USA Digital Forensics Information Intelligence

(DFII)

A National Science Foundation / University Cooperative Research Center

Industry Advisory Board Meeting

DFII Site Projects

August 13-15, 2017

Florida International University

Miami, FL

In the NSF Center for Advanced Research in Forensic Science (CARFS)

CARFS is a collaboration between Florida International University and the University of South Alabama

INTRODUCTION

As a Site in the Center for Advanced Forensic Science Research, our focus will be Digital Forensics Information Intelligence (DFII), which, broadly defined, is the development, testing, and implementation of novel approaches to understand not only how devices, information systems, and software can be compromised, but also how one can reliably determine how those compromises occurred. The digital forensics field encompasses a broad environment. This is due to society's increasing dependence on and amalgamation of new technology into all aspects of life, including a growing Internet of Things, information security is now critical to virtually every industry and sector of our economy, including automotive, healthcare, supply chains, and national defense. Previous publications highlight the fact that not all forensic tools are equal in their ability to verifiably retrieve data from both traditional and mobile devices. This problem is complicated with the ever increasing evolution of technology.

From a research perspective, Digital Forensics Information Intelligence (DFII) involves development, experimentation and testing of novel solutions that hinder and assist real-world digital investigations. Hence, to acquire a more in-depth understanding of how new technology stores data and how this technology can be abused, empirical studies need to be developed with industrial partners to test code development, data alterations, modifications and system corruption for malicious purposes. Then these systems and software artifacts must be analyzed to acquire an understanding of the effectiveness of proposed solutions in new and hostile environments. The proposed I/UCRC Site at USA strives to implement functional, cross-cutting strategies to support undergraduate, postgraduate thesis and doctoral research objectives.

CENTER OVERVIEW

The two research sites that constitute the CARFS are the University of South Alabama (USA) and Florida International University (FIU). The Center for Advanced Research and Forensic Sciences is a cooperative research center. A center must consist of a minimum of two universities. Each of the sites at USA and FIU are subsidized by the National Science Foundation through funds provided by the National Institute of Justice and the Department of Justice. This provides a benefit to industry advisory board members in that all of the money they provide goes directly to the researchers without the normal rates of overhead and other costs normally associated with research projects. The structure of an IUCRC encourages collaboration between industry advisory board members due to the competitive nature of the research.

Moreover, additional valuable collaborative opportunities are created because the IUCRC allows other non-NSF funded universities to participate within a research site. While these universities do not have their own research sites within the Center, their researchers are allowed to participate in research opportunities with IAB members under specific circumstances. For example, FIU has allowed George Washington University and Northeastern University to affiliate with their site, and South Alabama enjoys a mutually beneficial affiliate relationship with Texas A&M University.

Research Focus:

The proposed I/UCRC Site at USA strives to implement three functional, cross-cutting strategies to support undergraduate, postgraduate thesis and doctoral research objectives. These strategies are each applicable to a range of industries and governmental interest.

- **Malicious Software Analysis:** Support for software development and analysis of malware, rootkits and viruses in virtual environments.
- **Technology Evaluation:** Evaluation of current data extraction capabilities against new technologies.
- **Detection and Exploitation:** Development of code designed to aid in reverse engineering device activities and increasing residual data extraction capabilities on new technologies in order to understand the impact of malicious alterations and test detection mechanisms.

Broad Research Goal

Conduct fundamental real-world relevant research to further the body of knowledge in digital forensic science through malicious software analysis, technology evaluation, and detection and exploitation.

ADMINISTRATIVE STRUCTURE

Two University Sites

University of South Alabama
Shelby Hall,
150 Jaguar Drive
Mobile, Al. 36688-7274

Florida International University
Modesto A. Maidique Campus
Miami, Fl., 33199

CENTER CONTACTS

USA

Alec Yasinsac, Ph.D.

Professor
Dean of the School of Computing
251.460.6390
yasinsac@usouthal.edu

Les Barnet

Director Center for Forensics, Information Technology and Security
School of Computing
251-461-1601
hlbarnett@southalabama.edu

Brad Glisson, Ph.D.

Chief Research Scientist Digital Forensics Information Intelligence (DFII) Site
Associate Professor
School of Computing
251.460.7634
bglisson@southalabama.edu

FIU

Jose R Almirall, Ph.D.

Professor
Department of Chemistry and Biochemistry and Director International Forensic Research Institute
305.348.3917
almirall@fiu.edu

Executive Summary for Proposed Projects

Investigation of Smartphone Residual Data in Secondary Markets

PROJECT ID | DFII-5 TYPE | [X] New [ ] Continuing START DATE | January, 2018

PROJECT LEAD/PARTICIPANTS | Student; Dr. Brad Glisson; Dr. Todd McDonald

DESCRIPTION | This research seeks to identify trends in residual data in secondary markets, the performance capabilities of industry accepted tool-kits (Cellebrite, XRY, FTK Phone Examiner Plus) along with an analysis of the impact of improved storage capacities, garbage collection, and reset capabilities on newer devices. It is increasingly important for organizations and law enforcement to understand the residual data that can be gathered from mobile devices in terms of intellectual property leakage, residual data retention from social media apps and residual GPS data. According to the International Data Corporation (IDC), Android is the dominate OS with, approximately, 82% of the market. Hence, this research will also investigate the effeteness of the top three remote deletion apps available on Google Play.

EXPERIMENTAL PLAN | During the project period the team will work to:

- Purchase a range of smartphones from secondary markets, i.e. pawn shops and the Internet.
- Implement forensic analysis tools to extract data from smartphones purchased from secondary markets.
- Hard-rest acquired devices. Implement forensically analysis tools to extract data.
- Scrutinize the extraction results from the smartphones and the tools (XRY, and FTK Phone Examiner Plus).
- Download the top three remote data deletion apps for the Android OS
- Insert a defined dataset on-to the device & run the remote data deletion app
- Implement forensic analysis tools to extract data.

RELATED WORK | The foundation of this research is based on Glisson's real-world experience with collecting data from secondary markets and corporate environments using Cellebrite and XRY mobile device tool kits. While research has been conducted in the past on residual data retention by Glisson, the majority of the data collected several years ago in that experiment originated from relatively low-end featureless mobile devices. The research in this space would greatly benefit from repeated execution cycles with smartphones from multiple manufactures, secondary market analysis (devices purchased from eBay vs. local pawn shops), and up-dated mobile device forensics toolkits.

HOW OURS IS DIFFERENT | Initial secondary market research was performed by Glisson in 2010/2011. Additional research by Glisson investigated mobile devices in a Global Fortune 500 financial organization in 2012/2013. The later research utilized a single forensic toolkit. Both investigations worked with relatively low end mobile devices. Since then, mobile phones have increased substantially in storage capability, network connectivity and processing power. The continued evolution of these devices along with extraction toolkit capabilities generates new questions in terms of residual data retention, the effectiveness of hard-resets and remote deletion applications.

MILESTONES FOR YEAR |

4 months: Configure test environments, acquire smartphones from secondary markets and perform the initial processing. Investigate initial results from the perspective of the smartphones and the extraction tools.

8 months: Develop the data sets and execute the remote data deletion experiments on an Android mobile device. Examine the effectiveness of hard-resets on data extraction capabilities.

12 months: Compare these data deletion app results with the results from the hard-rest and the initial data extraction. Investigate tool performance results between the data deletion experiment, hard-rest and the initial data extraction.

DELIVERABLES | Detailed reports analyzing:

- General categories, data types and apps along with Logical & physical tool extraction performance
- The effectiveness of hard resets from a data extraction and tool performance perspective.
- The effectiveness of remote data deletion apps on an Android smartphone.

BUDGET FOR YEAR |

	1 Ph.D.
Students	31,818
Equipment	-
Travel	-
Overhead (10%)	3,182
Total	**35,000**

NECESSARY EXPERTISE | Coding Skills, Android OS and Digital Forensic Knowledge

ECONOMICS | With the proliferation of mobile devices in corporate and legal environments. It is critical for companies to have an understanding of the residual data resident on devices in order to mitigate data leakage risk. From a law enforcement perspective, it is helpful for a digital analyst to know the limitations of mobile device forensics toolkits. Both industry and law enforcement benefit from an understanding of the effectiveness of remote deletion applications and hard-rest activities.

POTENTIAL MEMBER COMPANY BENEFITS | This project provides immediate insight to both IAB members and their clients by detailing the type of residual data that is resident on smartphones, the effectiveness of digital forensics toolkits, and the viability of remote deletion applications. In the long run, this project identifies future research areas in residual data mediation solutions.

PROGRESS TO DATE | Previous research by Glisson investigated residual data resident on mobile devices in a Global Fortune 500 Financial Organization and in secondary markets. Both investigations focused on relatively featureless mobile devices.

KNOWLEDGE TRANSFER TARGET DATE | 12 months

Visual Analytics for Cloud Ecosystems

PROJECT ID | DFII-7	**TYPE |** [X] New [ ] Continuing	**START DATE |** January, 2018

PROJECT LEAD/PARTICIPANTS | Student; Dr. Jordan Shropshire; Dr. Ryan Benton

DESCRIPTION | This research develops a new, out-of-band model for monitoring the integrity of virtual machines. It uses visual analytics to identify malware embedded within guest operating systems, files, and software. The proposed model not only works with virtual machines, but also with containers and unikernels. The proposed approach renders a two-dimensional, colored depiction of each guest's disk image. The depictions are analyzed using a pattern recognition algorithm. The pattern recognition algorithm is trained to parse the depictions and identify individual files and software components. The detection process focuses on identifying elements which do not appear as expected. Three visual detection methods are proposed: (1) Anomaly detection: Compare each file or software component visualization against a trusted depiction of the same element in order to identify anomalies such as modifications, deletions, or additions to binary files. (2) Rule-based detection: Depictions of file or software components are compared against a rule set designed to flag signs of concealed malware such as compressed or encrypted data within the contents of certain files. (3) Signature-based detection: Compare virtual machine disk image depictions against a database of visualizations of known malware.

EXPERIMENTAL PLAN | During the project period the team will work to:

- Develop a supervised training process for the pattern recognition algorithm
- Analyze a subset of malware contain in the National Vulnerabilities Database
- Implement and test the effectiveness of the proposed visual detection methods
- Compare the effectiveness of visual detection against prevailing detection methods

RELATED WORK | This research is based on Shropshire's expertise in cloud security which was gained over half a decade of – federally funded research. Additionally, Dr. Benton has myriad experiences tuning machine learning algorithms to maximize their effectiveness at various tasks. This research breaks new ground by combining visualization with integrity analysis. However, a number of previous studies are contemplating more advanced methods for intrusion detection.

HOW OURS IS DIFFERENT | The current malware monitoring process for guest virtual machines is usually performed from a peering point within each guest operating system. This approach has several drawbacks: cloud tenants must consent to the installation of software within their virtual machines, the monitoring software onboard the virtual machine is itself subject to compromise, and the process is inefficient. This research develops a model for external evaluation which does not require within-guest monitoring or other invasive techniques. Because the proposed model is out-of-band, it is less susceptible to acts of subterfuge by malware. Efficiency gains are realized due to non-reliance on guest resource allocations and/or requiring sustained connectivity between guests and centralized analytical engines. Further, the proposed model works on virtual machines, container-based and unikernels systems without adaptation.

MILESTONES FOR YEAR |

- **4 months**: Configure Chef/ Puppet and develop scripts to support development of supervised learning database.
- **8 months**: Create database and implement machine learning algorithm in C or C++
- **12 months**: Train algorithm and assess effective anomaly detection, rule-based detection, and signature based detection.

DELIVERABLES | Detailed reports analyzing:

- An algorithm for efficient conversion of virtual machine / container / unikernels images into visual depictions.
- A database of tagged samples to support supervised training of machine learning algorithms for cloud security.
- An implementable Machine learning algorithm for detecting compromised guests.

BUDGET FOR YEAR |

	1 Ph.D.
Students	32,586
Equipment	2,800
Travel	2,800
Overhead (10%)	3,814
Total	**42,000.**

NECESSARY EXPERTISE | Hardware Understanding, Coding Skills and Digital Forensic Knowledge

ECONOMICS | Both cloud clients and cloud service providers benefit from the visual analytics model because clients receive personalized security services without giving up control of their virtual machines and cloud service providers can provide value-added services without the additional liabilities associated with direct access to client images.

POTENTIAL MEMBER COMPANY BENEFITS | Cloud service providers have a precarious relationship with their tenants. Tenants want managed security services for their virtual machines but do not wish to relinquish their privacy expectations. If successfully, this research will provide a commercially-viable solution to a problem with a defined market.

PROGRESS TO DATE | A test cloud computing system utilizing 25 servers and the Openstack platform have already been implemented. A Puppet image builder has been integrated into the orchestration layer. The basic testing requirements are met.

KNOWLEDGE TRANSFER TARGET DATE | 12 months

Hardware-Based Exploitation and Forensics Evaluation of iOS Devices

PROJECT ID | DFII-16	**TYPE |** [X] New [ ] Continuing	**START DATE |** August, 2017

PROJECT LEAD/PARTICIPANTS | Students; Dr. Todd McDonald; Dr. Todd Andel; Dr. Brad Glisson

DESCRIPTION | This research seeks to identify opportunities for exploitation of iOS (iPhone Operating System) devices for the purposes of information and data recovery relative to forensics investigations. We will investigate solutions to unlock/access iOS devices for data recovery purposes using available black box tools, known vulnerabilities, and hardware-based side channel techniques such as power glitching and Electromagnetic (EM) probing. These techniques may also allow exposure of keying material from cryptographic algorithms operating on iOS devices outside of an unlock scenario. The initial project would begin by investigating low-cost EM-based attacks to develop an understanding of the information leaked along with identification of opportunities for further exploitation and manipulation. The tools utilized for this attack would include: oscilloscopes, digital multi-meters, pcs with data acquisition boards and related interfacing. The need to recover data from locked iOS devices with automatic data erasure, particularly versions 8.x and above, has been of public and government interest for quite some time. For example, the FBI paid under $1 million to a contractor for a technique used to unlock the iPhone used by one of the San Bernardino shooters [1].

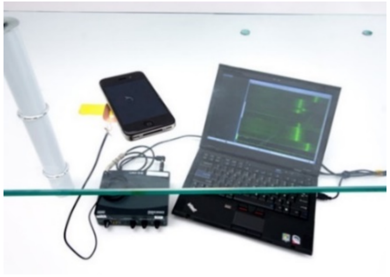

Figure 1: Improvised EM Probe of iPhone 4

EXPERIMENTAL PLAN | During the project period the team will work to:

- Develop a list of exploitation tools and vulnerabilities in iOS operating systems, categorized by version families.
- Develop a prototype tool that demonstrates breadboard data collection of iOS electromagnetic signals.
- Develop a prototype a tool that demonstrates breadboard control of iOS functions

RELATED WORK | The foundation of this research is based on McDonald and Glisson's previous experience with Mobile devices and Andel's prior work with side-channel analysis. Glisson has real-world experience with collecting data from secondary markets and corporate environments using Cellebrite and XRY mobile device tool kits. McDonald has extensive experience working with secure software engineering techniques for penetration testing. Andel has extensive experience working with hardware countermeasures.

HOW OURS IS DIFFERENT | iOS is one of the most secure operating systems based on policies created and enforced by Apple. Though not impervious or free from vulnerabilities, iOS has remained resilient to many attacks partially based on lower market share of devices, but primarily because of tight controls placed on iOS development and application deployment. Locked iOS devices pose a specific hard problem for both law enforcement and corporate IT dealing with malicious insiders or intrusion scenarios. Our approach seeks to provide functionality and services that can be used to specifically address this emerging opportunity by focusing on hardware-based and physical side-channel information.

MILESTONES FOR YEAR |

4 months: Research and acquire tools and techniques for iOS exploitation based on version families, focusing on unlock scenarios.

8 months: Setup breadboard functionality for EM analysis and input control of iPhone.

12 months: Gather relative data and build expertise for hardware-based exploit demonstration on iOS devices relative to unlock scenarios.

DELIVERABLES |

- A list of exploitation techniques for iOS based on version families including existing tools/products, vulnerabilities, software-based techniques, and hardware-based exploits.
- A demonstration framework for breadboard iPhone setup to conduct hardware-based studies relative to unlock scenarios.
- Data results for initial studies using low-cost EM-based probing of an iPhone device.

BUDGET FOR YEAR |

	1 Ph.D.
Students	31,818
Equipment	-
Travel	-
Overhead (10%)	3,182
Total	**35,000**

NECESSARY EXPERTISE | Hardware Understanding, Coding Skills and Digital Forensic Knowledge

ECONOMICS | There is a considerable market in law enforcement, government, and industry to assist forensic examiners faced with locked iOS devices. Identifying, categorizing, and realizing a demonstration framework for techniques that provide solutions for this unique iOS problem would provide great opportunity in these sectors.

POTENTIAL MEMBER COMPANY BENEFITS | This project will benefit IAB members and their clients by providing a list of potential iPhone vulnerabilities by iOS version family and provide a prototype demonstration capability for forensic data recovery from locked iPhone devices based on hardware manipulation and side-channel emanations.

PROGRESS TO DATE | Prior research by Andel and McDonald in side-channel analysis and hardware-based security properties. Prior work by Glisson investigated residual data resident on mobile devices in a Global Fortune 500 Financial Organization that identified policy breaches and opportunities for data leakage.

KNOWLEDGE TRANSFER TARGET DATE | 12 months

[1] http://www.reuters.com/article/us-apple-encryption-idUSKCN0XQ032

Investigating Methods to Capture Guest VM Memory

Project ID | DFII-18

Type | [X] New [ ] Continuing

Start Date | January, 2018

Project Lead/Participants | Dr. Jordan Shropshire; Dr. Michael Black; Students;

Description | Compromised virtual machines represent a growing risk to the stability and reliability of cloud-based operations. Not only are their onboard software, data, and services susceptible to exploitation, but they can potentially affect other guests co-located on the same hardware. There is a need for sophisticated processes for analyzing the software executing in virtual machines.

This research meets this need by developing automated methods for analyzing the memory in live VMs. It will result in software which allows for guest memory observation from a peering point within the hypervisor. The proposed methods have three desirable properties: (1) The proposed methods are invisible to the guest. They eliminates the need for software installation within the guest operating system. Further, no writes to guest memory will be required. The guest is not aware that memory collection takes place. (2) The software maintains an accurate map of guest virtual memory to host physical memory. This is achieved using passive guest introspection, a technique the investigators have refined over the past three years. (3) The proposed methods minimize memory smear by manipulating the hypervisor scheduler. The credit scheduling algorithm is retooled to synchronize with data collection apparatus. When memory is collected, the associated guest is under-scheduled and the hypervisor is overscheduled. The guest is effectively slowed but not halted.

Experimental Plan | During the project period the team will:

- Design, develop, and assemble the tooling for memory analysis of live VMs.
- Test the software using benign and compromised VMs.
- Create APIs to integrate the software with existing open source memory forensic tools.
- Benchmark the software's features (e.g., extent of smear) against other memory analysis tools.

Related Work | This research is based on Dr. Shropshire's expertise in virtualization and Dr. Black's professional and academic experience in forensic analytics. This research breaks new ground by combining methods for live VM memory analysis with forensics and cyber security. However, a number of previous studies have separately considered these concepts.

How Ours is Different | Several tools already exist for memory analysis, but they have a number of problems. First, they require the installation of software on the guest OS. They also require writes to guest memory. Together, this compromises the integrity of the guest memory and requires a significant modification of the guest's saved image. These changes are unacceptable in production cloud environments. Further, they often result in memory captures which contain high levels of smear. The proposed software overcomes these challenges by combining a number of emerging techniques for passive, unobtrusive memory analysis.

Milestones For Year |

4 months: Incorporate existing code for passive guest introspection. Develop standardized format for outputting VM memory collections

8 months: Develop methods for mapping guest virtual memory to host virtual memory

12 months: Assemble near-real-time mapping between guest virtual memory and host physical memory. Create tooling to identify and extract guest physical memory

Deliverables |

- Project codebase to be released via GitHub.
  - Includes documentation wiki
- Sample Outputs of VM captures
  - Library of memory outputs showing benign and malicious code
- Benchmark testing against other platforms

Budget for Year |

	1 Ph.D.
Students	32,586
Equipment	2,800
Travel	2,800
Overhead (10%)	3,814
Total	**42,000.**

Necessary Expertise | Hardware Understanding, VM Knowledge, Memory and Digital Forensic Knowledge

Economics | Organizations expect highly reliable cloud infrastructure. It imperative that compromised VMs are identified and isolated from healthy VMs so that they don't risk the stability of the cloud ecosystem. This research provides a low-cost method for guest analysis that won't violate most public cloud privacy agreements.

Potential Member Company Benefits | This project will result in software that members can implement within their own organizations to analysis guest memory and determine the extent of damage within virtual machines. Because it builds on open source hypervisors such as Xen, implementation is straightforward.

Progress to Date | Development infrastructure is already in place. A 25 node OpenStack installation with Xen hypervisors is available for testing. Further, several of the algorithms necessary for implementing the toolset have already been built.

Knowledge Transfer Target Date | 12 months

On-Device Detection via Anomalous Environmental Factors

Project ID | DFII-20 **Type |** [X] New [ ] Continuing **Start Date |** January, 2018

Project Lead/Participants | Students; Dr. Todd Andel; Dr. Todd McDonald; Dr. Ryan Benton

Description | This research seeks to develop real-time attacker detection capabilities through the use of device level measurements of environmental factors. We hypothesize that physical indicators from attacker activity can be statistically distinguished from normal operations. These indicators include measurements such as device temperature, power usage, CPU utilization, memory utilization, and network activity. While it is recognized an attacker may have the capability to report false data for on-chip readings, we hypothesize it is highly unlikely for an attacker to be able to provide reasonable readings for multiple sensors. For instance fabricated temperature readings may not correlate to CPU activity or current power draw. We aim to develop correlation algorithms between multiple sensors to distinguish between normal and malicious activities. This research will additionally utilize high fidelity side-channel analysis to determine correlation between the on-chip readings and physical electrical properties that emanate during system operation. The research objective during this phase is expected to determine if correlation between side-channel indicators due to attacker activities and a single on-chip sensor, such as temperature can be discovered. Identification of external side-channel correlation to an environmental factor vantage point may lead to detection algorithms that can be integrated without subsequent need for high fidelity side-channel capabilities. We will also investigate the integration of multiple redundancy sensors to provide secondary measurement channels not available to the attackers. This will provide a correlation source that cannot be subverted. One such approach is to develop multiple dispersed on-chip digital temperature sensors that are embedded during chip production and provide readings on out-of-band channels. As a research challenge, we recognize that any environmental factors may be dependent on the environment in which the device is itself deployed. It will therefore be vital to characterize a *“normal baseline*" for various *"deployment biomes"*.

Experimental Plan | During the project period the team will work to:

- Develop correlation algorithms for baseline signatures in multiple non-malicious environments.
- Evaluate and tune correlation engine against active attacker.
- Investigate sensor combinations, including multiple redundant sensors, for systems reporting false information.

Related Work | The foundation of this research is based on Andel and McDonald’s, previous experience with side-channel analysis, as well as exploit and rootkit detection. Andel has significant experience in side-channel countermeasures, McDonald has extensive experience working with exploit and rootkit detection mechanisms, and Benton is an expert in characterization algorithms. The team will additionally rely on Andel’s previous work in embedded digital temperature sensors.

How Ours is Different | Our solution serves as a forensic base capability that is not reliant in data capture from within the system (e.g., file modifications) itself, which can be subverted through rootkit level exploits that traditionally have full operating system control to provide false information. The ability to provide initial correlation research through high fidelity capability side-channel capture by the Riscure Inspector side channel analysis system is unmatched and available at very few sites. Riscure has approximately 13 U.S. based customers, mostly in government and industry.

Milestones For Year |

4 months: Develop correlation algorithms for baseline signatures from on-device environmental factor sensors and traditional side-channels

8 months: Tune correlation engine to detect anomalous attacker activity.

12 months: Develop redundant sensors with out-of-band reporting. Incorporate low-fidelity capture capabilities in place of high-level requirements.

Deliverables |

- Baseline signature dataset for non-malicious applications/hardware.
- Analysis and identification techniques to identify maliciously modified applications and systems.
- Techniques to detect unknown zero-day exploits and hardware implants.

Budget for Year |

	1 Ph.D.
Students	32,586
Equipment	2,800
Travel	2,800
Overhead (10%)	3,814
Total	**42,000.**

Necessary Expertise | Hardware Understanding, Statistics and Digital Forensic Knowledge

Economics | Advanced persistent threats can routinely subvert a system and provide false readings to cover their activity. The ability to determine such activity using on-device sensor correlations provides a cost effective detection capability.

Potential Member Company Benefits | This project will benefit IAB members and their clients by providing a identifying potential zero-day exploit and undiscovered rootkits, or even maliciously implanted hardware. Identification of such exploits from an external side-channel other correlated environmental factor vantage point may lead to detection algorithms that can be integrated without subsequent need for high fidelity side-channel capabilities.

Progress to Date | High level side-channel analysis capability being acquired though the funding of NSF Major Research Instrumentation (MRI) grant for $393K and a USA cost share of $170K.

Knowledge Transfer Target Date | 12 months

Anomalous Detection of Engine Data

Project ID | DFII-23	**Type** | [X] New [ ] Continuing	**Start Date** | January, 2018

Project Lead/Participants | Student, Dr. Tom Johnsten, Dr. Todd Andel, Dr. Ryan Benton, and Dr. Todd McDonald

Description | This research will investigate the feasibility of detecting anomalous events in the operation of aircraft engines based on data collected from externally mounted sensors. In particular, our primary emphasis will be to design and develop an anomaly detection method to distinguish normal and abnormal events in the operation of aircraft engines. The expectation is that the proposed method can provide support for preventive maintenance and retrospective investigations. In the case of preventive maintenance, the proposed method should be able to determine if an engine is acting abnormally. In the retrospective investigation, the method should be usable as a part of a post-analysis review to pinpoint when deviations from expected behavior began. Such results also have the potential to determine if the current models used for engine diagnosis and prognosis may need retraining. In addition, a secondary aim is to create datasets that can be used to evaluate the proposed methods as well as future anomaly detection methods. This will require, in conjunction with the IAB, the collection of data from externally mounted sensors deployed on aircraft engines that capture measurements such as air temperature, engine vibration, fuel usage, altitude, electrical draw, and air pressure. Since data collected from externally mounted sensors may be influenced by the environment in which they are deployed, it will therefore be important to define normal baselines for various environments.

Experimental Plan | During the project period the team will collaborate with our IAB to:

- Acquire, categorize, and pre-process engine sensor data
- Design and implement anomaly detection methods
- Conduct experiments to evaluate the methods' ability to detect anomalies using external sensors

Related Work | The foundation of this research is based on Andel and McDonalds' previous experience with side-channel analysis as-well-as Benton and Johnstens' experience in developing data mining and big data algorithms. The team will additionally rely on Andel's previous work in embedded digital temperature sensors.

How Ours is Different | Our solution serves as a forensic base capability that is not reliant in data capture from within the engine (e.g., temperature and pressure sensors) itself, which can be compromised through debris, dirty oil, and rust that traditionally provide false information.

Milestones For Year |

4 months: Acquire, categorize, and pre-process engine sensor data and select an appropriate learning approach such as clustering, classification and association mining.

8 months: Design, implement and conduct initial evaluations of proposed methods.

12 months: Refine methods and conduct an extensive evaluation

Deliverables |

- Pre-processed datasets for anomaly detection evaluation
- Code demonstrating the proposed functionality.
- Technical report detailing anomaly detection methods implemented to detect potential abnormal engine events and experimental results.

Budget for Year |

	1 Ph.D.
Students	31,818
Equipment	-
Travel	-
Overhead (10%)	3,182
Total	**35,000**

Necessary Expertise | Data Mining and Big Data Algorithms, Statistics, Side Channel Analysis and Digital Forensic Knowledge

Economics | Externally mounting sensors can lower engine production costs. Relying on particular condition sensors like oil pressure requires conclusions be drawn from these disparate sensors about engine condition and overall operational state. Internal engine conditions can compromise the operation of sensors and provide false readings. The ability to determine normal and abnormal engine operation using on-device sensor correlations provides a cost effective detection capability.

Potential Member Company Benefits | This project will benefit IAB members and their clients by providing a digital forensics tool that inherently provides engine health in real time. Identification of such conditions from an external side-channel other correlated environmental factor vantage point may lead to detection algorithms that can be integrated without subsequent need for expensive and less reliable internally embedded sensors. The technology could potentially replace current sensors.

Progress to Date | Investigators have acquired and installed various machine learning toolkits to support this investigation

Knowledge Transfer Target Date | 12 months