REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
21-02-2023	Final Report	15-Jun-2018 - 10-Jun-2019

4. TITLE AND SUBTITLE
Final Report: A Secure Data Processing Infrastructure for Research and Education in IoT, SCADA and SGX Systems

5a. CONTRACT NUMBER
W911NF-18-1-0249

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611103

6. AUTHORS

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
University of Texas at Dallas
800 West Campbell Road, AD15

Richardson, TX 75080 -3021

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
72247-NC-RIP.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
2 Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Bhavani Thuraisingham
UU	UU	UU	UU		19b. TELEPHONE NUMBER 972-883-2808

RPPR

as of 22-Feb-2023

Agency Code:

Proposal Number: **Agreement Number:**

Organization:
Address: , ,
Country:
DUNS Number: EIN:
Report Date: Date Received:
for Period Beginning and Ending
Title:
Begin Performance Period: **End Performance Period:**
Report Term: -
Submitted By: Email:
Phone:

Distribution Statement: -

STEM Degrees: **STEM Participants:**

Major Goals:

Accomplishments:

Training Opportunities:

Results Dissemination:

Plans Next Period:

Honors and Awards:

Protocol Activity Status:

Technology Transfer:

I certify that the information in the report is complete and accurate:
Signature:
Signature Date:

Research and Testbed Utilizing the Equipment Purchased Using the ARO DURIP Funds

Bhavani Thuraisingham
The University of Texas at Dallas

ABSTRACT

This report describes the sample research projects we have carried out and the IoT (Internet of Things) Testbed we are developing using the funds from the DURIP project.

1. Introduction

Our goal was to develop a secure and scalable infrastructure that can efficiently perform big data analytics that are robust to adversarial attacks, while providing a cryptographically secure mechanism for preserving data privacy and security. Learning from our past experiences of addressing privacy issues on Intel SGX, a hardware-based cryptographically secure mechanism that provides confidentiality and data integrity, we utilized our expertise in building this framework for addressing data security concerns from IoT devices.

We have purchased GPUs, SGX and other equipment and are subsequently developing the IoT and IIoT Test beds. In addition, we have also purchased equipment (e.g., drones and land vehicles) to integrate with this Testbed to conduct research in secure cyber physical systems. This report describes our sample research projects as well as the Testbed we are developing.

The organization of this report is as follows. Section 2 lists the equipment we have purchased from the DURIP funds. Section 3 provides the highlights of our research. Our sample research projects are listed in Section 4. Our Testbed is discussed in Section 5. Some directions are given in Section 6. The high level architecture of the testbed is given in Appendix A.

2. Equipment Purchased from DURIP Funds

Item	Quantity
Lambda Operating System: Ubuntu 18.04 + Lambda Stack	1
Dell w/2 TB	6
Dell with 128 GB Server	5
Intel SGX servers	7
DION/Puget machine Compact GPU workstation, Ubuntu 18.04 LTS	1
DION/Puget machines, 4 x Crucial DDR4-2666 16GB	4
Provantage SGX machines, V6 C236 16GB DDR4 8M	4
Lambda GPU Enabled Machine	5
AE DREAMS TURTLE MAIL DEVICE	1
ROBOCARS	4
HONEYWELL THERMOSTAT	1
DONKEY CAR	1
LEDDARTECH LiDAR sensor and single element sensor modules	2
Medical Alert Panic Button- WI	1
WOpet SmartFeeder Dog and Cat Feeder	1
Smarter SMCOF01-US SMCOF01-US Coffee Remote Brew	1
Dojo Smart internet security and privacy solution	1

Shark ION Robot Vacuum WIFI-Controlled	1
PetSafe Smart Feed Automatic Dog/Cat Feeder	1
Swann 1080P Indoor WiFi Wireless Camera	1
XBee 2mW Wire Antenna - Series	3
Ring Video Doorbell 2	1
Ring Chime, A Wi-Fi-Enabled	1
Foobot Indoor Air Quality Monitor	1
ECOVACS DEEBOT N79 Robotic Vacuum	1
Echo Dot Kids Edition	1
HackRF One Software Defined Radio	1
August Smart Lock, 3rd Gen door lock	1
CogniToys STEMosaur Educational Smart Toy	1
Matrice 100 Quadcopter DJM100P drone	1
Raspberry Pi Zero W (Wireless)	1
Rachio 16ZULW-C Sprinkler Controller	1
Ring Wi-Fi Enabled Video Doorbell	1
Foobot Indoor Air Quality Monitor	1
Element14 Raspberry Pi 3 B+ Model	1

3. Highlights of the Research Supported by the DURIP

Here are the highlights of the research carried out using the equipment purchased from the DURIP funds. Details are given in Section 4.

- We have designed and developed a framework to preserve data privacy utilizing a Trusted Execution Environment (TEE) such as Intel SGX, and end-to-end data encryption mechanism. We evaluate the framework by executing rule-based programs in the SGX securely with both simulated and real IoT device data.

- We have designed and developed a framework for enterprise system defenders to detect cross-platform attacks, such as Advanced Persistent Threats (APTs). Empirical evaluation and analysis on both real-world and synthetic datasets are performed to validate the effectiveness of our algorithm, comparing to state-of-the-art techniques. The implementation is carried out using GPUs.

- We use a trusted processor (e.g, Intel Software Guard eXtension (SGX)) to build a secure information retrieval system that provides better security guarantee and performance improvements. Unlike other related works, we focus on securely building the encrypted index in the cloud computing environment using the SGX, and show that the encrypted index could be used for executing keyword queries over text documents and face recognition detection in image documents.

- We have designed and developed a robust active learning technique using the DURIP equipment for situations where there are weak and adversarial oracles. Our work falls under the general umbrella of active learning in which training data is insufficient and oracles are queried to supply labels for the most informative samples to expand the training set.

- By introducing novel tools based on topological data analysis and functional data depth into Blockchain Data Analytics, we show that Ethereum network (one of the most popular blockchains for creating new crypto-tokens) can provide critical insights on price strikes of crypto-tokens that are otherwise largely inaccessible with conventional data sources and traditional analytic methods. We have used the DURIP equipment in this research.

- We have introduced SAVIOR: an architecture for securing autonomous vehicles with robust physical invariants. We implemented and validated our algorithms on two popular open-source controllers for aerial and ground vehicles, and demonstrated its effectiveness. We used the DURIP funds to purchase the equipment for this project.

- We have used the DURIP equipment to build an IoT laboratory focusing on the following two areas: (1) Our Industry 4.0 (Industry IoT) testbed has mainly comprised of Micro-controller units (MCUs), Programmable Logic Controller (PLC) boards, and various sensor and controller devices. Herein, we have created and actualized many advanced attack scenarios (e.g., Stuxnet, Triton/Trisys attack campaigns) against critical infrastructure. Industrial Control System/Cyber Physical Systems (ICS/CPS) testbed facilitates our security research to explore data-driven defense solutions. (2) We have also utilized devices for Unmanned Aerial/Ground Vehicles (UAV and UGV) to build stealthy attack scenarios and exercise their counterpart defenses.

4. Details of the Research Supported by the DURIP Equipment

4.1 Research Utilizing GPUs and SGX (Co-PI: Latifur Khan)

4.1.1 Projects related to IoT utilizing GPUs and SGXs

Project 1: The growing adoption of IoT devices in our daily life is engendering a data deluge, mostly private information that needs careful maintenance and secure storage system to ensure data integrity and protection. Also, the prodigious IoT ecosystem has provided users with opportunities to automate systems by interconnecting their devices and other services with rule-based programs. The cloud services that are used to store and process sensitive IoT data turn out to be vulnerable to outside threats. Hence, sensitive IoT data and rule-based programs need to be protected against cyberattacks. To address this important challenge, in this project, we have designed and developed a framework to maintain confidentiality and integrity of IoT data and rule-based program execution. We design the framework to preserve data privacy utilizing Trusted Execution Environment (TEE) such as Intel SGX, and end-to-end data encryption mechanism. We evaluate the framework by executing rule-based programs in the SGX securely with both simulated and real IoT device data.

Project 2: The growing adoption of IoT devices in our daily life created a need for secure systems to safely store and analyze sensitive data as well as the real-time data processing system to be as fast as possible. The cloud services used to store and process sensitive data are often turn out to be vulnerable to outside threats. Furthermore, to analyze streaming IoT data swiftly, they are in need of a fast and efficient system. This project envisioned the aspects of complexity dealing with real time data from various devices in parallel, building a solution to ingest data from different IoT devices, forming a secure platform to process data in a short time, and using various techniques of IoT edge computing to provide meaningful intuitive results to users. The project included two modules of building a real time data analytics system. In the first module, we maintain confidentiality and integrity of IoT data, which is of paramount importance, and manage large-scale data analytics with real-time data collection from various IoT devices in

parallel. We created a framework to preserve data privacy utilizing Trusted Execution Environment (TEE) such as Intel SGX, end-to-end data encryption mechanism, and strong access control policies. Moreover, we designed a generic framework to simplify the process of collecting and storing heterogeneous data coming from diverse IoT devices. In the second module, we proposed a drone-based data processing system in real-time using edge computing and on device computing. As, we know the use of drones is growing rapidly across many application domains including real-time monitoring, remote sensing, search and rescue, delivery of goods, security and surveillance, civil infrastructure inspection etc. This project demonstrated the potential drone applications and their challenges discussing current research trends and provide future insights for potential use cases using edge and on-device computing.

Project 3: The growing adoption of IoT devices in our daily life caused a need for secure systems to safely store or analyze sensitive data, as well as a decentralized data processing system to handle vast amounts of streaming data. The cloud services used to store data and process sensitive data are often turn out to be vulnerable to outside threats. Moreover, to analyze enormous streaming data swiftly, they are in need of a fast and efficient system. In this project we have designed and developed a framework to maintain confidentiality and integrity of IoT data, which is of paramount importance, and manage large-scale data anaytics. We designed the framework to preserve data privacy utilizing Trusted Execution Environment (TEE) such as Intel SGX, and end-to-end data encryption mechanism. In addition, we utilize Apache Spark for fast real-time streaming data processing from many IoT devices. We evaluate the framework by performing simple decision making in the SGX securely that involves multiple IoT devices, and a real-time anomaly detection in the streaming data from IoT devices using Spark.

4.1.2 Project utilizing GPUs and SGXs for Machine Learning Algorithms

Project 4: In this project, we addressed challenges of detecting instances from emerging classes over a non-stationary data stream during data classification. In particular, data instances from an entirely unknown class may appear in a data stream over time. Existing classification techniques utilize unsupervised clustering to identify emergence of such data instances. Unfortunately, they make strong assumptions which are typically invalid in practice; (i) Most instances associated with a class are closer to each other in feature space than instances associated with different classes, (ii) Covariates of data are normalized through an oracle to overcome the effect of a few data instances having large feature values, and (iii) Labels of instances from emerging classes are readily available soon after detection. To address the challenges that occur in practice when the above assumptions are weak, i.e., instances of each class are scattered and the true labels of novel class instances are sparsely available, we have designed and developed a practical semi-supervised emerging class detection framework. Particularly, we aim to identify similar data instances within local regions in feature space by incorporating a mutual graph clustering mechanism. We also perform online normalization along the data stream instead of assuming an oracle, and have designed and developed a classification technique that uses only a small amount of true labels for training and emerging class detection. Our empirical evaluation of this framework on real-world datasets demonstrates its superiority of classification performance compared to existing methods, while using significantly fewer labeled instances.

Project 5: Under a newly introduced setting of multistream classification, two data streams are involved, which are referred to as source and target streams. The source stream continuously generates data instances from a certain domain with labels, while the target stream does the same task without labels from another domain. Existing approaches assume that domains for both data streams are identical, which is not quite true in real world scenario, since data streams from different sources may contain distinct features. Furthermore, obtaining labels for every instance in

a data stream is often expensive and time-consuming. Therefore, it has become an important topic to explore whether labeled instances from other related streams can be helpful to predict those unlabeled instances in a given stream. Note that domains of source and target streams may have distinct features spaces and data distributions. Our objective is to predict class labels of data instances in the target stream by using the classifiers trained by the source stream. We have designed and developed a framework of multistream classification by using projected data from a common latent feature space, which is embedded from both source and target domains. This framework is also crucial for enterprise system defenders to detect cross-platform attacks, such as Advanced Persistent Threats (APTs). Empirical evaluation and analysis on both real-world and synthetic datasets are performed to validate the effectiveness of our algorithm, comparing to state-of-the-art techniques. Experimental results show that our approach significantly outperforms other existing approaches.

Project 6: Good quality similarity metrics can significantly facilitate the performance of many large-scale, real-world applications. Existing studies have identified various solutions to learn a Mahalanobis or bilinear metric in an online fashion by either restricting distances between similar (dissimilar) pairs to be smaller (larger) than a given lower (upper) bound or requiring similar instances to be separated from dissimilar instances with a given margin. However, these linear metrics learned by leveraging fixed bounds or margins may not perform well in real-world applications, especially when data distributions are complex. We aim to address the open challenge of "Online Adaptive Metric Learning" (OAML) for learning adaptive metric functions on-the-fly. Unlike traditional online metric learning methods, OAML is significantly more challenging since the learned metric could be non-linear and the model has to be self-adaptive as more instances are observed. In this project, we presented a new online metric learning framework that attempts to tackle the challenge by learning a ANN-based metric with adaptive model complexity from a stream of constraints. In particular, we have designed and developed a novel Adaptive-Bound Triplet Loss (ABTL) to effectively utilize the input constraints, and present a novel Adaptive Hedge Update (AHU) method for online updating the model parameters. We empirically validate the effectiveness and efficacy of our framework on various applications such as real-world image classification, facial verification, and image retrieval.

Project 7: In the problem setting of cross-domain sentiment classification, two different domains are introduced, and we refer to them as source and target domains respectively. For the source domain, sentiment labels are available, while those for the target domain are not available. This problem is critical and practical, as in the real world, data in some domains (source) are abundant, while those in other domains (target) may become scarce. In this project, we have designed and developed a cross-domain sentiment classification framework based on Generative Adversarial Networks (GANs) with the assistance of an attention mechanism, which aims to leverage the information available from the source domain to the target domain. Existing state-of-the-art methods mainly use multi-task learning to minimize the distance between the source and the target instances in a latent feature space. However, the projections may suffer as the deep model always tries to overfit the cross-domain adaptation task. In this project, we introduce a framework with multiple tasks, including adversarial example generation, cycle reconstruction, and cross-domain classification. Empirical evaluation and analysis on real-world datasets are being performed to validate the effectiveness of our algorithm compared to state-of-the-art (SOTA) techniques.

Project 8: For the real-world sentiment classification problem, most existing machine learning methods are biased towards majority class when the Imbalance Ratio (IR) is high for tasks such as sentiment classification. To address this problem, we have designed and developed our set convolution (SetConv) operation and episodic training strategy to extract a single representative for each class, so that classifiers can later be trained by a balanced class distribution. We prove

that our algorithm is permutation-invariant despite the order of inputs, and experiments on multiple large-scale benchmark text datasets show the superiority of our framework when compared to other SOTA methods.

Project 9: The aim of image-to-image translation algorithms is to tackle the challenges of learning a proper mapping function across different domains. Generative Adversarial Networks (GAN) have shown superior ability to handle this problem by both supervised and unsupervised ways. However, one critical problem of GAN in practice is that the discriminator is typically much stronger than the generator which could lead to failures such as mode collapse, diminished gradient, etc. To address these shortcomings, we have designed and developed a novel framework, which incorporates a powerful spatial attention mechanism to guide the generator. Specifically, our designed discriminator estimates the probability of realness or a given image, and provides an attention map regarding this prediction. The generated attention map contains the informative regions to distinguish the real and fake image, from the perspective of the discriminator. Such information is particularly valuable for the translation because the generator is encouraged to focus on those areas and produce more realistic images. We conduct extensive experiments and evaluations, and show that our method is both qualitatively and quantitatively better than other state-of-the-art image translation frameworks.

Project 10: One of the key challenges of performing label predictions over a data stream is concerned with the emergence of instances belonging to unobserved (or novel) classes over time. Although existing studies have proposed various solutions to address this challenge, they mostly focus on streams with low-dimensional data and strongly rely on the intrinsic cohesion and separation data property, i.e., instances belonging to the same class are closer to each other (cohesion) than those belonging to different classes (separation) in the observed feature space, to detect instances from unknown classes. Unfortunately, such a property is typically not inherent in high-dimensional data such as images and texts. Thus, to perform classification and novel class detection on high-dimensional data streams, we need to address two main problems: 1) Finding a feature space that exhibits cohesion and separation properties, and 2) Training with limited amount of labeled data. In this project, we have designed and developed a multi-task metric learning mechanism useful for identifying a latent space in which the cohesion and separation data property are valid and have designed a semi-supervised stream classifier called SIM based on this mechanism. We empirically measure the performance of SIM over multiple real-world image and text datasets, and demonstrate its superiority by comparing the performance with existing state-of-the-art frameworks.

4.2 Research Utilizing GPUs and SFXs and (Co-PI: Murat Kantarcioglu)

Project 11: To preserve the security and the privacy of the data need for cloud applications, encrypting the data before outsourcing has emerged as an important tool. Furthermore, to enable efficient processing over the encrypted data stored in the cloud, utilizing efficient searchable symmetric encryption (SSE) schemes became popular. Usually, SSE schemes require an encrypted index to be built for efficient query processing. If the data owner has limited power, building this encrypted index before data is outsourced to the cloud could become a computational bottleneck. At the same time, secure outsourcing of encrypted index building using techniques such as homomorphic encryption is too costly for large data. Instead, in this work, we use a trusted processor, e.g, Intel Software Guard eXtension (SGX), to build a secure information retrieval system that provides better security guarantee and performance improvements. Unlike other related works, we focus on securely building the encrypted index in the cloud computing environment using the SGX, and show that the encrypted index could be used for executing keyword queries over text documents and face recognition detection in image documents. Finally, we show the effectiveness of our system via extensive empirical evaluation.

Project 12: Recent proliferation of cryptocurrencies that allows for pseudo-anonymous transactions has resulted in a spike of various e-crime activities and, particularly, cryptocurrency payments in hacking attacks demanding ransom by encrypting sensitive user data. Currently, most hackers use Bitcoin for payments, and existing ransomware detection tools depend only on a couple of heuristics and/or tedious data gathering steps. By capitalizing on the recent advances in Topological Data Analysis, we have designed and developed a novel, efficient and tractable framework to automatically predict new ransomware transactions in a ransomware family, given only limited records of past transactions. Moreover, our new methodology exhibits high utility to detect emergence of new ransomware families, that is, detecting ransomware with no past records of transactions.

Project 13: The last decade has seen a growing interest in adversarial classification, where an attacker tries to mislead a classifier meant to detect anomalies. We study this problem in a setting where anomaly detection is being used in conjunction with differential privacy to protect personal information. We show that a strategic attacker can leverage the additional noise (introduced to ensure differential privacy) to mislead the classifier beyond what the attacker could do otherwise; we also design and develop countermeasures against such attacks. We then evaluate the impact of our attacks and defenses in road traffic congestion and smart metering examples.

Project 14: Blockchain technology and, in particular, blockchain-based cryptocurrencies offer us information that has never been seen before in the financial world. In contrast to fiat currencies, all transactions of cryptocurrencies and cryptotokens are permanently recorded on distributed ledgers and are publicly available. As a result, this allows us to construct a transaction graph and to assess not only its organization but to glean relationships between transaction graph properties and crypto price dynamics. The ultimate goal of this project is to facilitate our understanding on horizons and limitations of what can be learned on cryptotokens from local topology and geometry of the Ethereum transaction network whose even global network properties remain scarcely explored. By introducing novel tools based on topological data analysis and functional data depth into Blockchain Data Analytics, we show that the Ethereum network (one of the most popular blockchains for creating new crypto-tokens) can provide critical insights on price strikes of crypto-tokens that are otherwise largely inaccessible with conventional data sources and traditional analytic methods.

Project 15: With emergence of blockchain technologies and the associated cryptocurrencies, such as Bitcoin, understanding network dynamics behind Blockchain graphs has become a rapidly evolving research direction. Unlike other financial networks, such as stock and currency trading, blockchain based cryptocurrencies have the entire transaction graph accessible to the public (i.e., all transactions can be downloaded and analyzed). A natural question is then to ask whether the dynamics of the transaction graph impacts the price of the underlying cryptocurrency. We show that standard graph features such as degree distribution of the transaction graph may not be sufficient to capture network dynamics and its potential impact on fluctuations of Bitcoin price. In contrast, the new graph associated topological features computed using the tools of persistent homology, are found to exhibit a high utility for predicting Bitcoin price dynamics. Higher order interactions among the nodes in Blockchain graphs can be used to build much more accurate price prediction models. Using the persistent homology-based techniques, we offer a new elegant, easily extendable and computationally light approach for graph representation learning on Blockchain.

Project 16: We have designed and developed a robust active learning technique for situations where there are weak and adversarial oracles. Our work falls under the general umbrella of

active learning in which training data is insufficient and oracles are queried to supply labels for the most informative samples to expand the training set. On top of that, we consider problems where a large percentage of oracles may be strategically lying, as in adversarial settings. We present an adversarial active learning technique that explores the duality between oracle modeling and data modeling. We demonstrate on real datasets that our adversarial active learning technique is superior to not only the heuristic majority-voting technique but one of the state-of-the-art adversarial crowdsourcing technique—*Generative model of Labels, Abilities, and Difficulties* (GLAD), when genuine oracles are outnumbered by weak oracles and malicious oracles, and even in the extreme cases where all the oracles are either weak or malicious. To put our technique under more rigorous tests, we compare our adversarial active learner to the *ideal active learner* that always receives correct labels. We demonstrate that our technique is as effective as the ideal active learner when only one third of the oracles are genuine.

4.3 Research Utilizing Drones (Co-PI: Alvaro Cardenas)

Project 17: Autonomous Vehicles (AVs), including aerial, sea, and ground vehicles, assess their environment with a variety of sensors and actuators that allow them to perform specific tasks such as navigating a route, hovering, or avoiding collisions. So far, AVs tend to trust the information provided by their sensors to make navigation decisions without data validation or verification, and therefore, attackers can exploit these limitations by feeding erroneous sensor data with the intention of disrupting or taking control of the system. In this project we introduce SAVIOR: an architecture for securing autonomous vehicles with robust physical invariants. We implement and validate our proposal on two popular open-source controllers for aerial and ground vehicles, and demonstrate its effectiveness.

Project 18: Autonomous Vehicles (AVs), also known as self-driving cars, are becoming more prevalent in our daily lives. One application of AVs that has gained the interest of researchers is the concept of vehicle platooning, which is the mobilization of associated vehicles programmed to minimize distance between themselves with the goal of increasing fuel efficiency and maximizing road space. As technological advancements expand the usage of AVs to perform more complex tasks, it is imperative to secure the integrity of these devices against malicious external tampering. In this project, we have designed and developed a security framework for AVs by introducing the concept of a shared reality: verifying that sensors involved share the same physical reality. We implement our design on a custom hardware platform that uses the popular Robot Operating System (ROS) software. Our experiments show that platooning vehicles by utilizing our security framework ensures security with a low overhead while performing several autonomous tasks.

5. Development of the IoT Laboratory (Team Member: Kangkook Jee)

We are developing an IoT Laboratory to support several research projects. The lab is supported partially by the DURIP funds as well as some start-up funds. Specifically, we are developing the UTD security testbed environment, which we expect to be the driving force behind sustaining security research and community contribution. Thus far, we have utilized the DURIP equipment in the following two research areas. (1) Our Industry 4.0 (Industry IoT) testbed has mainly comprised of Micro-controller units (MCUs), Programmable Logic Controller (PLC) boards, and various sensor and controller devices. Herein, we have created and actualized many advanced attack scenarios (e.g., Stuxnet, Triton/Trisys attack campaigns) against critical infrastructure. ICS/CPS testbed facilitates our security research to explore data-driven defense solutions. (2) We have also utilized devices for Unmanned Aerial/Ground Vehicles (UAV and UGV) to build stealthy attack scenarios and exercise their counterpart defenses. Presuming APT-like attack

scenarios, we have focused on applying advanced defense approaches (e.g., ML-based) to resource-constrained IoT/IIoT domains. DURIP has played a critical role in exploring various design combinations to distribute the expensive data collection and learning tasks between the edge IoT and the cloud backend. For instance, Edge TPU devices carry out data summarization and model prediction (detection) tasks, whereas the GPU servers oversee the data to build/update models from the cloud. The high level diagram of the lab setup is given in Appendix A.

Following is some of the research we are carrying out using the lab. In the future, we will be examining IoBT (Internet of Battlefield Things) systems and plan to conduct research in this area using the testbed.

Detection of Sensor Reply and Forgery attack for ICS / CPS System: UT Dallas is carrying out an ambitious plan to build an ICS test environment. Using the testbed, we could develop many novel attack scenarios and counterpart defenses. I would like to showcase one promising research idea that detects the sophisticated ICS attacks without requiring additional resources. Stuxnet attack is a milestone campaign against the ICS system, which intercepted and forged the data pipeline to deceive the operators while disrupting centrifuge operation at the physical layer. Inspired by the Stuxnet attack, many advanced attack approaches have followed to demonstrate the advanced form of replay and forgery attacks that manipulate sensory inputs (e.g, man-in-the-HMI, and man-in-the-PLC attacks). To counter, we propose a zero-cost, yet effective detection approach that leverages an inherent asymmetry between defender and attacker. For instance, to launch a man-in-the-PLC attack, the attacker must model a system for a given control state and control input. The computational budget of the device (e.g., microcontroller PLC) and the operation deadline would significantly limit the attacker's capability to forge the accurate output. In contrast, the defender can utilize more than enough computational resources; The most supervisory hosts (e.g., HMI, EWS) do not fully use their computational resources. Our approach thus spares these resources and promises to conduct expensive but accurate modeling. In this research, we will build a generic framework that would challenge lower-level components' integrity by asking hard-to-answer questions with no disruption to the physical operation. To generalize the approach, we plan to expand the research to cover various ICS sectors to implement domain-specific modeling requirements.

6. Directions

This report has discussed the ways we have used the equipment purchased from the DURIP funds to carry out research in (i) Machine Learning and Cyber Security, (ii) Trusted Execution Environments, (iii) Blockchain, (iv) Secure Cyber Physical Systems, and (v) developing a Secure IoT Testbed. We will continue to develop the laboratory as well as continue with the research in these areas as well as explore new areas including IoBT and the Internet of Transportation Systems.

Our proposed IIoT testbed can help us design algorithms that can be used by devices we deploy in foreign networks in order to hijack the application-level objectives of the enemy IoBT. Our assumption is that the enemy will deploy attack-detection algorithms and therefore our attacks will have to be designed subject to the constraints that they will not raise any alerts. Our proposed infrastructure can help us design and understand scenarios where the U.S. Army would deploy devices in foreign IoBT networks (e.g., by launching an infiltrator device to an enemy IoBT network) and then find the limits of how the device can hijack or affect these operations while bypassing the defenses from the enemy.

Appendix (Testbed being developed partially supported by DURIP – Kangkook Jee)

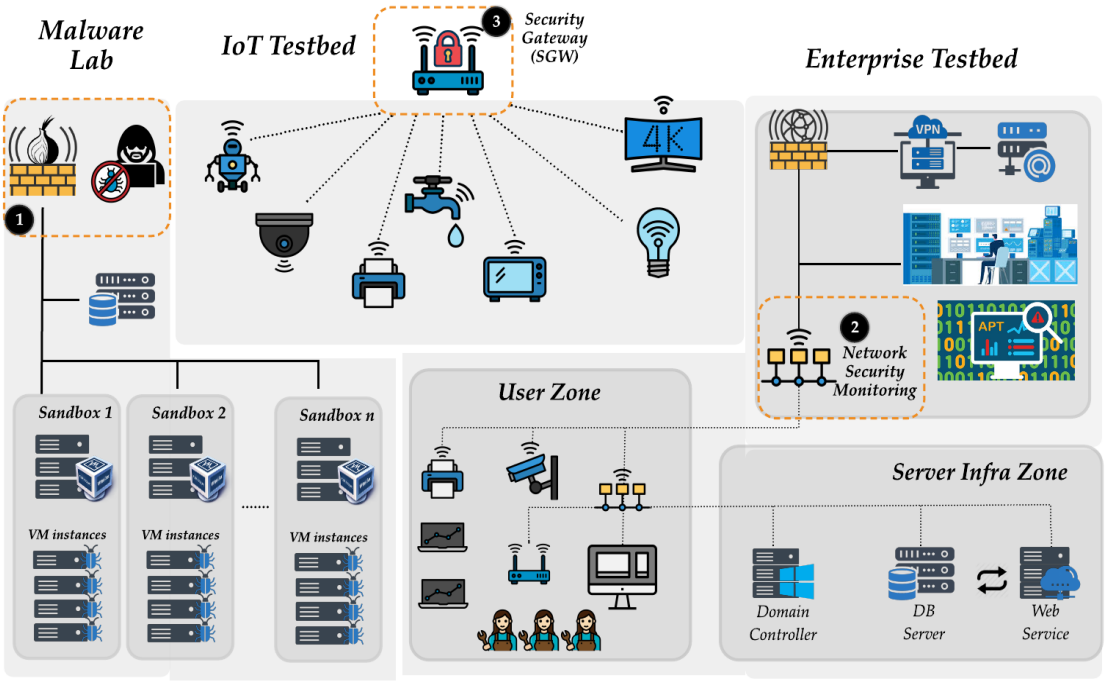

Figure 1. UTD Security lab *comprises of three segments for Enterprise testbed, IoT testbed, and Malware lab. Some labs are designed to use testbed facilities. Malware execution traces are collected from Malware lab (❶) and Lab4a and lab4b use network security monitoring (❷) to instrument in-network communication. Finally, lab 5a uses the Security Gateway (❸).*