs high pulse followed by a 0.5 µs low pulse, and a logical 0 by the low-to-high transition [18]. The BC and RTs send bits over the bus in frames called words that encompass a three-bit sync, then 16 encoded bits, and a final parity bit for a total of 20 bits. The sync is encoded in three-bit invalid Manchester format that alerts the terminals on the bus that a new word is being sent and allows triggering for the devices to synchronize. The parity bit is the final bit and provides an error check on the words transmitted.

There are three types of words (command, status, and data) that combine to form different messages and transfer data from one device to another. There are a total of 10 distinct message formats, labeled by what the terminal is transmitting and receiving [1]. For example, one of the most common message formats is a BC-RT message that transmits data from the BC to the RT. Other common messages are RT-BC and RT-RT. The other seven combinations involve mode codes, which the protocol uses for error checking and bus operation commands, and the broadcast versions of the combinations already discussed. All messages captured for this experiment were RT-BC messages and triggered on the status word. Fig. 3 displays the format of these messages.

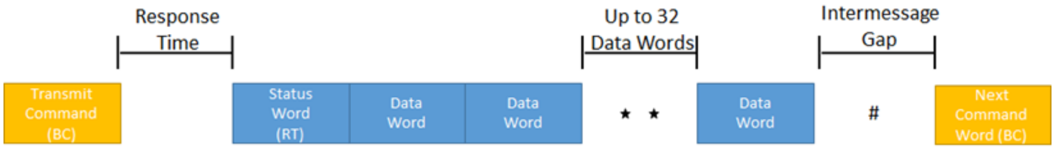


Fig. 3 - RT-BC Message Format.

2.3 Alta Data Technology Cards

The experiments were conducted using an Alta Data Technologies (Alta) interface card to collect message signals for three RT devices. The Alta card is a PCIE4L-1553 inserted into a Windows Desktop and controlled through C++ code and Microsoft Visual Studio. This type of interface card is often used in laboratory environments and can emulate a BC, RT, or BM. The following tests used the Alta card as the BM. The card was set-up to trigger on the start of a status word response to an RT-BC message and to begin collecting bus voltages using its signal capture feature. The signal capture feature allows the card to collect up to 2048 bus voltage samples at a rate of 20 MHz, or 50 ns per sample. The card stores each voltage value in an 8-bit register buffer [19].

The number of samples and size of the Signal Capture Data register present two limitations in collecting the signals. First, each signal capture only collects up to 2048 samples at a rate of 20 Mega-samples/s. Therefore, it can only obtain a total of 102.4 µs of data, which equates to just over five total contiguous words for an RT-BC message, far less than the maximum possible 33-word message response (i.e., one status and 32 data words) sent by each device. The second limitation is that the 8-bit register introduces quantization effects and hard limits on the resolution of the voltage. The signal capture was designed to measure signals between -15 and 15 V. The 256 possible values of the 8-bit register are quantized uniformly, which makes it have a quantile interval of 0.117 V per least-significant bit [20]. For cards with the capability to gather samples faster or with higher voltage resolution, the classification and verification explored in this paper would be better. This relationship is evident in other WS-DNA studies [21] [22] that capture signals using oscilloscopes and more sophisticated signal capturing methods. However, using such equipment requires additional hardware to be added to the platform which in most cases is undesirable or infeasible.

3.0 DATA COLLECTION METHOD

RF-DNA fingerprinting is a physical layer security measure that captures domain specific attributes from devices by computing statistics on their waveforms.

3.1 Region of Interest

The first step in the fingerprinting process is finding a region of interest (ROI). The ROI is the section of the waveform over which the statistical fingerprints will be developed. It is essential that collection is taken over a region that is consistent and shared between all the devices that will be discriminated against. For this study, three different ROIs were chosen from the status word response. The first was the sync portion of the status word and was chosen because it is consistent across all RTs on a bus. The next region selected was the sync portion and the following RT address. This section was chosen because it is the maximum number of bits that are consistent in a transmitted status word for an RT. While this region would not be ideal as a system-wide classifier, on an RT level, it works well to spot a masquerading device. The last region was all 20 bits (20 µs) of the status word. This region is longer and gives more samples to help refine a model, but it is not necessarily stable for each RT. The bits following the RT address in a status word (bits 9-20) are used to indicate an error (bit 9), send notification flags from the system (bits 10-11 and 16-19), and are reserved (bits 12-15). These bits are not guaranteed to be static, but they are typically always zero and used to alert the system of events which need attention. Thus, while monitoring the whole status word may raise false-alarms to intrusion, it will likely correspond to occurrences that require attention anyway. Fig. 4 shows the 20-bit status word for reference.

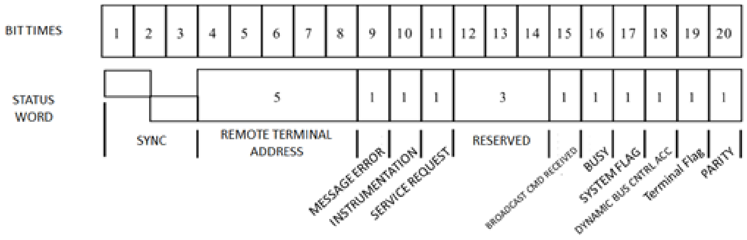


Fig. 4 - MIL-STD-1553 Status Word. Figure modified from [1].

3.2 Sub-Regions

Each of the regions is then further subdivided into sub-regions to gather more representative statistics. The ROIs are divided by three times the number of bit times the region overlaps. The dissection results in nine sub-regions for the sync region, 24 for the RT region, and 60 for the status region. This division helps to isolate the steady portions of the sync bits and to separate the rise and fall transitions of the other bits. Leading samples were also added to give an equivalent number of samples to each region, which results in the sync word having eight samples/sub region, and the RT and status regions have seven samples/sub region. Fig. 5 shows each region and the corresponding sub regions.

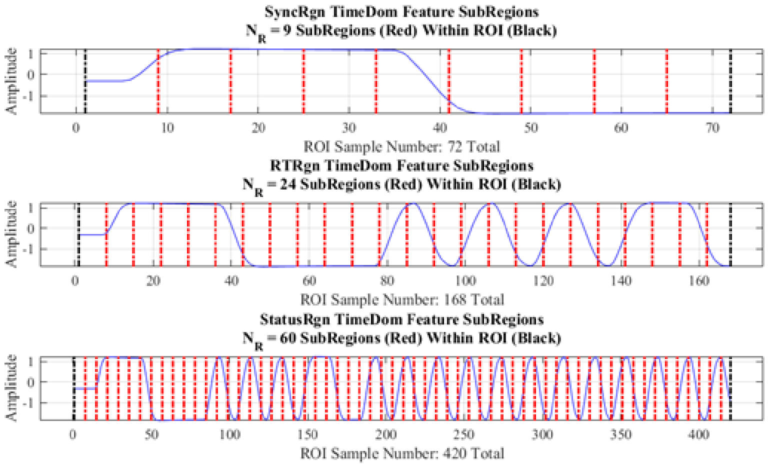


Fig. 5 - ROI and Sub-regions.

3.3 Statistical Fingerprint Calculation

Fingerprints are then generated in the Time Domain (TD) from the signal's instantaneous amplitude $a[n]$, phase $\phi[n]$, and frequency $f[n]$. Each instantaneous feature then has its mean value removed (centered) and then is normalized (division by the maximum value). This process helps to ensure that the waveform is being analyzed rather than the losses from the transmission medium. After centering and normalization, the standard deviation, variance, skewness, and kurtosis are calculated for each sub region to form a fingerprint for the device under test (DUT).

4.0 EXPERIMENT CONFIGURATIONS

Two configurations were used to test the viability of WS DNA to discriminate between devices and detect a rogue device masquerading on the bus. Models were developed using three 1553 devices that were all the same make and model and only differed by serial number. The first configuration, Fig. 6, is the most straightforward bus configuration possible and was used to test the characterization methodology without effects from cabling and other couplers. The second configuration, Fig. 7, was a 150 ft test bus that was used to represent the size of an operational bus as would be found on an aircraft or other vehicle. For every test, a total of 5,000 status word responses were captured.

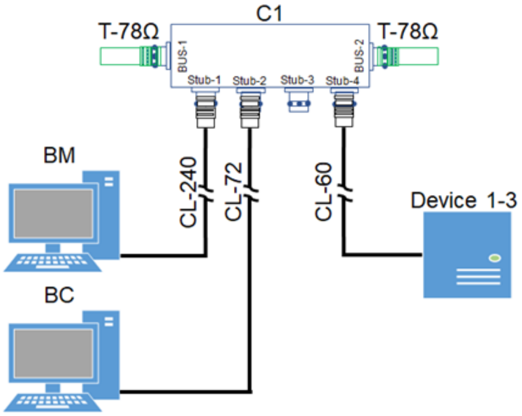


Fig. 6 - Configuration 1.

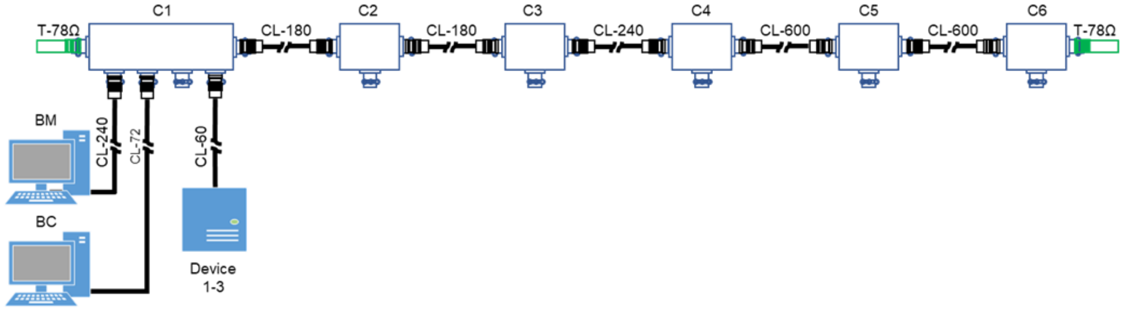


Fig. 7 – Configuration 2.

4.1 Configuration 1 Tests

Two different tests were conducted using configuration one. The first test involved adding a coupler, C2, off of coupler C1 at increasing cable distances. Five scenarios were tested: (1) the classification without C2, (2) the addition of C2 at six inches, (3) 36 inches, (4) 180 inches, and (5) 240 inches. This test was done to see whether the addition of a second coupler had effects on the gathered fingerprints. The configuration for this test can be seen in Fig. 8.

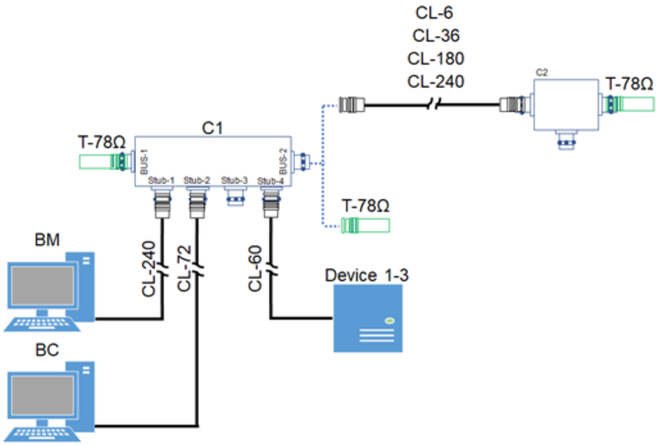


Fig. 8 - Configuration 1, Test 1.

The second test was conducted by taking samples at increasing time intervals to determine whether device heating had any effects on the gathered fingerprints. The equipment layout for this test did not change from Fig. 6. Collections were taken at 15, 30, 45, and 60 minutes during the system's runtime. Each capture took approximately three minutes to collect which gave time to reset the triggers for the collection process.

4.2 Configuration 2 Tests

Three tests were conducted using configuration two. The first test moved a device along the bus, at different coupler distances, to see how well the model developed from the default configuration worked. The test took a collection at each coupler location and analyzed to see whether it would fall within a 90% verification threshold established for the C1 results. Figure 9 shows the arrangement of this test.

The second test used an unmodeled rogue device as the communicating RT. The device was of the same make and model as Devices 1-3 to determine whether the model was sophisticated enough to reject it as a rogue device. Once again, the device was tested at each coupler. Fig. 10 shows the layout of this test.

The final test was to determine whether the loading effects of a rogue bus listener could be detected when an authorized device was communicating. The communicating device remained connected to coupler C1, and a collection was done with the rogue device at each location. The configuration of the final test can be seen in Fig. 11.

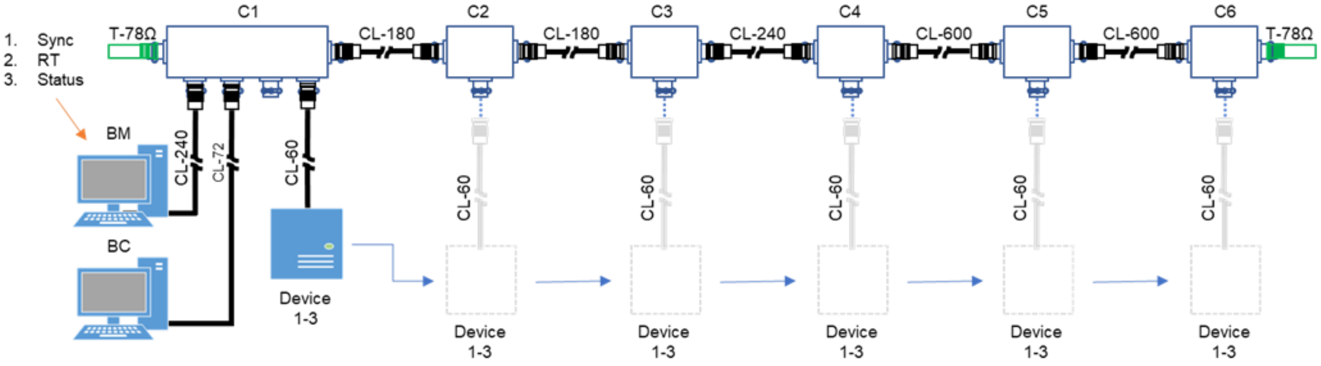


Fig. 9 - Configuration 2, Test 1.

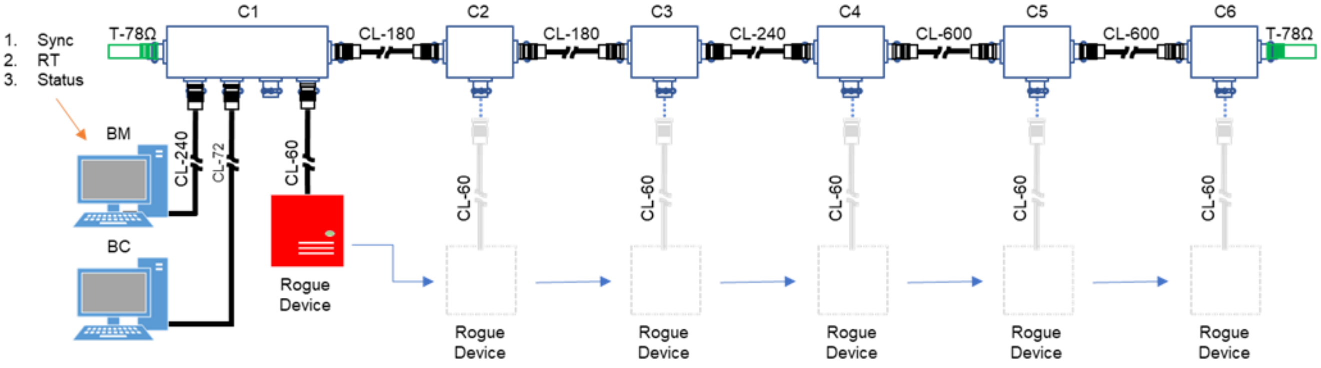


Fig. 10 - Configuration 2, Test 2.

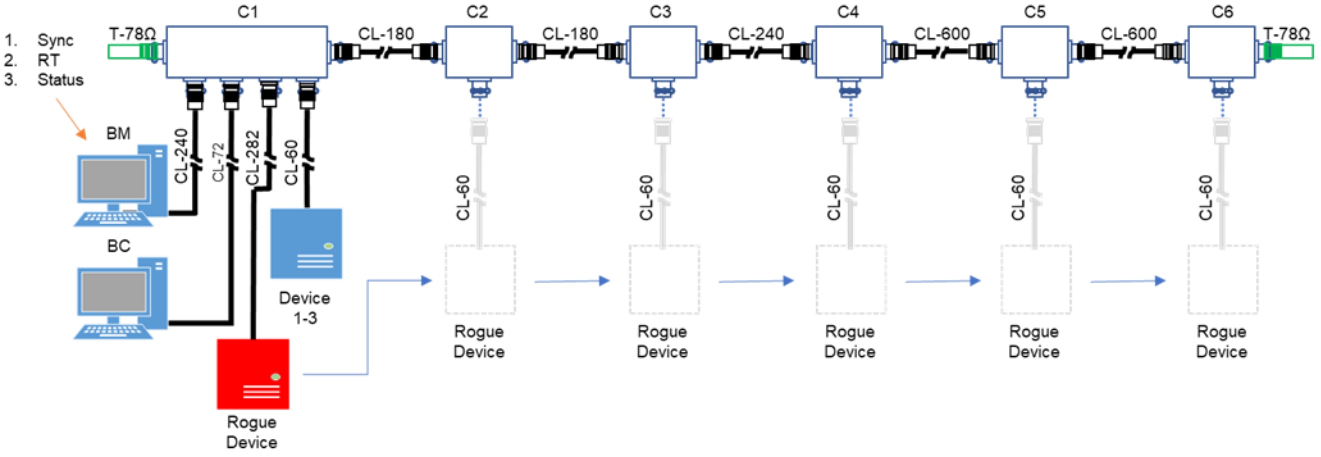


Fig. 11 - Configuration 2, Test 3.

5.0 RESULTS

The results of each test are represented within tables that show either the classification or verification rate (whichever is applicable). The classification tables are displayed as confusion matrices, where the rows represent the instances of the actual class, and each column represents the instances predicted class. The classifier is designed to make a single determination on which device the waveform “looks most like”. This means the rows of each confusion block will add up to 100% (5000 instances). The tested variable in each experiment has its confusion matrix for each value and that is appended unto the others to form the table.

In conventional RF-DNA studies, the established benchmark for a successful classifier is a 90% true-positive rate. Keeping this benchmark, the tables represent any instance that meets this requirement highlighted in green and anything that falls below 90% highlighted yellow. A few instances are highlighted in red, corresponding to large configuration alterations overly degrading the model causing it to incorrectly classify the device more often.

The verification tables have the same layout and represent the actual instances as rows and whether the instance passed an established threshold as columns. The threshold helps determine whether the device is consistent enough to be considered authentic or should be rejected as a rogue device. Within the verification tables, a single waveform instance may pass the threshold values established for zero, one, two, or three devices; this means that unlike the classification tables there is no significance to adding the columns of a row together. Each instance in the acceptance table is just a percentage measure on whether a message sent from the actual device should be considered as authentic for the predicted device. The verification thresholds are based on the maximum Euclidian distance the waveform’s projection can be from the modeled means in the Fisher space. The threshold limits were established so that the training data used in the development of the model would have a 90% acceptance rate. This decision criteria means that it is unlikely that the acceptance rate will ever be much higher than 90% and anything in the 80% range will be considered successful. All true-acceptance measures above 80% will be highlighted in green and measures below in yellow. An instance will only be highlighted red if the false-positive rate for another device exceeds the true-positive rate. For configuration 2, test 2 the highlighting convention will be any values below 50% are green and above red; this is because the objective of these test is to reject situations where there is a rogue device.

5.1 Configuration 1, Test 1

Table 1 shows the classification results for this test. There are some trends that can be ascertained from the results. The first is the classification results are not what would be considered successful in conventional WS-DNA, where it expects a 90% discrimination accuracy; there are a couple of statistics that exceed 90%, but this is not the typical result.

The second trend is that expanding the region of interest had a significant effect on the classifier’s accuracy; moving from the sync to RT ROI there was an average gain of 17.09% and from RT to status a 7.90% gain in true-positive classifications. The last trend identified deals with how the Fisher Projections shift with added cabling. For the RT and status ROIs, the longer the cable added the higher the likelihood that the device is classified as Device 3, whether it is Device 3 or not. Fig. 12 shows the Fisher Projection fingerprint results without any additional cabling (same configuration from which the model was developed) and Fig. 13 shows how adding additional cabling, while using the same model, causes each of the projected fingerprints to shift down and right away from the model’s means for each class. This shifting effect results in a greater number of the fingerprints looking more like Device 3. This shift is a model specific phenomenon and involves how Multiple Discriminant Analysis dimensionally reduces from N_f features to $N_c - 1$ features, where N_c is the number of classes. It is unlikely this relationship would remain consistent if the system were retrained, the system may gravitate towards a different device or the gains from adding more cabling could be increased or reduced.

Table 1: Configuration 1, Test 1 Classification. G – True Positive ≥ 90%; Y – True Positive < 90%; R – False Positive > True Positive.

			Classified As														
			C1 - No Coupler			6 in Cable & C2			36 in Cable & C2			180 in Cable & C2			240 in Cable & C2		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	74.29	19.9	5.81	69.63	25.06	5.31	70.86	23.28	5.86	79.21	16.03	4.76	82.8	7.56	9.64
		RT	83.98	13.6	2.42	79.37	16.33	4.3	77.27	16.45	6.28	73.13	9.62	17.25	52.97	3.82	43.21
		Status	92.02	7.27	0.71	89.47	9.69	0.84	88.92	10.02	1.06	89.06	6.66	4.28	70.8	2.55	26.65
	Dev 2	Sync	24.46	70.2	5.34	18.81	73.5	7.69	20.36	71.14	8.5	39.01	52.76	8.23	44.65	46.34	9.01
		RT	10.84	87.93	1.23	8.62	87.84	3.54	9.47	86.19	4.34	19.68	73.12	7.2	22.68	65.45	11.87
		Status	6.42	93.25	0.33	5.57	93.51	0.92	6.23	92.64	1.13	15.67	81.17	3.16	19.11	74.77	6.12
	Dev 3	Sync	28.81	12.2	58.99	45.99	12.97	41.04	44.84	10.19	44.97	35.96	5	59.04	35.31	5.09	59.6
		RT	13.78	4.12	82.1	12.49	4.18	83.33	9.55	3.15	87.3	3.62	0.99	95.39	3.56	1.03	95.41
		Status	8.42	2.64	88.94	9.07	3.49	87.44	5.98	2.59	91.43	1.57	0.58	97.85	1.28	0.64	98.08

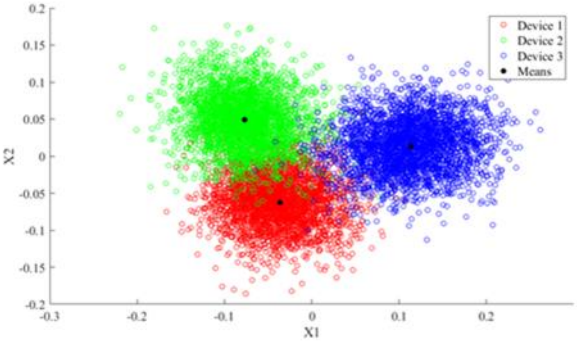


Fig. 12 - Configuration 1, Test 1 No Coupler. Status Region Dimensionally Reduced Fisher Plane.

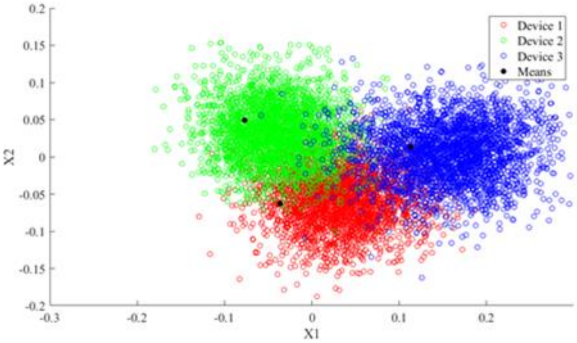


Fig. 13 - Configuration 1, Test 1 Coupler and 240 inch Cable. Status Region Fisher Plane.

The verification results, Table 2, also have some trends that can be analyzed. This table has an analogous trend where more cabling leads to a lower verification rate, no matter the device. There are a few percentages that increase from the 6- to the 36-inch test but not significantly. It appears the addition of the coupler had a greater effect on acceptance than the lengthening of the cable between couplers C1 and C2. Expanding the ROI does not seem to have a clear impact on raising or lower the acceptance rate for true-device acceptance when looked at by acceptance rate on the longest cable tested. Where the expanded ROIs excel is in their ability to more reliably reject the devices that are not communicating. On average the acceptance rate for false devices is 51.53% for the sync region, 18.50% for the RT region, and 5.84% for the status region.

Table 2: Configuration 1, Test 1 Verification. G – True Verification Rate (TVR) ≥ 80%; Y - TVR < 90%; R - False Verification Rate (FVR) > TVR.

			Acceptance Rate														
			C1 - No Coupler			6 in Cable & C2			36 in Cable & C2			180 in Cable & C2			240 in Cable & C2		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	89.59	63.95	50.16	84.35	61.77	39.46	84.29	61.28	41.45	82.67	48.8	38.71	80.64	36.92	43.86
		RT	89.58	31.43	7	89.05	36.47	11.28	87.31	35.57	14.61	80.21	22.84	30.47	56.03	10.07	52.85
		Status	83.12	10.23	0.32	85.83	14.09	0.47	84.34	14.84	0.67	81.98	10.49	2.61	53.72	3.69	17.26
	Dev 2	Sync	57.9	88.12	44.42	47.64	82.27	35.22	49.47	83.17	37.58	71.19	83.84	51	76.29	81.15	55.16
		RT	30.1	88.85	2.56	25.02	85.18	5.25	26.6	84.88	6.2	42.76	81.71	12.94	46.11	75.3	20.14
		Status	10.42	85.95	0.18	9.09	85.28	0.42	10.33	85.73	0.52	21.18	78.41	1.43	23.59	70.51	3.18
	Dev 3	Sync	60.25	52.24	90.31	76.35	54.93	83.23	73.22	48.89	83.32	56.78	30.36	80.13	52.93	27.72	76.7
		RT	20.84	6.42	87.12	18.88	6.37	87.48	14.73	4.34	87.81	5.53	1.31	73.36	5.08	1.3	66.19
		Status	4.97	1.22	80.8	5.65	1.72	79.06	3.51	1.13	82.22	0.91	0.2	75.1	0.66	0.22	67.83

5.2 Configuration 1, Test 2

The classification results seen in Table 3 exhibit many of the same trends discussed with Test 1—specifically, the increased classification rate as the ROIs were expanded and a shifting of the Fisher Projections (this time in Device 2's favor). RT, emulator card, and component heating are the likely causes of this shift. A temperature probe took readings of the outside of the three RT devices during the test, and they fluctuated from 73.7 °F to 93.2 °F.

Table 3: Configuration 1, Test 2 Classification. G – True Positive ≥ 90%; Y – True Positive < 90%; R – False Positive > True Positive.

			Classified As														
			0 min			15 min			30 min			45 min			60 min		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	61.7	25.53	12.77	60.9	27.04	12.06	59.21	29.53	11.26	59.4	31.27	9.33	59.8	31.62	8.58
		RT	79.25	14.47	6.28	77.09	18.78	4.13	75.21	21.4	3.39	73.47	23.56	2.97	73.16	24.53	2.31
		Status	89.85	7.01	3.14	87.11	10.46	2.43	86.12	11.76	2.12	84.93	13.57	1.5	84.19	14.64	1.17
	Dev 2	Sync	18.75	74.49	6.76	16.86	76.45	6.69	15.89	78.34	5.77	14.84	79.25	5.91	12.51	83.02	4.47
		RT	10.25	88.2	1.55	8.62	89.38	2	7.75	90.54	1.71	7.31	91.04	1.65	6.19	92.48	1.33
		Status	5.43	93.85	0.72	5.72	93.22	1.06	4.85	94.47	0.68	5.3	93.96	0.74	3.85	95.7	0.45
	Dev 3	Sync	14.26	8.08	77.66	18.47	11.25	70.28	21.15	12.11	66.74	23.34	13.85	62.81	23.5	14.64	61.86
		RT	8.86	2	89.14	16.41	3.63	79.96	19.31	4.57	76.12	22.54	6.01	71.45	23.88	6.7	69.42
		Status	5.31	1	93.69	11.62	2.85	85.53	14.67	3.74	81.59	18.34	4.73	76.93	19.52	5.85	74.63

The verification results, Table 4, were slightly better than those reported in Test 1. The average true-acceptance rate in Test 2 was 84.82%, compared to Test 1's 81.19%. Both tests had similar false acceptance rates of 25.83% and 25.29%. These measures signify that the WS-DNA model is more robust against temperature variation than configuration change. As in configuration 1, test 1, expanding the ROI lowered the false-acceptance rate but unlike test 1, in test 2 it also lowered the true-acceptance rate in each case.

Table 4: Configuration 1, Test 2 Verification. G – TVR ≥ 80%; Y - TVR < 90%; R - FVR > TVR.

			Acceptance Rate														
			0 min			15 min			30 min			45 min			60 min		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	89.05	76.28	45.41	88.22	76.23	43.75	87.01	77.58	41.57	86.44	78.8	37.87	86.05	78.2	35.3
		RT	88.73	38.49	12.31	87.84	43.69	8.85	86.36	46.09	7.17	85.59	48.07	6.28	84.09	48.36	5.33
		Status	86.4	12.69	2.85	83.68	17.38	2.11	82.7	18.34	1.91	81.73	20.16	1.41	80.07	21.25	1.13
	Dev 2	Sync	61.99	90.94	25.86	56.7	87.58	24.47	53.41	86.09	20.53	51.59	84.82	20.46	47.47	83.94	15.39
		RT	36.26	89.59	2.24	31.92	87	2.3	28	85.76	2.08	27.5	83.65	1.86	24.58	83.98	1.4
		Status	12.03	87.95	0.3	11.37	85.22	0.49	10.07	84.18	0.35	10.82	84.24	0.32	7.78	82.98	0.21
	Dev 3	Sync	45.47	36.2	89.86	55.31	44.81	90.08	58.73	48.77	89.77	62.44	52.8	88.75	62.7	53.23	87.12
		RT	14.73	3.6	88.52	26.05	6.77	85.88	30.44	9.16	83.5	34.29	11.61	80.08	35.78	12.43	78.87
		Status	4.48	0.64	85.73	10.09	1.73	79.82	12.6	2.83	76.14	16.52	3.8	71.74	17.61	4.6	68.97

5.3 Configuration 2, Test 1

This test was done to study the robustness of the model when the RT was measured off different couplers, using the model developed off C1. The classification and verification results, Table 5 and 6, show that moving the RT has a far more prominent effect than either test conducted in configuration one. The Fisher plane shift favors Device 2 to the degree that any measurement taken after C2 will be classified as Device 2. Even at C2 the model begins to degrade with only one instance of Device 1 and 3 being classified correctly. The verification results also show a significant effect on the model's ability to verify devices that traverse along a bus. Any ability to reliably authenticate the device disappears after the second coupler.

Table 5: Configuration 2, Test 1 Classification. G – True Positive ≥ 90%; Y – True Positive < 90%; R – False Positive > True Positive.

			Classified As																	
			C1			C2			C3			C4			C5			C6		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	63	13.9	23.1	58.4	23.1	18.5	44.6	55.4	0	5.7	94.1	0.1	0	91.6	8.4	0	68.8	31.2
		RT	78.9	7	14.1	0.6	99.4	0	0	100	0	0	100	0	0	100	0	0	100	0
		Status	90.3	2.1	7.6	0	100	0	0	100	0	0	100	0	0	88.5	11.5	0	90.1	9.9
	Dev 2	Sync	21.7	66.7	11.6	34.1	47	18.9	16.8	83.2	0	1.9	97.9	0.3	0	93.2	6.8	0	78.5	21.5
		RT	16.1	82.6	1.3	0	100	0	0	100	0	0	100	0	0	100	0	0	100	0
		Status	8.1	91.6	0.3	0	100	0	0	100	0	0	100	0	0	95.8	4.3	0	94	6
	Dev 3	Sync	11.8	3.6	84.6	34.5	34	31.5	43.1	56.8	0.1	1.5	98.5	0	0	82.5	17.5	0	38.4	61.6
		RT	5.7	0.2	94.1	3.1	96.4	0.5	0	100	0	0	100	0	0	100	0	0	100	0
		Status	1.3	0	98.7	1.3	98.7	0	0	100	0	0	100	0	0	61.7	38.3	0	84.1	15.9

Table 6: Configuration 2, Test 1 Verification. G – TVR ≥ 80%; Y - TVR < 90%; R - FVR > TVR.

			Acceptance Rate																	
			C1			C2			C3			C4			C5			C6		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	87.1	52.7	56.7	70.4	48.2	37.2	2.1	2.3	0.1	1.4	5.5	0.3	0	0	0	0	0	0
		RT	87.8	16.8	23.6	1.5	30.	0	0	0	0	0	0	0	0	0	0	0	0	0
		Status	80.5	1.6	4.2	0	1.8	0	0	0	0	0	0	0	0	0	0	0	0	0
	Dev 2	Sync	60.9	88.4	31.4	63.5	71.6	36.8	1.9	5.3	0.1	1.3	7.7	0.3	0	0	0	0	0	0
		RT	33	86.8	1.6	0	5.9	0	0	0	0	0	0	0	0	0	0	0	0	0
		Status	5.8	79.7	0	0	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0
	Dev 3	Sync	41.3	18.5	87.8	61.	56.4	47.6	11.	12.9	0.8	0.6	4.8	0.1	0	0	0	0	0	0
		RT	13	0.3	87.6	7.5	65.9	0.3	0	0.1	0	0	0	0	0	0	0	0	0	0
		Status	0.6	0	82.7	0.9	53.3	0	0	0.2	0	0	0	0	0	0	0	0	0	0

5.4 Configuration 2, Test 2

This test was done to study the ability of the system to reject (not accept) a rogue device as valid. Only the verification results table was included (Table 7) because inclusion of the classification results would be non-informative and counter intuitive. Any incident value below 50% was decided to be the benchmark for success of this test, and the system performed well with only the sync and RT ROIs off C1 exceeding the established benchmark. Also, as happened in the previous test, all the verification rates after C2 were extremely small. While this may not be a desirable result for authenticating a moving device, it is for detecting a rogue one.

Table 7: Configuration 2, Test 2 Verification. G – TVR < 50%; R - FVR ≥ 50%.

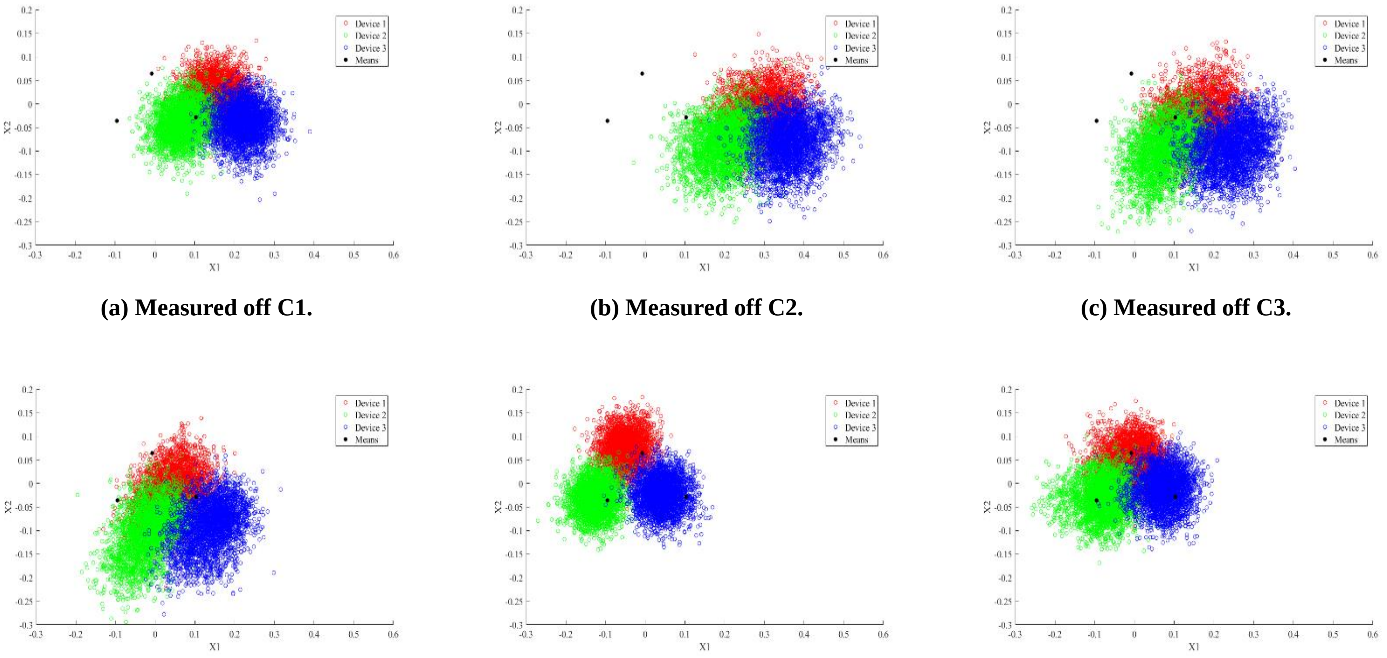
			Acceptance Rate																	
			C1			C2			C3			C4			C5			C6		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Act Dev	Dev 4	Sync	34.5	83	26.9	48.3	44.8	12.5	1.3	10	0.5	1.5	9.6	2.8	0	0	0	0	0	0
		RT	37.1	63	12.3	0.2	9.9	0	0	0	0	0	0	0	0	0	0	0	0	0
		Status	31.7	22.2	8.4	0	6.4	0	0	0	0	0	0	0	0	0	0	0	0	0

5.5 Configuration 2, Test 3

The final test was to study the impact a rogue listening device would have on the system. The bus monitor acts as a passive element on the system. Its effect on the system will be from the loading at each coupler and how this affects the transmission characteristics of the waveform along the bus. The classification results in Table 8 show an interesting result where the classification appears to get worse as the listener is moved to C2 but after that improves with measurements taken off C5 correctly classifying each of the devices. The final collection off C6 then reversed the trend again producing lowered measured values. Fig. 14 shows how the results deviate from the model means for the middle couplers and then rebounds for the last two. This effect could be a result of either how the model dimensionally reduces or the rogue listener having less disturbance towards the bus edges.

Table 8: Configuration 2, Test 3 Classification. G– True Positive ≥ 90%; Y – True Positive < 90%; R – False Positive > True Positive.

			Classified As																	
			C1 - Listener			C2 - Listener			C3-Listener			C4 - Listener			C5 - Listener			C6 - Listener		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	26.7	4.8	68.5	0	2.7	97.3	0	15.6	84.5	0.3	43.	56.7	71.8	25.6	2.6	61.1	36.7	2.2
		RT	0.8	0	99.2	0	0	100	0	0	100	0.3	0.5	99.2	67.7	32.1	0.2	59.9	39.7	0.4
		Status	3.2	0	96.8	0	0	100	1.3	0.2	98.5	26.9	16.7	56.4	91.5	8.5	0.1	87	8.5	4.6
	Dev 2	Sync	9.6	39.2	51.3	0	16.8	83.2	0	49.4	50.6	0	78.2	21.9	11.6	86.9	1.5	14.4	84	1.5
		RT	4.7	4.5	90.9	0	0	100	0	0.6	99.4	0.2	10.8	89.1	1.4	98.6	0	2.2	97.8	0
		Status	5.8	4.3	89.9	0	0.1	99.9	0.6	10.2	89.2	3.5	64.2	32.3	0.7	99.3	0	11.7	86.8	1.5
	Dev 3	Sync	2.3	0.6	97.1	0	0.8	99.2	0	2.8	97.2	0	15.	85.1	30	18.7	51.2	40	30.7	29.3
		RT	0	0	100	0	0	100	0	0	100	0	0	100	38.5	10	51.5	54.	15.8	30.2
		Status	0	0	100	0	0	100	0	0	100	0	0.7	99.3	30.4	8.8	60.7	16.9	2.7	80.4



(a) Measured off C1.

(b) Measured off C2.

(c) Measured off C3.

(d) Measured off C4.

(e) Measured off C5.

(f) Measured off C6.

Fig. 14 - Configuration 2, Test 3 Rogue Listener. Status Region Dimensionally Reduced Fisher Plane.

The verification results (Table 9) had a similar phenomenon, which like the previous test showed low acceptance rates for any incident off the couplers following C1. The results also start to peak at C5 and reverse and begin to reverse at C6.

Table 9: Configuration 2, Test 3 Verification. G – TVR ≥ 80%; Y - TVR < 90%; R - FVR > TVR.

			Acceptance Rate																	
			C1 - Listener			C2 - Listener			C3-Listener			C4 - Listener			C5 - Listener			C6 - Listener		
			Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3	Dev 1	Dev 2	Dev 3
Actual Device	Dev 1	Sync	54.1	21.3	87.7	0.2	0.3	18.8	0.2	1.1	11.4	3.8	20.7	32.1	81.9	55.7	22.5	76.4	63.1	19.7
		RT	1.7	0	53.7	0	0	0	0	0	3.9	0.9	0.1	56.8	64.5	32.4	0.9	58.7	36.4	1.3
		Status	0.5	0	30.4	0	0	0.6	0.4	0.01	41.9	15.3	4.9	33.	61.2	3.4	0	70.7	4.5	2
	Dev 2	Sync	38.5	58.9	74.7	0	0.3	4.8	0.	0.2	1.3	0.1	4.9	3.2	32.1	83.7	9.9	32.2	79.2	10.5
		RT	8.7	1.4	81.1	0	0	0.7	0	0	5.8	0.4	0.4	30.8	3.8	68.8	0	5.1	66.8	0
		Status	2	0.4	65.	0	0	12.8	0.2	0.3	33.8	1.1	8.8	8.5	0.4	71.	0	7.5	63.3	0.2
	Dev 3	Sync	10.4	3.5	69.9	0	0	2.7	0	0.02	1.4	0.1	1.7	7.7	63.7	50.7	82.1	74.6	66.2	68.5
		RT	0	0	6	0	0	0	0	0	0	0	0	2.2	50.7	11.9	59.7	67.8	20.7	40.2
		Status	0	0	5.5	0	0	0	0	0	2.6	0	0	28.8	16.5	1.9	37.02	8.6	0.6	59.8

6.0 DISCUSSION

There is currently no capability on MIL-STD-1553 systems for authenticating a device that is communicating. Considering this lack of protection, the classification and verification capabilities discussed are promising. Additionally, there is the benefit that this capability can likely be implemented without additional equipment, through coding and have the models updated without the need for extra oscilloscopes and collection devices.

Also, while some of the verification numbers are low, the perceived weakness can be used to help detect certain types of attacks. The most apparent example deals with the lower verification rate in the tests where the bus was modified, or the rogue BM attached. The model fails to correctly classify/verify the device when large changes are made to the configurations; however, correct identification is not a desirable feature with aircraft security where alteration to the physical bus should raise alarm. Additionally, while the model seems susceptible to the loading effects of adding on couplers or other devices, this feature is ideally suited for detecting a rogue listening device that could be used to exfiltrate data from a system.

7.0 CONCLUSION AND FUTURE WORK

These results proved promising and encourage further study. To expand on this research and provide a more representative real system many improvements can be made. The addition of an out-of-class option could help combat the Fisher Plane shifts by characterize anything that shifts too far beyond the training means as an unrecognized device. The configuration of the tests and their execution are part of the reason for the small variation in the waveforms. In a true 1553 system, the devices will be modeled off different couplers, likely have a fixed location on the bus, and probably be of different makes and models. These facts are likely significant enough to give enough separation in the classification model to significantly increase the classification accuracy obtained. To the final point on the slower sampling and lower resolution used to take the measurements, any additional capacity available will make the model better and more reliable.

Each of the improvements discussed could expand on this work. It would also be useful to further explore if the effects of loading and timing could be modeled to predict what the fingerprints should look like in different situations. Lastly, to be genuinely representative of a real-world bus it would be useful to expand the test to multiple RTs on a single bus and supply it with real-operational data.

8.0 DISCLAIMER

The views expressed in this paper are those of the authors and do not reflect the official policy or position of the United States Air Force, the U.S. DoD, or the U.S. Government.

9.0 REFERENCES

[1] Condor Engineering Inc, “MIL-STD-1553 Tutorial,” 2000.

[2] A. Nourian and S. Madnick, “A Systems Theoretic Approach to the Security Threats in Cyber Physical Systems Applied to Stuxnet Arash,” IEEE Trans. Dependable Secur. Comput., 2015.

[3] D. Kushner, “The Real Story of Stuxnet,” IEEE Spectrum, 2013. [Online]. Available: http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet. [Accessed: 13-Jun-2017].

[4] S. Checkoway et al., “Comprehensive Experimental Analyses of Automotive Attack Surfaces,” USENIX Secur., 2011.

[5] Y. Liu, Y. Peng, B. Wang, S. Yao, and Z. Liu, “Review on cyber-physical systems,” IEEE/CAA J. Autom. Sin., vol. 4, no. 1, pp. 27–40, 2017.

[6] Department of Defense, “MIL-STD-1553B 22Jan79.pdf.” 1979.

[7] K.-T. Cho and K. G. Shin, “Fingerprinting Electronic Control Units for Vehicle Intrusion Detection Fingerprinting Electronic Control Units for Vehicle Intrusion Detection,” 25th USENIX Secur. Symp. is, 2016.

[8] M. Enev, A. Takakuwa, K. Koscher, and T. Kohno, “Automobile Driver Fingerprinting,” Proc. Priv. Enhancing Technol., vol. 2016, no. 1, pp. 34–51.

[9] A. Taylor, S. Leblanc, and N. Japkowicz, “Anomaly Detection in Automobile Control Network Data with Long Short-Term Memory Networks.”

[10] A. Taylor, N. Japkowicz, and S. Leblanc, “Frequency-based anomaly detection for the automotive CAN bus,” in 2015 World Congress on Industrial Control Systems Security, WCICSS 2015, 2016.

[11] O. Stan, Y. Elovici, A. Shabtai, G. Shugol, R. Tikochinski, and S. Kur, “Protecting Military Avionics Platforms from Attacks on MIL-STD-1553 Communication Bus.”

[12] C. M. Talbot, M. A. Temple, T. J. Carbino, and J. A. Betances, “Detecting rogue attacks on commercial wireless insteon home automation systems,” no. 17, 2017.

[13] T. J. Bihl, K. W. Bauer, and M. A. Temple, “Feature Selection for RF Fingerprinting with Multiple Discriminant Analysis and Using ZigBee Device Emissions,” IEEE Trans. Inf. Forensics Secur., 2016.

[14] Alta Data Technologies LLC, “MIL-STD-1553 Tutorial and Reference,” 2014.

[15] Department of Defense, “MIL-STD-1553B: Interface Std for Digital Time Division Command/Response Multiplex Data Bus,” 1996.

[16] D. K. (David K. Cheng, Field and wave electromagnetics. Addison-Wesley, 1989.

[17] N. H. Modi, J. R. Armstrong, J. G. Tront, and M. Z. Khan, “Modeling and simulation of 1553 bus for upset tolerance experiments,” in Seventh Annual International Phoenix Conference on Computers an Communications. 1988 Conference Proceedings, 1998, pp. 131–135.

[18] Department of Defense, “MIL-HDBK-1553A Multiplex Applications Handbook,” 1988.

[19] Alta Data Technologies LLC, “AltaAPI Software User’s Manual,” 2016.

[20] B. Sklar, Digital Communications: Fundamentals and Applications. Prentice-Hall PTR, 2001.

[21] J. Lopez, N. C. Liefer, R. Busho, Colin, and M. A. Temple, “Enhancing Critical Infrastructure and Key Resources (CIKR) Level-0 Physical Process Security Using Field Device Distinct Native Attribute Features,” IEEE Trans. Inf. FORENSICS Secur.

[22] J. Lopez, M. A. Temple, and B. E. Mullins, “Exploitation of HART Wired Signal Distinct Native Attribute (WS-DNA) Features to Verify Field Device Identity and Infer Operating State .,” 9th Int. Conf. Crit. Inf. Infrastructures Secur., 2014.