AFRL-RI-RS-TR-2020-029

SCIENCE OF SECURE AND RESILIENT CYBER-PHYSICAL SYSTEMS

VANDERBILT UNIVERSITY

FEBRUARY 2020

FINAL TECHNICAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE

■ **AIR FORCE MATERIEL COMMAND** ■ **UNITED STATES AIR FORCE** ■ **ROME, NY 13441**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09. This report is available to the general public, including foreign nations. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2020-029 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE CHIEF ENGINEER:

/ S /
WILLIAM E. MCKEEVER
Work Unit Manager

/ S /
RICHARD MICHALAK
Acting Technical Advisor
Computing & Communications Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
FEBRUARY 2020	FINAL TECHNICAL REPORT	AUG 2014 – AUG 2019

4. TITLE AND SUBTITLE

SCIENCE OF SECURE AND RESILIENT CYBER-PHYSICAL SYSTEMS

5a. CONTRACT NUMBER
FA8750-14-2-0180

5b. GRANT NUMBER
N/A

5c. PROGRAM ELEMENT NUMBER
31011G

6. AUTHOR(S)

Xenofon Koutsoukos

5d. PROJECT NUMBER
SURE

5e. TASK NUMBER
14

5f. WORK UNIT NUMBER
VU

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Vanderbilt University
The Office of Contract & Research Administration
110 21st Avenue S STE 710
Nashville TN 37203-2416

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Air Force Research Laboratory/RITA
525 Brooks Road
Rome NY 13441-4505

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/RI

11. SPONSOR/MONITOR'S REPORT NUMBER
AFRL-RI-RS-TR-2020-029

12. DISTRIBUTION AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited. This report is the result of contracted fundamental research deemed exempt from public affairs security and policy review in accordance with SAF/AQR memorandum dated 10 Dec 08 and AFRL/CA policy clarification memorandum dated 16 Jan 09

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The objective of the project was to develop the Science of Secure and Resilient Cyber-Physical Systems. In the area of hierarchical coordination and control, the project developed tools for threat and attack modeling in order to understand the impact of cyber-attacks to system performance and operation, methods to improve resilience in CPS monitoring, resilient distributed control and coordination algorithms in the presence of malicious agents, and algorithms for improving the resilience and security of CPS flow networks such transportation, water, and power networks. In the area of decentralized security, the project developed methods for secure decentralized control, investigated in-depth multidefender games and high-resolution multi-stage security games. Our work investigated practical reasoning techniques for addressing security problems that affect CPS including actor-networks as a formal model of computation, malware detection using active learning, adversarial machine learning, privacy in CPS, optimal personalized filtering, and methods for human-in-the-loop visual analytics for exploring document collections. The project developed a suite of simulation and hardware-in-the-loop testbeds and conducted comprehensive evaluation of both system and and security properties. Finally, outreach and education activities included summer camps for high school students and community outreach supporting multiple conferences and workshops and the science of security virtual organization (SOS-VO).

15. SUBJECT TERMS
Cyber risk analysis, Resilient monitoring and control of cyber-physical systems, Decentralized security, Transportation, power, and water networks, Modeling and simulation integration for evaluation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			**WILLIAM MCKEEVER**
U	U	U	UU	58	19b. TELEPHONE NUMBER (*Include area code*) **N/A**

TABLE OF CONTENTS

Section **Page**

LIST OF FIGURES

1. SUMMARY

According to one of the widely accepted definitions, cyber-physical systems (CPS) are engineered systems where functionality emerges from the networked interaction of computational and physical processes. The tight integration of physical and computational components creates new generations of smart systems whose impacts are revolutionary; this is evident today in emerging autonomous vehicles, military platforms, intelligent buildings, smart energy systems, intelligent transportation systems, robots, and smart medical devices. Emerging industrial platforms such as the Internet of Things (IoT), Industrial Internet (II) in the US and Industrie 4.0 in Europe are triggering a gold rush toward new markets and are creating societal-scale systems.

Complex CPS abound in modern society and it is not surprising that many of these systems are safety and mission critical that makes them a target for attacks. High-profile attacks have been reported in a broad range of CPS. For example, researchers have demonstrated the ability to compromise unmanned aerial vehicles and modern automobiles with cyber attacks that can lead to catastrophic physical consequences. Even under normal conditions, CPS face complex issues crosscutting many disciplines with significant implications on essential system functions. Adding cyber-attacks in all their insidious variety creates a massive challenge that cannot be neglected due to the potential consequences. There is a profound revolution driven by technology and market forces that turns whole industrial sectors into producers of CPS. This is not about adding computing and communication equipment to conventional products where both sides maintain separate identity. This is about merging computing and networking with physical systems to create new capabilities and product qualities. Whether we recognize it or not, we are in the midst of a pervasive, profound shift in the way humans engineer physical systems and manage their physical environment using networking and information technology. Because of these disruptive changes, physical systems can now be attacked through cyberspace and cyberspace can be attacked through physical means.

Because of its significance, security and resilience have attracted considerable attention in many CPS application domains. Because of the heterogeneity and complexity, methodologies that improve CPS security are very diverse with different objectives, specifications, and constraints resulting in a broad body of knowledge. Research efforts are starting to use scientific methods and results to shape technology, practice, and policy in protecting systems from attackers, detecting intrusions, and recovering from compromises. However, scientific methods remain underutilized and they do not adequately address the involved interdisciplinary socio-technical aspects. Beyond the complex structure and interactions, security and resilience properties emerge from complex interrelationships between engineered systems and humans, they are not explained by understanding the individual elements of the system, and are highly dynamic in response to changing environment and circumstances.

The objective in this SURE project was to develop the Science of SecUre and Resilient CybEr-Physical Systems (SURE). The SURE project was organized into five thrusts:

1. *Hierarchical Coordination and Control* which is further organized into the following research projects: (1) *Cyber risk analysis and incentive design* that aimed performing risk analysis and developing strategies for security and resilience and (2) *Resilient monitoring and control* of the networked control system of the CPS infrastructure.
2. *Science of decentralized security* which aimed to develop a framework that will enable reasoning about the security of all the integrated constituent CPS components.
3. *Reliable and practical reasoning about secure computation and communication* which aims to

resolve many practical issues about security and resilience in CPS.

4. *Evaluation and experimentation* where algorithms were tested using simulation and experimentation testbeds to gain greater understanding of CPS attacks and defenses.
5. *Education and Outreach* component that aimed at educating the next generation of researchers in the field of security and resilience of CPS.

The key accomplishments and findings of the project are:

1. Developed languages and tools for threat and attack modeling in CPS in order to understand the impact of cyber-attacks to system performance and operation. The tools allow the representation of novel attack models and the characterization of the actions of that are needed to compromise the operation of the CPS. The threat and attack models developed were used for vulnerability analysis of transportation networks.
2. Developed methods to improve resilience in CPS that include methods for sensor selection and placement in adversarial settings in the context of transportation networks and water distribution systems and integrative methods that combine redundancy, diversity, and hardening for improving security and resilience.
3. Developed techniques for resilient monitoring of CPS that include scheduling and configuring intrusion detection systems, protecting multi-agent systems through mobile heterogeneous guards, investigating secure state estimation methods for multi-agent CPS such unmanned aerial vehicles, and modeling and monitoring the spread of viruses and malware.
4. Developed resilient distributed control and coordination algorithms in the presence of malicious agents. Innovations in this area include the use of trusted nodes for improving resilience of consensus algorithms in multi-agent systems and resilient distributed diffusion for multi-task state estimation that can be used for tracking multiple targets.
5. Developed methods for improving the resilience and security of CPS flow networks such transportation, water, and power networks. Our work addressed monitoring large scale systems, network routing under link disruptions, improving the detection of abnormal events in the presence of cyber attacks, and efficient emergency response.
6. In the area of decentralized security, the project developed methods for secure decentralized control, investigated in depth multidefender games considering real-world scenarios, and developed a framework for high-resolution multi-stage security games.
7. Investigated practical reasoning techniques for addressing security problems that affect CPs including actor-networks as a formal model of computation in heterogenous networks of computers, malware detection using active learning strategies, adversarial machine learning, privacy in CPS, optimal personalized filtering, and methods for human-in-the-loop visual analytics for exploring document collections,
8. Developed a suite of simulation and hardware-in-the-loop testbeds and conducted comprehensive evaluation of both system and security properties. The testbeds include a modeling and simulation integration platform that enables evaluation of resilience of transportation networks in the presence of cyber attacks based on attacker-defender games using simulations of sufficient fidelity and a hardware-in-the-loop which connects physics simulators with networked embedded computers for performing analysis of realistic cyber-attack effects in networked CPS.
9. Outreach and education: The main activity was the organization of summer camps on CPS and CPS security for high school students. The camps used RoboScape, a team-based, hands-on curriculum for teaching cybersecurity, distributed programming, and robotics. In addition,

community outreach included the support of multiple conference and workshops and the science of security virtual organization (SoS-VO).

2. INTRODUCTION

While CPS research addresses the tight interaction between the physical and cyber parts of from performance point of view, in-depth consideration of security and resilience in an integrated manner is still in early stages. Much of the cyber-security related studies and efforts in the past have focused on the challenges of adversarial environments as well as the scientific foundations and technology of networking and information technology. However, the full scope of the required research is much wider and deeper than a restructuring focusing on the cyber side; there is a profound revolution driven by technology and market forces that turns whole industrial sectors into producers of CPS. This is not about adding computing and communication equipment to conventional products where both sides maintain separate identity. This is about merging computing and networking with physical systems to create new capabilities and product qualities. Because of these disruptive changes, physical systems can now be attacked through cyberspace and cyberspace can be attacked through physical means. The ultimate objective of the project is to develop a system science for security and resilience in order to address these challenges.

The project aims at developing the principles governing secure and resilient CPS in adversarial environments and using these principles for system design and management. Systems approaches require a mix of methods and tools. The project developed integrated solutions that increase our understanding of complex interrelationships in CPS and allow design of defense strategies that improve security and resilience.

The proposed research of the SURE project was organized into five thrusts: (1) Hierarchical Coordination and Control which is further organized into Cyber risk analysis and incentive design and (2) Resilient monitoring and control, (2) Science of decentralized security (3) Reliable and practical reasoning about secure computation and communication, (4) Evaluation and experimentation and (5) Education and Outreach.

Section 3 describes the main areas and the assumptions of our research. The key accomplishments and findings of the project are organized based on these thrusts and are described in Section 4. Finally, Section 5 concludes with the main results of the project. Detailed technical results can be found in the references listed in Section 6.

3. METHODS, ASSUMPTIONS AND PROCEDURES

The rapidly emerging new research in Cyber Physical Systems (CPS) has drawn significant attention from academia, industry and government in the USA and elsewhere. CPS are engineered systems created as networks of interacting physical and computational processes. Most modern products in major industrial sectors, such as automotive, avionics, medical devices or energy production and distribution already are or rapidly becoming CPS driven by new requirements and competitive pressures. However, science and technology advancements in the 20th century have produced methods and tools for designing computational and physical systems in isolation and cannot support tight integration.

The past twenty years provided ample evidence that the separation of information and physical sciences has created a divergence in scientific foundations that has become strongly limiting to achieve progress in CPS security and resilience. For example, dominant abstractions in programming languages typically avoid the explicit representation of time, power, electro-magnetic interference, and all other physical aspects. These are simply lumped together as “non-functional” behaviors and either ignored or be dealt with separately. This leads to vulnerabilities

such as covert channels or lack of real-time reactivity. On the physical side, although engineering increasingly relies on computer-based implementations, systems science has developed and evolved abstractions that largely neglect salient properties of computing and communication platforms (such as scheduling, resource management, network delays) and considers those only as secondary implementation issues. Consequently, physical design flows typically neglect the potential impact of security breaches on the physical behavior and assume that the cyber components (networks and computing) will be correct and sufficiently protected.

While CPS research addresses the tight interaction between the physical and cyber parts of from performance point of view, in-depth consideration of security and resilience in an integrated manner is still in early stages. Much of the cyber-security related studies and efforts in the past have focused on the challenges of adversarial environments as well as the scientific foundations and technology of networking and information technology. However, the full scope of the required research is much wider and deeper than a restructuring focusing on the cyber side; there is a profound revolution driven by technology and market forces that turns whole industrial sectors into producers of CPS. This is not about adding computing and communication equipment to conventional products where both sides maintain separate identity. This is about merging computing and networking with physical systems to create new capabilities and product qualities. Whether we recognize it or not, we are in the midst of a pervasive, profound shift in the way humans engineer physical systems and manage their physical environment using networking and information technology. Because of these disruptive changes, physical systems can now be attacked through cyberspace and cyberspace can be attacked through physical means. What we need is a system science for security and resilience.

Our goal is to equip CPS designers and operators with the theoretical foundations and theory-based comprehensive tools in order to improve resilience against faults and intrusions. The overall proposed approach of the SURE project is organized as follows:

Hierarchical Coordination and Control

Networked CPS can be designed using a hierarchical coordination and control architecture that ensures resilient distributed dynamics. Resilient dynamics generalize functional performance by augmenting design concerns to attain robustness against faults and cyber attacks. The effects of failures and intrusions are usually modeled as uncertainties and casted as adversarial games.

Science of Decentralized Security

Extension of the functional design space with resilient dynamics is an important tool for designers to achieve intrusion and fault tolerant behavior of CPS. However, these results without considering the actual implementation of the "cyber side" of a CPS (typically networked monitoring and control functions) are inadequate. For example, an intrusion detection system utilizing physical invariances represents good progress, but the system will be implemented on network and computation resources that can be attacked and compromised.

Reliable and practical reasoning about secure computation and communication. In addition to the thrusts describe above, resilience and security of CPS must resolve many practical issues and consider the interactions with cyber systems and humans.

Evaluation and Experimentation The applicability of results produced by the proposed research program requires theories, concepts, and methods translated into reusable tools and software

components that can be integrated into CPS design flows. It also requires that the new design approach is tested and validated in meaningful test-beds.

4. RESULTS AND DISCUSSION

4.1 Hierarchical Coordination and Control

4.1.1 Cyber risk analysis and incentive design

Compared to general-purpose applications computing systems, CPS have the following characteristics: (1) overall behavior that emerges from the integration of the cyber and physical dynamics, (2) highly interdependent, that is failures and attacks in some parts of the infrastructure can have catastrophic consequences for other parts, and (2) stringent end-to-end requirements, such as timeliness and availability. The goal of this research is to understand the impact of cyber attacks to system performance and operation. Attacks may occur at different levels that include attacks exploit the vulnerabilities of the system platforms and application-level attacks exploit the vulnerabilities of the CPS and its services. The objectives of our work are (1) to develop modeling tools for cyber attacks allowing the representation of novel attack models, and the characterization of the actions of that are needed to compromise the operation of the CPS and (2) to perform vulnerability analysis for CPS leveraging that the attack models developed.

Modeling Tools for Cyber Attacks in CPS

Security in general and Security of Cyber-Physical Systems (CPS) in particular greatly depend on our understanding of possible attacks as well as of the attack properties. The prerequisite for quantitative and qualitative analysis of various attack properties is the existence of a knowledge base containing attack descriptions. In turn, the structure of the attack descriptions is the indispensable foundation of such knowledge base. In this work, we introduce a Cyber-Physical Attack Description Language (CP-ADL), which lays a cornerstone for the structured description of various attacks on CPS [1]. The core of the proposed language is a taxonomy for description of attacks on CPS shown in Figure 1. This taxonomy specifies which semantically distinct aspects of an attack on CPS should be described. The proposed CP-ADL language extends the taxonomy with the means to describe relationships between these aspects, despite complex cardinality relationships typical for attacks on CPS. Furthermore, the proposed CP-ADL is capable of capturing various relationships between attack descriptions, including linking between attack steps and folding of attack details.

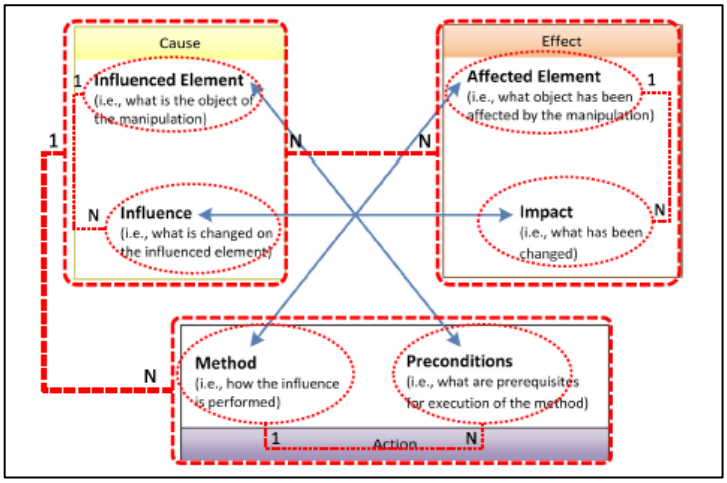

Figure 1. Taxonomy of Cyber-Physical Attacks

CP-ADL has several application areas. This section outlines three important areas: structured descriptions of known attacks on cyber-physical systems, qualitative and quantitative analyses of known attacks on cyber-physical systems and cyber-physical system vulnerability analysis.

- *Structured attack descriptions*: Structured attack descriptions are a prerequisite for many activities, including attack analysis. A structured description ensures that all the informational aspects needed for analysis are documented. At the same time, it prevents the description of irrelevant information, thereby simplifying and speeding up various types of attack analyses.
- *Qualitative and quantitative attack analyses*: Attack analyses should reveal structural attack properties as well as their frequencies. In its simplest form, a comparison of two attack descriptions should be able to determine whether or not the two descriptions are equivalent. If they differ, it is important to identify the properties that differentiate the attack descriptions. In the case of a description of a known attack, it should be possible to identify whether the attack is principally a new one or merely a known attack applied to a different cyber-physical system or cyber-physical component. As a result, it should be possible to identify and document a variety of distinct attack aspects.
- *Vulnerability analysis*: Vulnerability analysis is a common approach for enhancing system security. The insights gained via a qualitative attack analysis can be used to discover vulnerabilities existing in a system. The frequency of a specific type of attack (provided by quantitative analysis) multiplied by the estimated cost of a successful attack provides a value that can be used to rank vulnerabilities. Such a ranking is a prerequisite for cost-effective decision making regarding the vulnerabilities that should be mitigated. Vulnerability analyses have been successfully automated in the domain of computer and network security. The automation feature of CP-ADL enhances the scalability and objectivity of vulnerability analyses, which can significantly improve the security properties of cyber-physical systems.

In order to secure CPS, it is important to identify potential vulnerabilities and minimize the overall cost involved in providing and maintaining resilience. One of the ways in which this risk analysis can be performed at the design stage is threat modeling. In this context, various approaches have been proposed in the literature. Attack tree-based approaches are widely used mainly due to their simplistic design, however, static nature and state space explosion considerably restrict their applicability to CPS. Our objective is to develop a threat modeling approach that can be used not only to perform a systematic and comprehensive analysis of a wide range of threats to a variety of CPS but also develop tools to support this analysis. We develop a tool to perform systematic analysis of threat modeling for CPS using real-world examples as case studies [2]. Our work is based on an adaptation of Microsoft's SDL Threat Modeling Tool, primarily used for analyzing threats in web applications, for threat identification in the CPS domain. This work focuses on the development of an attack centric threat modeling tool using the Generic Modeling Environment (GME) developed at the Institute for Software Integrated Systems at Vanderbilt University. The model allows placing and connecting various sensors, actuators, and controllers in a system. Each object in the model contains vulnerability attributes associated with the STRIDE (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege) threat model. An integrated library of attacks represents the possible exploits by an attacker. An interpreter is developed in conjunction with the Fast Downward Planning Library in

order to generate an attack plan.

One step involved in the security engineering process is threat modeling. Threat modeling involves understanding the complexity of the system and identifying all of the possible threats, regardless of whether or not they can be exploited. Proper identification of threats and appropriate selection of countermeasures reduces the ability of attackers to misuse the system. Our work presents a quantitative, integrated threat modeling approach that merges software and attack centric threat modeling techniques [3]. The threat model is composed of a system model representing the physical and network infrastructure layout, as well as a component model illustrating component specific threats. Component attack trees allow for modeling specific component contained attack vectors, while system attack graphs illustrate multi-component, multi-step attack vectors across the system. The Common Vulnerability Scoring System (CVSS) is leveraged to provide a standardized method of quantifying the low-level vulnerabilities in the attack trees. As a case study, a railway communication network is used to illustrate the approach.

Vulnerability of Transportation Systems

Traffic signals were originally standalone hardware devices running on fixed schedules, but by now, they have evolved into complex networked systems. As a consequence, traffic signals have become susceptible to attacks through wireless interfaces or even remote attacks through the Internet. Indeed, recent studies have shown that many traffic lights deployed in practice have easily exploitable vulnerabilities, which allow an attacker to tamper with the configuration of the signal. Due to hardware-based fail safes, these vulnerabilities cannot be used to cause accidents. However, they may be used to cause disastrous traffic congestions. Building on Daganzo's well-known traffic model, we introduce an approach for evaluating vulnerabilities of transportation networks, identifying traffic signals that have the greatest impact on congestion and which, therefore, make natural targets for attacks. While we prove that finding an attack that maximally impacts congestion is NP-hard, we also exhibit a polynomial-time heuristic algorithm for computing approximately optimal attacks. We then use numerical experiments to show that our algorithm is extremely efficient in practice. Finally, we also evaluate our approach using the SUMO traffic simulator with a real-world transportation network, demonstrating vulnerabilities of this network. These simulation results extend the numerical experiments by showing that our algorithm is extremely efficient in a micro-simulation model as well. Our results are reported in [4].

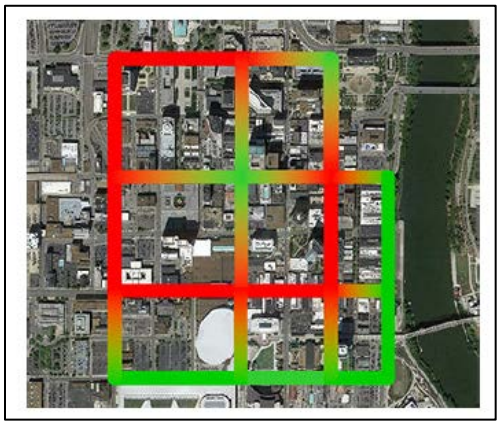

Figure 2. Traffic Heatmap for a Network Accumulation Attack (Green represents normal traffic and red congested traffic)

Recent experimental studies have shown that traffic management systems are vulnerable to cyber-attacks on sensor data. In this work, we study the vulnerability of fixed- time control of signalized intersections when sensors measuring traffic flow information are compromised and perturbed by an adversary [5] (Figure 2). The problems are formulated by considering three malicious objectives: 1) *worst-case network accumulation*, which aims to destabilize the overall network as much as possible; 2) *worst-case lane accumulation*, which aims to cause worst- case accumulation on some target lanes; and 3) *risk-averse target accumulation*, which aims to reach a target accumulation by making the minimum perturbation to sensor data. The problems are solved using bilevel programming optimization methods. Finally, a case study of a real network is used to illustrate the results.

Sensor Selection and Placement in Adversarial Settings

Our ability to dynamically control any system hinges on having accurate information about its evolving state, obtained through persistent system monitoring. In many applications, such as electric power grids or traffic networks, the system to be monitored can extend over a vast area, with many possible points of observation. Although these areas can be very large, the number of sensors that can be deployed is limited by financial and/or technological constraints. Consequently, we are faced with a problem of finding locations for placing a limited number of sensors so as to minimize our posterior uncertainty about the quantities being monitored. Due to its importance, this problem of sensor placement (or, more generally, observation/feature selection) and associated predictions about unobserved state variables has received considerable attention, particularly using learning techniques

Monitoring large areas using sensors is fundamental in a number of CPS applications. However, the number of sensors that can be deployed is often limited by financial or technological constraints. This problem is further complicated by the presence of strategic adversaries, who may disable some of the deployed sensors in order to impair the operator's ability to make predictions. Based on a Gaussian-process-based regression model, we formulate the problem of attack-resilient sensor placement as the problem of selecting a subset from a set of possible observations, with the goal of minimizing the posterior variance of prediction [6]. We show that both finding an optimal resilient subset and finding an optimal attack against a given subset are NP-hard problems. Since both the design and the attack problems are computationally complex, we propose efficient heuristic algorithms for solving them and present theoretical approximability results. Finally, we show that the proposed algorithms perform exceptionally well in practice using numerical results based on real-world datasets.

In a smart city, real-time traffic sensors may be deployed for various applications, such as route planning. Unfortunately, sensors are prone to failures and attacks, which result in erroneous traffic data. Erroneous data can adversely affect applications such as route planning, and can cause increased travel time and environmental impact. To minimize the impact of sensor failures, we must detect them promptly and with high accuracy. However, typical detection algorithms may lead to a large number of false positives (i.e., false alarms) and false negatives (i.e., missed detections), which can result in suboptimal route planning. In this work, we devise an effective detector for identifying faulty traffic sensors using a prediction model based on Gaussian Processes [7]. Further, we present an approach for computing the optimal parameters of the detector which minimize losses due to false positive and false-negative errors. We also characterize critical sensors, whose failure can have high impact on the route planning application. Finally, we implement our method and evaluate it numerically using a real-world dataset and the route

planning platform OpenTripPlanner.

Sensors deployed in cyber-physical systems (CPS) for monitoring and control purposes are prone to anomalies (e.g., reliability failures and cyber-attacks). To detect anomalies and prevent their harmful effects, anomaly detection systems (ADS) are used. However, ADS suffer from false positives (i.e., false alarms) and false negatives (i.e., missed detections), which may result in high performance degradation in CPS applications. Such detection errors can cause incorrect measurements being transmitted to a controller, and thus result in obtaining non-optimal or even destabilizing control decisions, which may compromise the performance of the system. To address the challenges caused by detection errors, it is necessary to take into account the CPS application when designing anomaly detectors, and to quantify the losses in the application caused by potential detection errors. To minimize the losses, it is desirable to reduce the detection errors as much as possible. However, there exists a trade-off between them (i.e., decreasing the rate of false alarms may increase the rate of missed faults, and vice versa), which can be changed through a detection threshold. Therefore, the performance loss caused by detection errors can be minimized by selecting the right detection threshold. Our goal is to perform these steps using a novel approach, which takes an existing anomaly detector and configures it considering the behavior of the controller. We call our framework Application-Aware Anomaly Detection (AAAD) [8]. This framework takes into account the interactions between the controller and the application, and so it can compute how each detection decision may affect the underlying application. Knowing this, the detector attempts to make detection decisions that will result in the least performance loss in the underlying application if the detection decision is not accurate due to false positive and false negative errors.

In [8], we propose a framework for minimizing the impact of detection errors in CPS. We devise an effective detector for identifying anomalies in sensor measurements using machine learning regression and an approach to recover from anomalies in order to maintain operation when detection alerts are triggered. We formulate the AAAD problem, in which a detector is optimally configured such that the performance loss in the presence of detection errors is minimized. In particular, the thresholds are selected so that the performance of the system in the presence of detection errors is as close as possible to the performance that could have been achieved if there were no detection errors. We show that the AAAD problem is computationally challenging, and then we present an efficient algorithm to find near-optimal solutions. We evaluate AAAD using simulation experiments on a case study of real-time control of traffic signals. The evaluation results demonstrate the benefits of the approach compared with standard anomaly detection techniques. We believe that the proposed approach can be useful in any system where there are a significant number of sensors with high variations in sensor values, which may cause many false-positive and false-negative errors.

Extending the work described above, we study the sensor placement problem in urban water networks that maximizes the localization of pipe failures given that some sensors give incorrect outputs [9]. False output of a sensor might be the result of degradation in sensor's hardware, software fault, or might be due to a cyber attack on the sensor. Incorrect outputs from such sensors can have any possible values which could lead to an inaccurate localization of a failure event. We formulate the optimal sensor placement problem with erroneous sensors as a set multi-cover problem, which is NP-hard, and then discuss a polynomial time heuristic to obtain efficient solutions. In this direction, we first examine the physical model of the disturbance propagating in the network as a result of a failure event, and outline the multi-level sensing model that captures several event features. Second, using a combinatorial approach, we solve the problem of sensor

placement that maximizes the localization of pipe failures by selecting m sensors out of which at most e give incorrect outputs. We propose various localization performance metrics, and numerically evaluate our approach on a benchmark and a real water distribution network. Finally, using computational experiments, we study relationships between design parameters such as the total number of sensors, the number of sensors with errors, and extracted signal features.

Integrating Redundancy, Diversity, and Hardening to Improve Security of CPS

Securing large-scale CPS systems against cyber- and cyber-physical attacks is a complicated task since these systems often face a variety of threats, have large attack surfaces, comprise heterogeneous components, may contain a number of undiscovered vulnerabilities, and have constrained resources. As a result, traditional security and resilience mechanisms ranging from redundant deployments to diversifying and hardening systems components may be useful but are not sufficient by themselves. In fact, there is clearly no “silver bullet" technique that could protect such complex systems against the entire broad spectrum of possible attacks. Thus, to provide security solutions that are capable of supporting the continued expansion of such systems, we need a multi-pronged and holistic approach. In other words, instead of relying on a single technique, defenders must employ multi-pronged solutions, which combine multiple techniques for improving the security and resilience of CPS. We can divide many of existing techniques into three canonical approaches:

- Redundancy for deploying additional redundant components in a system, so that even if some components are compromised or impaired, the system may retain normal (or at least adequate) functionality;
- Diversity for implementing components using a diverse set of component types (e.g., diverse hardware and software implementations) so that vulnerabilities which are present in only a single type have limited impact on the system; and
- Hardening for reinforcing individual components or component types (e.g., tamper-resistant hardware and firewalls), so that they are harder to compromise or impair.

A straightforward way to combine these approaches is to design and implement them independently of each other. However, for the improved security and resilience of CPS, the real benefit of combining these approaches is exhibited when they are integrated and optimized simultaneously in the design. Various defense techniques if deployed carefully can complement each other in elevating the ability of a system to resist malicious attacks. Unfortunately, a sound framework and methodology for combining techniques from different approaches is lacking. In lieu of a unified framework or methodology, defenders must follow best practices and intuition when integrating techniques, which can result in the deployment of ineffective or even vulnerable combinations.

In [10][11][12], we propose a framework for integrating redundancy, diversity, and hardening techniques for designing secure and resilient CPS. The objective is to develop a systematic framework for prioritizing investments for reducing security risk. Our contributions are:

- Establishing a system model (illustrated in Figure 3) that can capture (1) a wide variety of components that are found in CPS as well as the interactions between them, (2) a security investment model for redundancy, diversity, and hardening, and (3) a security risk model which quantifies the impact of attacks and defense mechanisms.
- Formulating the resilient CPS design problem as an optimization problem for prioritizing security investments and showing that the problem is NP-hard.

- Developing an efficient meta-heuristic design algorithm based on simulated annealing for finding near-optimal designs in practice.
- Evaluating the applicability of the approach using two case studies in canonical CPS domains of water distribution and transportation systems.

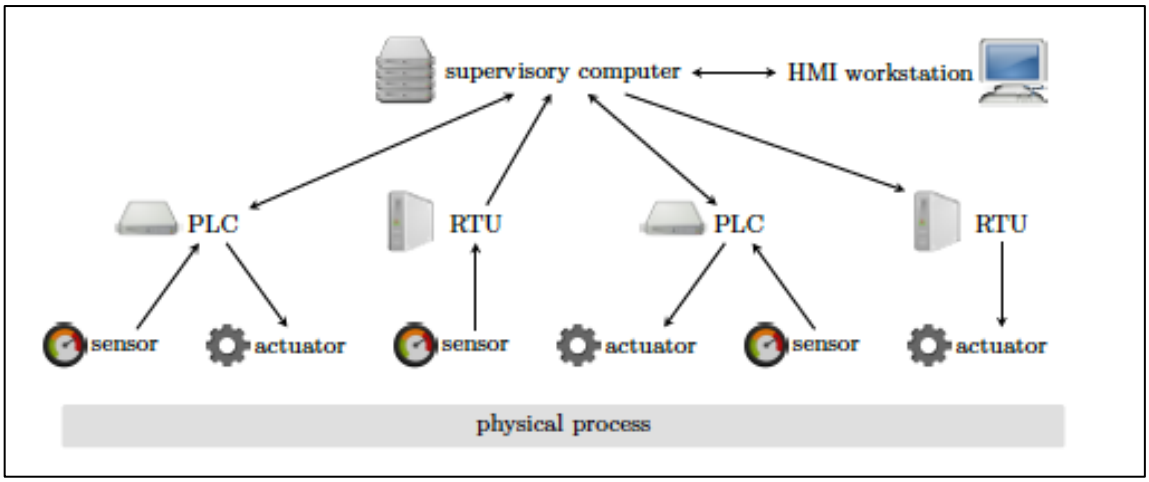

Figure 3. CPS System Model with Sensor and Control Signals

4.1.2 Resilient monitoring and control

Resilient Monitoring and Detection

Traditionally, cyber-security research has focused primarily on preventing attacks from successfully penetrating sensitive systems. However, as recent examples have shown, motivated and resourceful attackers may be able to compromise even highly secure and secluded systems. In light of these examples, we must focus not only on the "first lines" of defense but also on mitigating the effects of successful compromises, thereby increasing our system's resilience to attacks. Mitigating the effects of successful compromises is possible only if attackers are not able to inflict substantial damage immediately after compromising the system, but only after some delay. This delay allows us to implement countermeasures and prevent the system from sustaining significant losses, which is the key to increasing resilience. Due to the unalterable physical attributes of CPS, many attacks against CPS are inherently limited in how quickly they can cause substantial damage. Consequently, it is imperative that we are able to detect and react to attacks against our systems. To this end, we can deploy intrusion detection systems (IDS), which may detect attacks and alarm human operators, who can then intervene. These systems try to detect attacks by looking for signatures of known attacks (e.g., known exploits) or for anomalies (i.e., suspicious activities). Since attackers may try to stay covert until they inflict damage, detection is a challenging task, and designing detection methods for CPS is a challenging task. In the following, we summarize the major advances of our work in the project in attack detection and diagnosis of CPS. We emphasize our work on game-theoretic methods as well as scheduling of intrusion detection systems.

A well-known method that can be used for detecting anomalies in dynamical systems is sequential change detection. This method considers a sequence of measurements that starts under the normal hypothesis and then, at some point in time, changes to the anomaly hypothesis. In sequential change detection, the detection delay is the time difference between when an anomaly

occurs and when an alarm is raised. Detection algorithms may induce false positives that are alarms raised for normal system behavior. In general, it is desirable to reduce detection delay as much as possible while maintaining an acceptable false-positive rate. There exists a trade-off between the detection delay and the rate of false positives, which can be controlled by changing the sensitivity of the detector. A typical way to control the sensitivity is by changing the detection threshold. By decreasing (increasing) the detection threshold, a defender can decrease (increase) the detection delay and increase (decrease) the false-positive rate. Consequently, the detection threshold must be carefully selected, since a large value may result in large detection delays, while a small value may result in wasting resources on investigating false alarms.

Finding an optimal threshold, which optimally balances the trade-off between detection delay and rate of false positives is a challenging problem. The problem is exacerbated when detectors are deployed in systems with dynamic behavior and when the expected damage incurred from undetected attacks depends on the system state and time. Hence, defenders need to incorporate time-dependent information in computing optimal detection thresholds when facing strategic attackers. In dynamic systems, potential damage from attacks changes over time, which implies that optimal thresholds must also change with time. However, if we have to select a different threshold for each time period, then the number of possible solutions grows exponentially with the time-horizon. An adversary can attack a system in multiple ways, and each of these may cause a different amount of damage or may be detected with a different delay. To account for these differences, attack types available to the adversary must be explicitly modeled.

In [13], we study the problem of finding detection thresholds for multiple IDSes in the face of strategic attacks. We model strategic (i.e., rational) attacks against a set of computer systems that are equipped with IDSes as a two-player game between a defender and an attacker. We study the computational complexity of finding optimal attacks and defenses (i.e., optimal detection thresholds) and propose efficient heuristics. Finally, we compare our heuristic IDS thresholds to two baselines using numerical examples based on real-world intrusion detection data. The first baseline, which we call locally optimal, is configuring each IDS optimally but independently of the other IDSes. The second baseline, which we call uniform, is configuring all the IDSes in the same way, that is, having the same threshold. Our numerical results show that our approach, which optimizes multiple thresholds at the same time, outperforms the baselines, which optimize only one threshold at a time [13].

In [14], we study the problem of finding optimal detection thresholds for anomaly-based detectors implemented in dynamical systems in the face of strategic attacks. We model rational attacks against a system that is equipped with a detector as a two-player game between a defender and an attacker. We assume that an attacker can attack a system at any time. Considering that the damage is time-dependent, the attacker's objective is to choose the optimal time to launch an attack to maximize the damage incurred. On the other hand, the defender's objective is to select the detection thresholds to detect an attack with minimum delay while maintaining an acceptable rate of false positives. To this end, first we present an algorithm that selects an optimal threshold for the detector that is independent of time (i.e., fixed). We call it as a fixed threshold strategy. Next, we allow the defender to select a time-varying threshold while associating a cost with the threshold change. For this purpose, we present a polynomial time algorithm that computes thresholds that may depend on time. We call this approach the adaptive threshold strategy. We present a detailed analysis of the computational complexity and performance of both the fixed and adaptive threshold strategies. Finally, we evaluate our results using a water distribution system as a case study. Since expected damage to the system by an attack is time-dependent, the adaptive threshold strategy

achieves a better overall detection delay-false positive trade-off, and consequently minimize the defender's losses. Our simulations indicate that this is indeed the case, and adaptive thresholds outperform the fixed threshold [14].

In [15], we extended our approach for finding optimal thresholds for anomaly-based detection in dynamical systems in the face of strategic attacks. We formulate a two-player Stackelberg game between a defender and an adversary. We assume that the adversary attacks the system, choosing both the time and type of the attack to maximize the inflicted damage. On the other hand, the defender selects detection thresholds to minimize both damage from best-response attacks and the cost of false alarms. We present a dynamic-programming based algorithm to solve the game, thereby computing optimal time-dependent thresholds. We call this approach the time-dependent threshold strategy. We analyze the performance of the proposed algorithm and show that its running time scales polynomially as the length of the time horizon of interest increases, which is important in practice from the perspective of scalability. We also provide and study a polynomial-time algorithm for the problem of computing optimal fixed thresholds, which do not change with time. In addition, we study the problem of finding optimal thresholds in the presence of random faults and attacks, and present an algorithm that computes the optimal thresholds. The running time of the algorithm scales polynomially as the length of the time horizon of interest increases. Finally, we evaluate and apply our results to the detection of contamination attacks in a water-distribution system as a case study. Since expected damage to the system by an attack is time-dependent as water demand changes throughout the day, the time-dependent threshold strategy can achieve much lower losses than a fixed-threshold strategy. Our simulation results confirm this, showing that time-dependent thresholds significantly outperform fixed ones.

Detection and mitigation of attacks can be formulated as a multi-stage security game. In transportation networks, state-of-the-art traffic-control devices have evolved from standalone hardware to networked smart devices. Smart traffic control enables operators to decrease traffic congestion and environmental impact by acquiring real-time traffic data and changing traffic signals from fixed to adaptive schedules. However, these capabilities have inadvertently exposed traffic control to a wide range of cyber-attacks, which adversaries could mount easily through wireless networks or even through the Internet. Indeed, recent studies have found that a large number of traffic signals that are deployed in practice suffer from exploitable vulnerabilities, which adversaries may use to take control of the devices. Thanks to hardware-based failsafes, adversaries cannot cause traffic accidents directly by setting compromised signals to dangerous configurations. Nonetheless, an adversary could cause disastrous traffic congestion by changing the schedule of compromised traffic signals, thereby effectively crippling the transportation network. To provide theoretical foundations for the protection of transportation networks from these attacks, we introduce a game-theoretic model of launching, detecting, and mitigating attacks that tamper with traffic-signal schedules. In [4], we introduced an approach for evaluating the vulnerability of transportation networks to cyber-attacks that tamper with traffic-control devices. In [16], we extend this approach by considering detectors and countermeasures that operators can implement to mitigate these attacks. In particular, we introduce a game-theoretic model, in which an operator can setup anomaly-based detectors and mitigate ongoing attacks by reconfiguring traffic control. Similar to [4], we build on the cell-transmission model introduced by Daganzo. We formulate a multi-stage security game that models the detection and mitigation of cyber-attacks against transportation networks. We propose a metaheuristic search algorithm for finding effective detector configurations, which minimize losses in the face of strategic attacks. We introduce an anomaly-based detector for attacks against traffic control, which is built on a Gaussian-process

based model of normal traffic. We evaluate our detector and algorithms based on detailed simulations of traffic using SUMO.

We also study how we can efficiently schedule intrusion detection systems in resource-bounded CPS. In order to be resilient to attacks, a CPS must be able to detect attacks before they can cause significant damage. To achieve this, intrusion detection systems (IDS) may be deployed, which can detect attacks and alert human operators, who can then intervene. However, the resource-constrained nature of many CPS poses a challenge, since reliable IDS can be computationally expensive. Consequently, computational nodes may not be able to perform intrusion detection continuously, which means that we have to devise a schedule for performing intrusion detection. While a uniformly random schedule may be optimal in a purely cyber system, an optimal schedule for protecting CPS must also take into account the physical properties of the system, since the set of adversarial actions and their consequences depend on the physical systems. We study IDS scheduling problems in two settings and under the constraints on the available battery supplies [17]. In the first problem, the objective is to design, for a given duration of time *T,* scheduling schemes for IDS so that the probability of detecting an attack is maximized within that duration. We propose efficient heuristic algorithms for this general problem and evaluate them on various networks. In the second problem, our objective is to design scheduling schemes for IDS so that the overall lifetime of the network is maximized while ensuring that an intruder attack is always detected. Various strategies to deal with this problem are presented and evaluated for various networks.

Scheduling IDS in resource bounded CPS must consider energy and power constraints. Sensor networks monitoring spatially-distributed physical systems often comprise battery-powered sensor devices. To extend lifetime, battery power may be conserved using sleep scheduling: activating and deactivating some of the sensors from time to time. Scheduling sensors with the goal of maximizing average coverage, that is the average fraction of time for which each monitoring target is covered by some active sensor has been studied extensively. However, many applications also require time-critical monitoring in the sense that one has to minimize the average delay until an unpredictable change or event at a monitoring target is detected. We study the problem of sleep scheduling sensors to minimize the average delay in detecting such time-critical events in the context of monitoring physical systems that can be modeled using graphs, such as water distribution networks [18]. We provide a game-theoretic solution that computes schedules with near optimal average delays. We illustrate that schedules that optimize average coverage may result in large average detection delays, whereas schedules minimizing average detection delays using our proposed scheme also result in near optimal average coverage.

In the area of resilient monitoring, we consider also the problem of guarding multi-agent systems against a sequence of intruder attacks through mobile heterogeneous guards (guards with different ranges). We use graph theoretic abstractions of such systems in which agents are the nodes of a graph and edges represent interconnections between agents. Guards represent specialized mobile agents on specific nodes with capabilities to successfully detect and respond to an attack within their guarding range. Using this abstraction, we address the problem in the context of eternal security problem in graphs [19]. Eternal security refers to securing all the nodes in a graph against an infinite sequence of intruder attacks by a certain minimum number of guards. We employ heterogeneous guards and address all the components of the eternal security problem including the number of guards, their deployment and movement strategies. In the proposed solution, a graph is decomposed into clusters and a guard with appropriate range is then assigned

to each cluster. These guards ensure that all nodes within their corresponding cluster are being protected at all times, thereby achieving the eternal security in the graph.

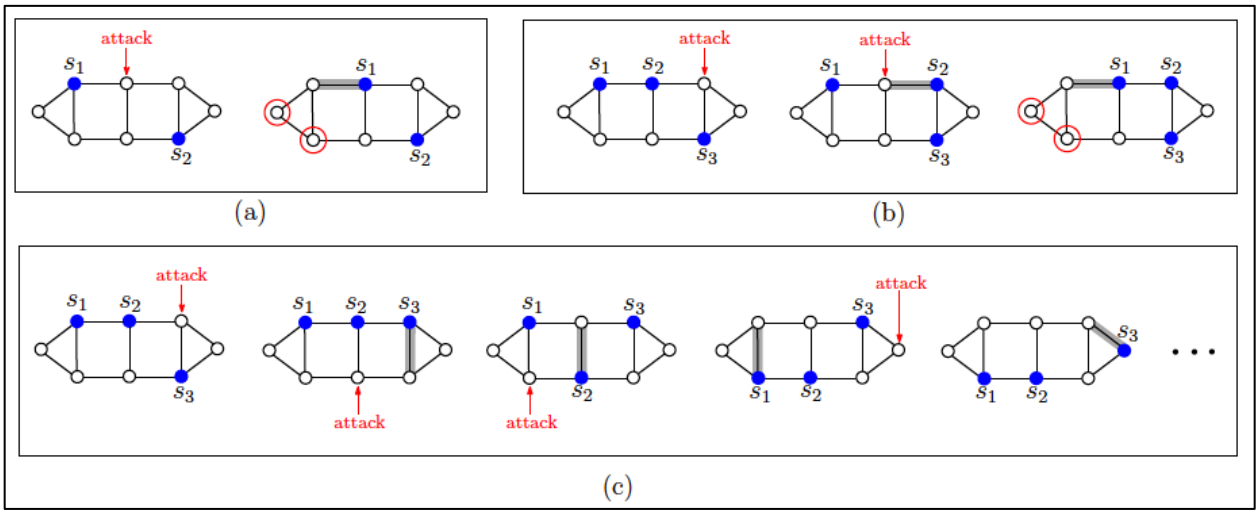

Figure 4. Guarding Networks Through Heterogeneous Mobile Guards

Figure 4 illustrates the results: (a) Two guards, *s1* and *s2*, each capable of detecting and responding to an attack on an adjacent vertex are securing the vertices of a graph through an initial secure configuration. In the case of an attack on a vertex indicated by an arrow, s1 moves towards it to counter the attack. The resulting configuration of guards is unsecure as the circled vertices have no guard in their neighborhoods. Here, the problem is that the number of guards is not sufficient. (b) Three guards *s*1, *s2*, and *s3*, each having the range 1 are deployed. After two intruder attacks, guards' configuration is unable to secure all the vertices. Though the number of guards are sufficient in this case, the strategy for the movement of guards to counter attacks is not appropriate to eternally secure a graph against an arbitrary sequence of attacks. (c) Guards move to counter attacks such that the resulting configuration is always secure, i.e., for every vertex there always exists a guard to secure it. This makes a graph eternally secure against any sequence of attacks.

Due to their low deployment costs, wireless sensor networks (WSN) may act as a key enabling technology for a variety of spatially-distributed CPS applications, ranging from intelligent traffic control to smart grids. However, besides providing tremendous benefits in terms of deployment costs, they also open up new possibilities for malicious attackers, who aim to cause financial losses or physical damage. Since perfectly securing these spatially distributed systems is either impossible or financially unattainable, we need to design them to be resilient to attacks: even if some parts of the system are compromised or unavailable due to the actions of an attacker, the system as a whole must continue to operate with minimal losses. In a CPS, control decisions affecting the physical process depend on the observed data from the sensor network. Any malicious activity in the sensor network can therefore severely impact the physical process, and consequently the overall CPS operations. These factors necessitate a deeper probe into the domain of resilient WSN for CPS. We provide an overview of various dimensions in this field, including objectives of WSN in CPS, attack scenarios and vulnerabilities, notion of attack-resilience in WSN for CPS, and solution approaches towards attaining resilience in [20]. We also highlight major challenges, recent developments, and future directions in this area.

In the context of sensor deployment and placement for event and attack detection, we investigate the coverage efficiency of a sensor network consisting of sensors with circular sensing footprints of different radii. The objective is to completely cover a region in an efficient manner through a controlled (or deterministic) deployment of such sensors. In particular, it is shown that when sensing nodes of two different radii are used for complete coverage, the coverage density is increased, and the sensing cost is significantly reduced as compared to the homogeneous case, in which all nodes have the same sensing radius. Configurations of heterogeneous disks of multiple radii to achieve efficient circle coverings are analyzed in detail. Our results are reported in [21].

In addition, we investigate the problem of secure estimation for Unmanned Aerial Vehicles (UAVs) against adversarial cyber attacks. In the coming years, usage of UAVs is expected to grow tremendously. Maintaining security of UAVs under cyber attacks is an important yet challenging task, as these attacks are often erratic and difficult to predict (Figure 5). Secure estimation problems study how to estimate the states of a dynamical system from a set of noisy and maliciously corrupted sensor measurements. The fewer assumptions that an estimator makes about the attacker, the larger the set of attacks it can protect the system against. In this work, we focus on sensor attacks on UAVs and attempt to design a secure estimator for linear time-invariant systems based on as few assumptions about the attackers as possible.

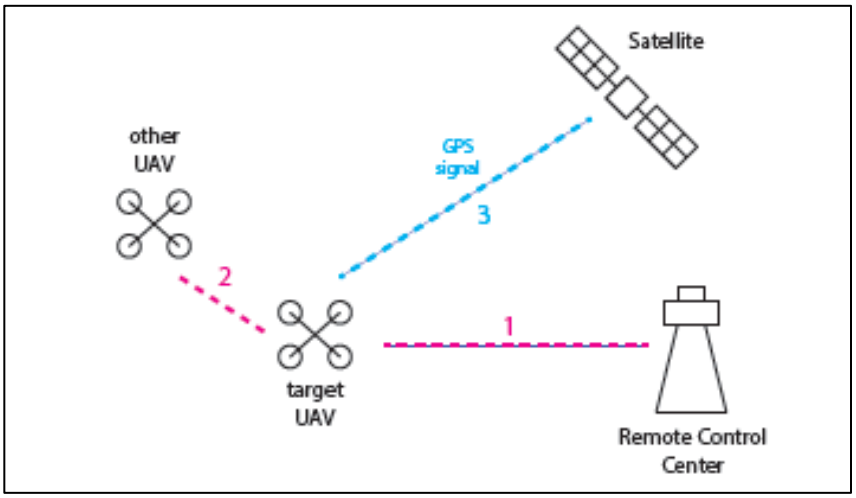

Figure 5. Different Communication Channels that are Subject to Cyber Attacks

We propose a computationally efficient estimator that protects the system against arbitrary and unbounded attacks, where the set of attacked sensors can also change over time. In addition, we propose to combine our secure estimator with a Kalman Filter for improved practical performance and demonstrate its effectiveness through simulations of two scenarios where an UAV is under adversarial cyber attack. Our approach is described in [22].

A broad variety of problems, such as targeted marketing and the spread of viruses and malware, have been modeled as selecting a subset of nodes to maximize diffusion through a network. In cyber-security applications, however, a key consideration largely ignored in this literature is stealth. In particular, an attacker often has a specific target in mind, but succeeds only if the target is reached (e.g., by malware) before the malicious payload is detected and corresponding countermeasures deployed. The dual side of this problem is deployment of a limited number of monitoring units, such as cyber-forensics specialists, so as to limit the likelihood of

such targeted and stealthy diffusion processes reaching their intended targets. We investigate the problem of optimal monitoring of targeted stealthy diffusion processes, and show that a number of natural variants of this problem are NPhard to approximate. On the positive side, we show that if stealthy diffusion starts from randomly selected nodes, the defender's objective is submodular, and a fast greedy algorithm has provable approximation guarantees. In addition, we present approximation algorithms for the setting in which an attacker optimally responds to the placement of monitoring nodes by adaptively selecting the starting nodes for the diffusion process. Our experimental results show that the proposed algorithms are highly effective and scalable [23][24].

Finally, participatory sensing can enable individuals, each with limited sensing capability, to share measurements and contribute towards develop a complete knowledge of their environment. The success of a participatory sensing application is often measured in terms of the number of users participating. In most cases, an individual's eagerness to participate depends on the group of users who already participate. For instance, when users share data with their peers in a social network, the engagement of an individual depends on its peers. Such engagement rules have been studied in the context of social networks using the concept of k-core, which assumes that participation is determined solely by network topology. However, in participatory sensing, engagement rules must also consider user heterogeneity, such as differences in sensing capabilities and physical location. To account for heterogeneity, we introduce the concept of (r ; s)-*core* to model the set of participating users. We formulate the problem of maximizing the size of the (r ; s)-*core* using 1) anchor users, who are incentivized to participate regardless of their peers, and by 2) assigning capabilities to users. Since these problems are computationally challenging, we study heuristic algorithms for solving them. Based on real-world social networks as well as random graphs, we provide numerical results showing significant improvement compared random selection of anchor nodes and label assignments. Preliminary results are presented in [25].

Resilient Distributed Consensus

CPS have witnessed a paradigm shift from centralized to distributed system design, propelled by advances in networking and low-cost, high performance embedded systems. This shift has led to significant interest in the design and analysis of multi-agent networks. A multi-agent network consists of a set of individuals called agents, or nodes, equipped with some means of sensing or communicating along with computational resources and possibly actuation. Through a medium, which is referred to as the network, the agents share information in order to achieve specific group objectives. Some examples of group objectives include consensus, synchronization, surveillance, and formation control. In order for the group objectives to be achieved, distributed algorithms are used to coordinate the behavior of the agents. There are several advantages to using multiple agents over a single one. First, the objective may be complex and challenging, or possibly even infeasible for a single agent to achieve. Second, employing many agents can provide robustness in the case of failures or faults. Third, networked multi-agent systems are flexible and can support reconfigurability. Finally, there are performance advantages that can be leveraged from multiple agents. Along with the advantages come certain challenges. Perhaps the most fundamental challenge in the design of networked multi-agent systems is the restriction that the coordination algorithms use only local information, i.e., information obtained by the individual agent through sensor measurements, calculations, or communication with neighbors in the network. In this manner, the feedback control laws must be distributed. A second challenge lies in the fact that not only is each agent typically a dynamical system, but the network itself is dynamic. This challenge arises because the agents may be mobile and the environment may be changing, thus giving rise

to dynamic (or switching) networks. Since the distributed algorithms depend directly on the network, this additional source of dynamics can affect the stability and performance of the networked system. An especially important challenge is that multi-agent networks, like all large-scale distributed systems, have many entry points for malicious attacks or intrusions. For the success of the group objective, it is important that the cooperative control algorithms are designed in such a way that they can withstand the compromise of a subset of the nodes and still guarantee some notion of correct behavior at a minimum level of performance. We refer to such a multi-agent network as being resilient to adversaries. Given the growing threat of malicious attacks in large-scale cyber-physical systems, this is an important and challenging problem. It is important that consensus algorithms be resilient to various forms of uncertainty, whether the source of uncertainty is caused by implementation effects, faults, or security breaches. We summarize the advances of our work in the project in resilient distributed control and coordination algorithms.

In our previous work, we introduced an Adversarial Resilient Consensus Protocol (ARC-P) capable to achieve consensus despite false information provided by a limited number of malicious nodes. In order to resilient to false information, the algorithm requires a mesh-like topology, so that multiple alternative information flow paths exist. However, these assumptions are not valid in some practical CPS. For instance, in Smart Grid, an emerging distributed CPS, the node connectivity is expected to resemble the scale free network topology. Especially closer to the end customer, in home and building area networks, the connectivity graph resembles a tree structure.

We have developed a scheme for a resilient distributed consensus problem through a set of trusted nodes within the network [26][27]. Currently, algorithms that solve resilient consensus problem demand networks to have high connectivity to overrule the effects of adversaries, or require nodes to have access to some non-local information. In our scheme, we incorporate the notion of trusted nodes to guarantee distributed consensus despite any number of adversarial attacks, even in sparse networks. A subset of nodes, which are more secured against the attacks, constitute a set of trusted nodes. It is shown that the network becomes resilient against any number of attacks whenever the set of trusted nodes form a connected dominating set within the network. Figure 6 illustrates the results for a graph with eight nodes which include two malicious and two trustworthy nodes. We also study a relationship between trusted nodes and the network robustness.

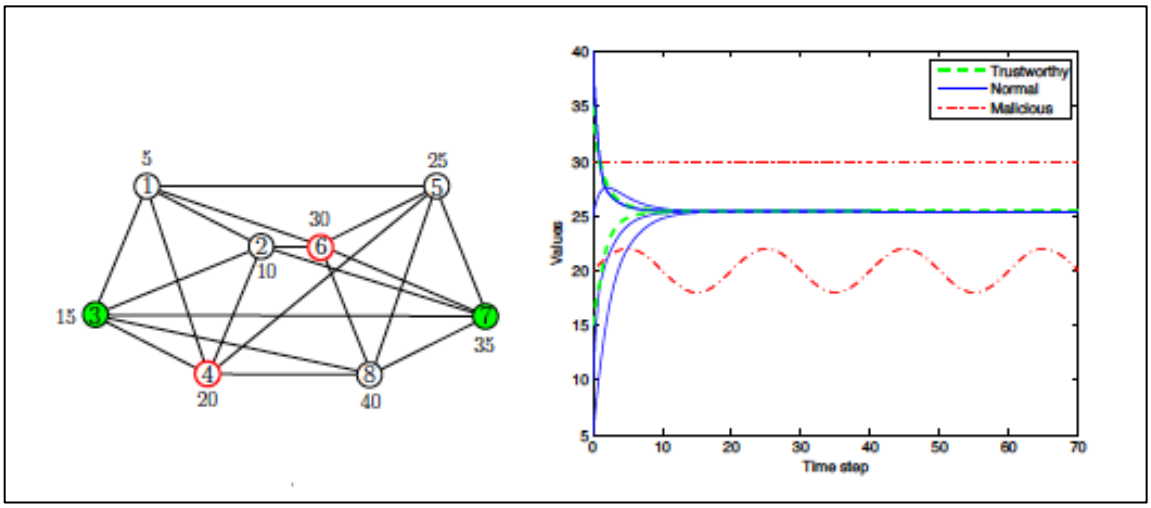

Figure 6. Two Trusted Nodes Ensure Consensus in the Presence of Adversaries

Network connectivity is a primary attribute and a characteristic phenomenon of any networked system. A high connectivity is often desired within networks for instance to increase robustness to failures, and resilience against attacks. A typical approach to increasing network connectivity is to strategically add links; however, adding links is not feasible in many situations. In this work, we propose an alternative approach to improving network connectivity, that is by making a small subset of nodes and edges "trusted," which means that such nodes and edges remain intact at all times and are insusceptible to failures. We then show that by controlling the number of trusted nodes and edges, any desired level of network connectivity can be obtained. Figure 7 illustrates the approach using a graph with ten nodes. Along with characterizing network connectivity with trusted nodes and edges, we present heuristics to compute a small number of such nodes and edges. Finally, we illustrate our results on various networks.

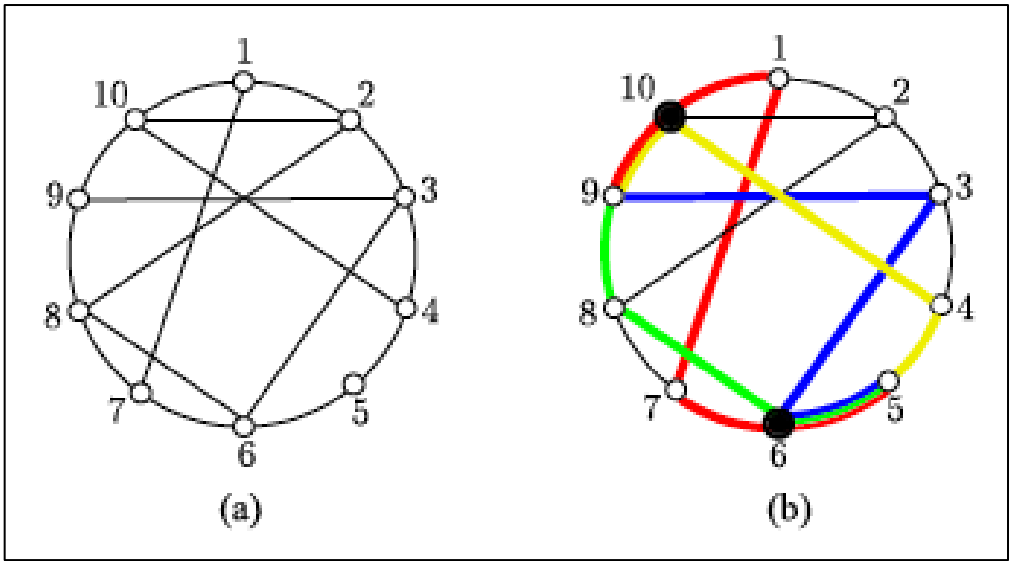

Figure 7. (a) The Graph is 2-connected.
(b) The Graph Becomes 4-connected with 2 Trusted Nodes

To observe and control a networked or distributed system, especially in failure-prone circumstances, it is imperative that the underlying network structure is robust against node or link failures. A common approach for increasing network robustness is redundancy: deploying additional nodes and establishing new links between nodes. However, the number of additional network elements and the resulting cost can be prohibitively high for a wide range of applications, such as distributed consensus. This work addresses the problem of improving network connectivity and structural robustness without adding any extra links between nodes. The main idea is to ensure that a small subset of nodes, referred to as the trusted nodes, remain intact and function correctly at all times. We show that as a result of hardening these trusted nodes, the overall network connectivity and robustness is improved significantly. In this direction, we extend two fundamental metrics with the notion of trusted nodes, network connectivity and r-robustness, the latter of which is particularly useful in characterizing the resilience of various dynamical processes over networks. We show that by controlling the number and location of trusted nodes, any desired connectivity and robustness can be achieved without adding extra links. Figure 8 illustrates the approach. We also show that finding an optimal set of trusted nodes is computationally hard. We

then provide results on constructing robust networks and relating trusted nodes to the addition of redundant elements. Finally, we apply these concepts and present a resilient consensus algorithm with trusted nodes and analyze its performance. We show that, unlike existing algorithms, resilient consensus is possible in sparse networks containing few trusted nodes. Our results are presented in [26][27].

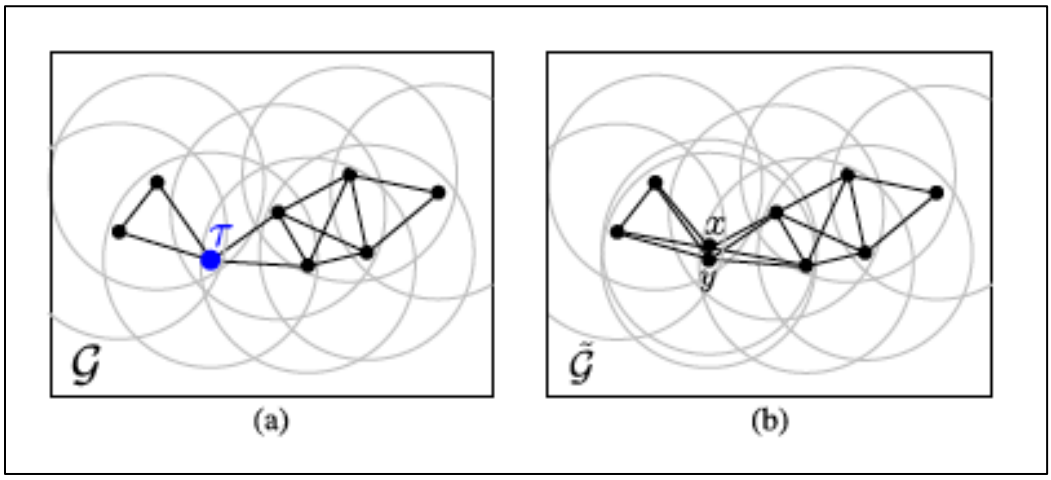

Figure 8. (a) A Unit-disk Proximity Network that is 2-robust with a Trusted Node τ. (b) A 2-robust Network Obtained by Replacing the Node τ with two Other Nodes

Local interaction rules for consensus and synchronization are vital for many applications in distributed control of Cyber-Physical Systems (CPS). However, most research in this area assumes all nodes (or agents) in the networked system cooperate. Our approach considers local interaction rules for resilient consensus and synchronization whenever a subset of the agents in the network do not follow the prescribed rules. The agents in the networked system influence one another by sharing state or output information according to a directed, time-varying graph. The normal agents, which follow the prescribed rules, have identical dynamics modeled by continuous-time, linear, time-invariant (LTI) systems. The adversary agents, which do not follow the prescribed rules, are assumed to satisfy three assumptions: (i) all transmitted signals must be uniformly continuous functions of time, (ii) each adversary node conveys the same information to each out-neighbor in the network at any given point in time, and (iii) there are at most F adversary nodes in the network. Under these assumptions, we design a consensus protocol that enables the normal agents to achieve consensus asymptotically despite the negative influence of the adversary agents. Then, we design dynamic state and output control laws for the normal agents, which utilize the consensus protocol, to facilitate asymptotic synchronization of the normal agents to a common zero-input state trajectory, again despite the influence of the adversary agents. Necessary and sufficient conditions on the network topology for resilient consensus and synchronization are articulated based on the property of network robustness [28]. Network robustness is a connectivity property of graphs that captures the notion of sufficient redundancy of directed edges between subsets of nodes.

In addition to distributed consensus, Diffusion Least-Mean Squares (DLMS) is a powerful algorithm for distributed state estimation. It enables networked agents to interact with neighbors to process streaming data and diffuse information across the network to perform the estimation tasks. Compared to a centralized approach, diffusion offers multiple advantages including robustness to drifts in the statistical properties of the data, scalability, reliance on the local data,

and fast response among others. Applications of distributed diffusion include spectrum sensing in cognitive networks, target localization, distributed clustering, and biologically inspired designs for mobile networks.

Diffusion strategies are known to be robust to node and link failures as well as to high noise levels. In our work, we consider distributed diffusion for multi-task estimation where networked agents must estimate distinct, but correlated states of interest by processing streaming data [29]. We are interested in understanding if adaptive weights introduce vulnerabilities that can be exploited by an attacker. The first problem we consider is to analyze if it is possible for an attacker to compromise a node, and make other nodes in its neighborhood converge to a state selected by the attacker. Then, we consider a network attack and determine a minimum set of nodes to compromise to make all nodes within the network converge to attacker's desired states. We know that without cooperation, all nodes can estimate their states with a known error as measured by the mean-square-deviation (MSD). However, nodes can do better if they cooperate and utilize the information collected from others while computing their own estimates. However, if there is even a single compromised node due to an attack, it can drive its neighbors' estimates to a desired value. Thus, cooperation is good on the one hand when there is no attack, but on the other hand could be adverse even if one of the nodes is compromised. We analyze this issue in detail and we propose a resilient diffusion algorithm that performs better than the non-cooperative case even in the presence of adversarial nodes [30].

By exploiting the adaptive weights used for diffusing information, we develop attack models that drive normal agents to converge to states selected by an attacker. The attack models can be used to deceive a specific node or the entire network and are applicable to both stationary and non-stationary state estimation. We develop a resilient distributed diffusion algorithm under the assumption that the number of compromised nodes in the neighborhood of each normal node is bounded by F. By selecting appropriate F, the proposed algorithm ensures that compromised nodes are not able to drive normal nodes to incorrect estimates. If the parameter F selected by the normal agents is large, the resilient distributed diffusion algorithm degenerates to non-cooperative estimation. Thus, we also analyze tradeoff between the resilience of distributed diffusion and its performance degradation in terms of MSD.

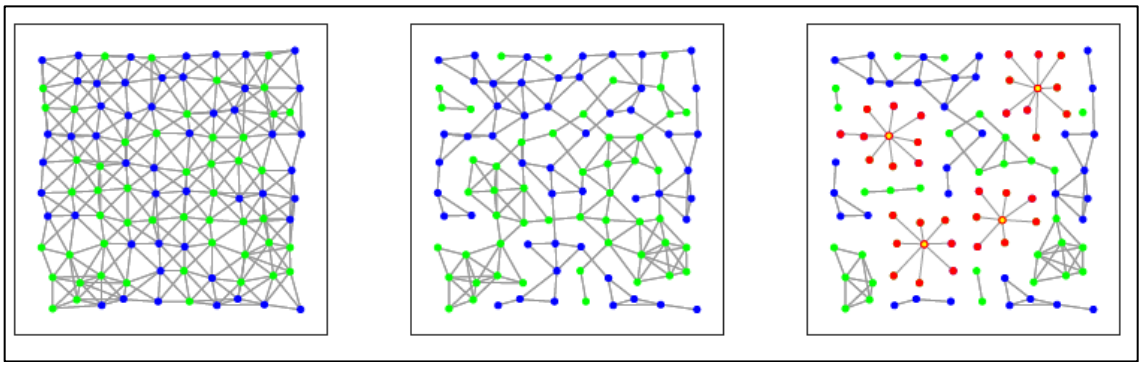

Figure 9. (a) Initial Network Topology,
(b) Distributed Diffusion with no Attack,
(c) Distributed Diffusion under Attack

We evaluate the proposed attack models and the resilient estimation algorithm using both stationary and non-stationary multi-target localization. The simulation results are consistent with

our theoretical analysis and show that the approach provides resilience to attacks while incurring performance degradation which depends on the assumption about the number of compromised nodes. Representative results are shown in Figure 9 and 10.

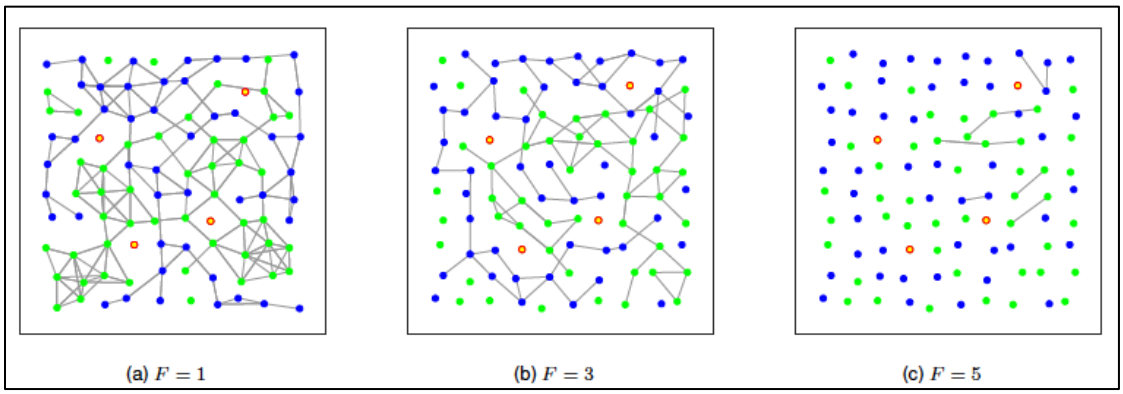

Figure 10. Network Topologies for Resilient Distributed Diffusion

4.1.3 Resilience and Security in CPS Flow Networks

According to one of the widely accepted definitions, CPS are engineered systems where functionality emerges from the networked interaction of computational and physical processes. The tight integration of physical and computational components creates new generations of smart systems whose impacts are revolutionary; this is evident today in intelligent transportation systems, power grids, and water distribution networks. It is not surprising that many of these systems are safety and mission critical that makes their resilience against faults and cyber-attacks an essential problem. H-CPS design processes use abstraction layers dictated by the heterogeneity of their component technology. Figure 11 shows a simplified view of abstraction layers that have distinct architectures and vulnerabilities [31]. In the following, we outline several studies in the project that aim at addressing such vulnerabilities for CPS flow networks that include transportation, power, and water networks.

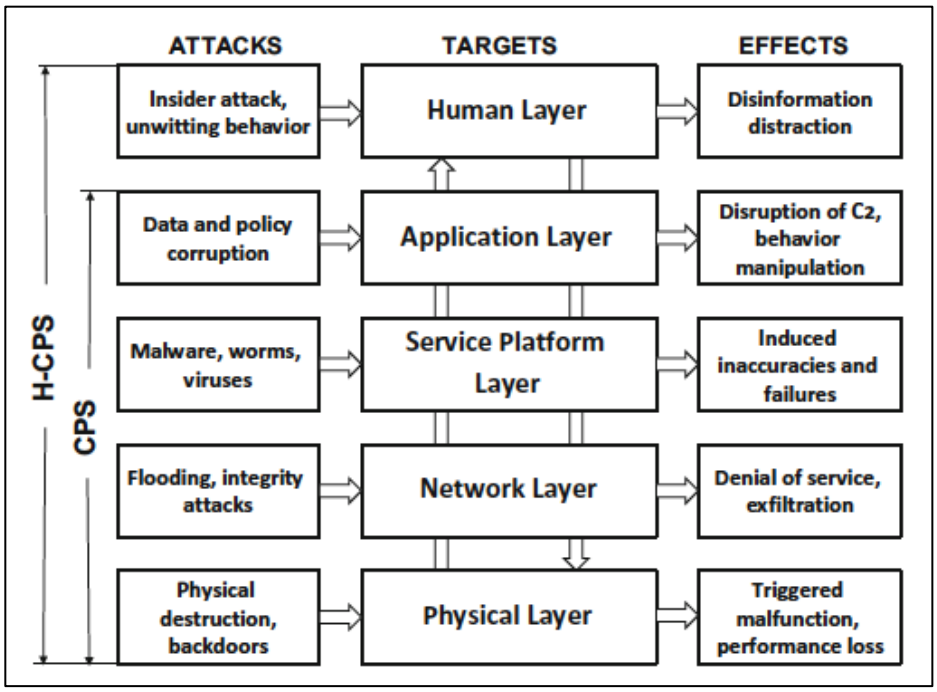

Figure 11. CPS Abstraction Layers and Types of Attacks

Monitoring large areas using sensors is fundamental in a number of CPS applications including electric power grid, traffic networks, and sensor-based pollution control systems. However, the number of sensors that can be deployed is often limited by financial or technological constraints. This problem is further complicated by the presence of strategic adversaries, who may disable some of the deployed sensors in order to impair the operator's ability to make predictions. Based on a Gaussian-process-based regression model, we formulate the problem of attack-resilient sensor placement as the problem of selecting a subset from a set of possible observations, with the goal of minimizing the posterior variance of predictions [6]. We show that both finding an optimal resilient subset and finding an optimal attack against a given subset are NP-hard problems. Since both the design and the attack problems are computationally complex, we propose efficient heuristic algorithms for solving them and present theoretical approximability results. Finally, we show that the proposed algorithms perform exceptionally well in practice using numerical results based on real-world datasets.

In addition, we consider a 2-player strategic game for network routing under link disruptions. Player 1 (defender) routes flow through a network to maximize her value of effective flow while facing transportation costs. Player 2 (attacker) simultaneously disrupts one or more links to maximize her value of lost flow but also faces cost of disrupting links. This game is strategically equivalent to a zero-sum game. Linear programming duality and the max-flow min-cut theorem are applied to obtain properties that are satisfied in any mixed Nash equilibrium. In any equilibrium, both players achieve identical payoffs. While the defender's expected transportation cost decreases in attacker's marginal value of lost flow, the attacker's expected cost of attack increases in defender's marginal value of effective flow. Interestingly, the expected amount of effective flow decreases in both these parameters. These results can be viewed as a generalization of the classical max-flow with minimum transportation cost problem to adversarial environments. Technical details can be found in [32].

We investigate in depth urban water networks and we develop several methods for improving security and resilience in presence of faults and attacks. The objective of this work is to develop an efficient and practical sensor placement method for the failure detection and localization in water networks. We formulate the problem as the minimum test cover problem (MTC) with the objective of selecting the minimum number of sensors required to uniquely identify and localize pipe failure events. First, we summarize a single-level sensing model and discuss an efficient fast greedy approach for solving the MTC problem. Simulation results on benchmark test networks demonstrate the efficacy of the fast greedy algorithm. Second, we develop a multi-level sensing model that captures additional physical features of the network, such as the distance between the failure event and the sensor location. Our sensor placement approach using MTC extends to the multi-level sensing model and an improved identification performance is obtained via reduced number of sensors (in comparison to single-level sensing model). In particular, we investigate the bi-level sensing model to illustrate the efficacy of employing multi-level sensors for the identification of failure events. Finally, we suggest extensions of our approach for the deployment of heterogeneous sensors in water networks by exploring the trade-off between cost and performance (measured in terms of the identification score of pipe/link failures). Our approach has been presented in [33].

In addition, our work focuses on the optimal sensor placement problem for identification of pipe failure locations in large-scale urban water systems. The problem involves selecting the minimum number of sensors such that every pipe failure can be uniquely localized. This problem can be viewed as a minimum test cover (MTC) problem, which is NP-hard. We consider two approaches to obtain approximate solutions to this problem. In the first approach, we transform the MTC problem to a minimum set cover (MSC) problem and use the greedy algorithm that exploits the submodularity property of the MSC problem to compute the solution to the MTC problem. In the second approach, we develop a new augmented greedy algorithm for solving the MTC problem. This approach does not require the transformation of the MTC to MSC. Our augmented greedy algorithm provides in a significant computational improvement while guaranteeing the same approximation ratio as the first approach. We propose several metrics to evaluate the performance of the sensor placement designs. Finally, we present detailed computational experiments for a number of real water distribution networks [34]. Figure 12 illustrates sensor placement for a real world water distribution network.

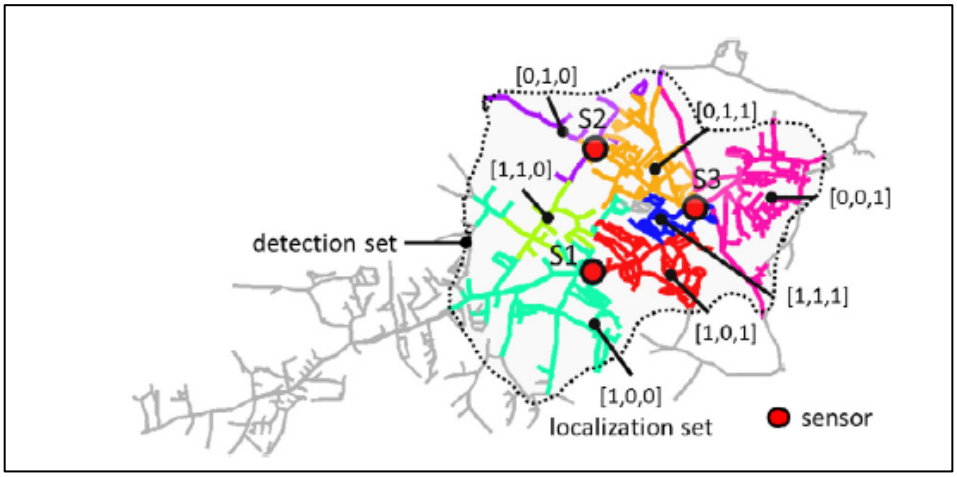

Figure 12. Water Network and Example of Detection and Localization for Three Sensors

Extending the work described above, we study the sensor placement problem in urban water networks that maximizes the localization of pipe failures given that some sensors give incorrect outputs [9]. False output of a sensor might be the result of degradation in sensor's hardware, software fault, or might be due to a cyber attack on the sensor. Incorrect outputs from such sensors can have any possible values which could lead to an inaccurate localization of a failure event. We formulate the optimal sensor placement problem with erroneous sensors as a set multi-cover problem, which is NP-hard, and then discuss a polynomial time heuristic to obtain efficient solutions. In this direction, we first examine the physical model of the disturbance propagating in the network as a result of a failure event, and outline the multi-level sensing model that captures several event features. Second, using a combinatorial approach, we solve the problem of sensor placement that maximizes the localization of pipe failures by selecting m sensors out of which at most e give incorrect outputs. We propose various localization performance metrics, and numerically evaluate our approach on a benchmark and a real water distribution network. Finally, using computational experiments, we study relationships between design parameters such as the total number of sensors, the number of sensors with
errors, and extracted signal features.

In networked systems, monitoring devices such as sensors are typically deployed to monitor various target locations. Targets are the points in the physical space at which events of some interest, such as random faults or attacks, can occur. Most often, these devices have limited energy supplies, and they can operate for a limited duration. As a result, energy efficient monitoring of various target locations through a set of monitoring devices with limited energy supplies is a crucial problem in networked systems. In this work, we study optimal scheduling of monitoring devices to maximize network coverage for detecting and isolating events on targets for a given network lifetime [35]. The monitoring devices considered could remain active only for a fraction of the overall network lifetime. We formulate the problem of scheduling of monitoring devices as a graph labeling problem, which unlike other existing solutions, allow us to directly utilize the underlying network structure to explore the trade-off between coverage and network lifetime. In this direction, first we propose a greedy heuristic to solve the graph labeling problem, and then provide a game-theoretic solution to achieve optimal graph labeling. Moreover, the proposed setup can be used to simultaneously solve the scheduling and placement of monitoring devices, which yields improved performance as compared to separately solving the placement and scheduling problems. Finally, we illustrate our results on various networks, including real-world water distribution networks.

Efficient emergency response is a related problem and a major concern in densely populated urban areas. Numerous techniques have been proposed to allocate emergency responders to optimize response times, coverage, and incident prevention. Effective response depends, in turn, on effective prediction of incidents occurring in space and time, a problem which has also received considerable prior attention. We formulate a non-linear mathematical program maximizing expected incident coverage, and propose a novel algorithmic framework for solving this problem. In order to aid the optimization problem, we propose a novel incident prediction mechanism. Prior art in incident prediction does not generally consider incident priorities which are crucial in optimal dispatch, and spatial modeling either considers each discretized area independently, or learns a homogeneous model. We bridge these gaps by learning a joint distribution of both incident arrival time and severity, with spatial heterogeneity captured using a hierarchical clustering approach. Moreover, our decomposition of the joint arrival and severity distributions allows us to

independently learn the continuous-time arrival model, and subsequently use a multinomial logistic regression to capture severity, conditional on incident time. We use real traffic accident and response data from the urban area around Nashville, USA, to evaluate the proposed approach, showing that it significantly outperforms prior art as well as the real dispatch method currently in use [36].

Further, we focus on the question of distributed control of electricity distribution networks in the face of security attacks to Distributed Energy Resources (DERs). Our attack model includes strategic manipulation of DER set-points by an external hacker to induce a sudden compromise of a subset of DERs connected to the network. We approach the distributed control design problem in two stages. In the first stage, we model the attacker-defender interaction as a Stackelberg game. The attacker (leader) disconnects a subset of DERs by sending them wrong set-point signals. The distribution utility (follower) response includes Volt-VAR control of non-compromised DERs and load control. The objective of the attacker (resp. defender) is to maximize (resp. minimize) the weighted sum of the total cost due to loss of frequency regulation and the cost due to loss of voltage regulation. In the second stage, we propose a distributed control (defender response) strategy for each local controller such that, if sudden supply-demand mismatch is controllers automatically respond based on their respective observations of local fluctuations in voltage and frequency. This strategy aims to achieve diversification of DER functions in the sense that each uncompromised DER node either contributes to voltage regulation (by contributing reactive power) or to frequency regulation (by contributing active power). We illustrate the effectiveness of this control strategy on a benchmark network. Our results are reported in [37].

4.2 Science of decentralized security

Science of decentralized security aims to develop a framework that will enable reasoning about the security of all the integrated constituent CPS components. Complex traffic networks, for example, include a number of controlled intersections, and, commonly, multiple districts or municipalities. The result is that the overall traffic control problem is extremely complex computationally. Moreover, given that different municipalities may have distinct, non-aligned, interests, traffic light controller design is inherently decentralized, a consideration that is almost entirely absent from related literature. Both complexity and decentralization have great bearing both on the quality of the traffic network overall, as well as on its security.

Although adaptive, state-aware strategies can offer tremendous gains in traffic control efficiency, they expose an attack surface that can be exploited to substantially increase congestion. For example, a common kind of adaptive control logic utilizes vehicle queue lengths in each direction, with light switching between red and green as a function of relative queue lengths. While such feedback-based switching can significantly increase efficiency, it also exposes a vulnerability of controllers to attacks on sensors from which queue length information is derived. An additional consideration which is crucial in modern transportation networks is that traffic lights on the network are often managed by multiple actors (e.g., municipalities).

We demonstrate how we can efficiently explore these challenges in multi-intersection closed-loop traffic light control, where (1) traffic light controllers take into account relative queue lengths to determine red-green state of the traffic lights at an intersection; (2) controllers for all lights must be designed to work jointly so as to optimize overall traffic network performance; (3) sensors feeding data into the controllers are vulnerable to DoS attacks; and 4) intersections can be partitioned among a set of players, with own goals pertaining to congestion within their local municipal region, which are in general misaligned with global interests of the entire traffic network (Figure 9). Details of the theoretical approach can be found in [38].

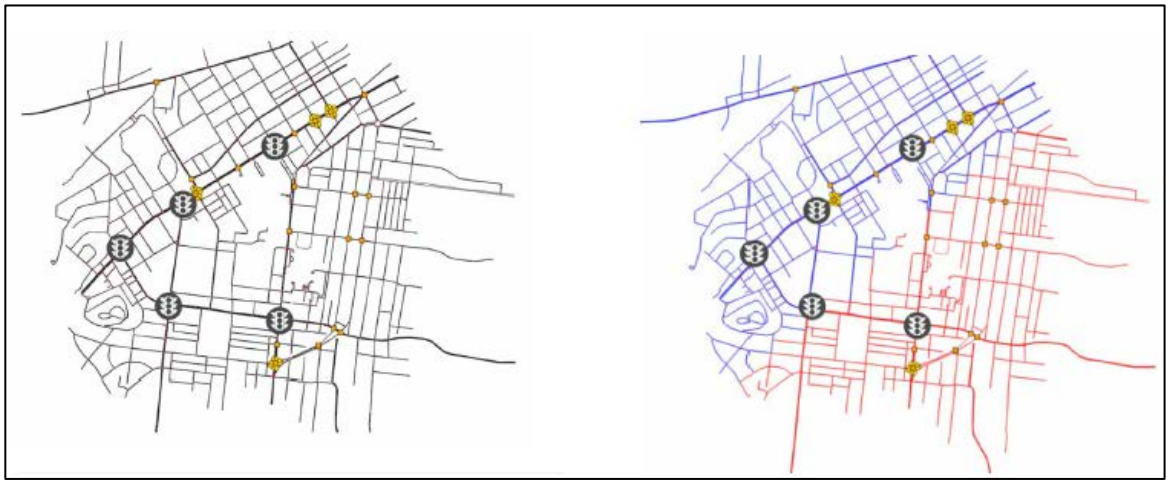

Figure 13. Centralized and Decentralized Control Traffic Light Control

We consider both of these issues in a dynamic traffic network. First, we propose an effective local search algorithm to efficiently design system-wide control logic for a collection of intersections. Second, we propose a game theoretic (Stackelberg game) model of traffic network security in which an attacker can deploy denial-of-service attacks on sensors, and develop a resilient control algorithm to mitigate such threats. Finally, we propose a game theoretic model of decentralization, and investigate this model both in the context of baseline traffic network design, as well as resilient design accounting for attacks. Our methods are implemented and evaluated using a simple traffic network scenario in SUMO [39].

Another challenge that needs to be address for all integrated constituent CPS components is integrity assurance. Assuring communication integrity is a central problem in security of CPS. However, overhead costs associated with cryptographic primitives used towards this end introduce significant practical implementation challenges for resource-bounded systems, such as cyber-physical systems. For example, many control systems are built on legacy components that are computationally limited but have strict timing constraints. If integrity protection is a binary decision, it may simply be infeasible to introduce into such systems; without it, however, an adversary can forge malicious messages, which can cause significant physical or financial harm. We propose a formal game-theoretic framework for optimal stochastic message authentication, providing provable integrity guarantees for resource-bounded systems based on an existing message authentication scheme [40][41]. Based on our threat model and objectives, we formulate stochastic message authentication as an attacker–defender game, considering both short term and long-term conflicts. –We study the adversary's best responses characterize when the adversary can be deterred from attacking, discuss finding an optimal defense when deterrence is impossible, study the defender's best responses, and characterize when the game has a Nash equilibrium. We propose two schemes for implementing stochastic authentication in practice and demonstrate their viability using experiments.

Typically, system and security design of various CPS components are performed by different agents, having varying and often conflicting interests. Our goal is to develop this framework of multidefender games, and associated computational tools to address security holistically, accounting for incentives of all the parties. Current Stackelberg security game models primarily

focus on isolated systems in which only one defender is present, despite being part of a more complex system with multiple players. However, many real systems such as transportation networks and the power grid exhibit interdependencies among targets and, consequently, between decision makers jointly charged with protecting them. To understand such multidefender strategic interactions present in security scenarios, the authors investigate security games with multiple defenders. Unlike most prior analyses, they focus on situations in which each defender must protect multiple targets, so even a single defender's best response decision is, in general, nontrivial. Considering interdependencies among targets, the authors develop a novel mixed-integer linear programming formulation to compute a defender's best response, and approximate Nash equilibria of the game using this formulation. Their analysis shows how network structure and the probability of failure spread determine the propensity of defenders to over- or underinvest in security. Our work is reported in [42].

Stackelberg game models of security have received much attention, with a number of approaches for computing Stackelberg equilibria in games with a single defender protecting a collection of targets. In contrast, multi-defender security games have received significantly less attention, particularly when each defender protects more than a single target. We fill this gap by considering a multidefender security game, with a focus on theoretical characterizations of equilibria and the price of anarchy. We develop the analysis of three models of increasing generality, two in which each defender protects multiple targets. In all models, we find that the defenders often have the incentive to overprotect the targets, at times significantly. Additionally, in the simpler models, we find that the price of anarchy is unbounded, linearly increasing both in the number of defenders and the number of targets per defender. Surprisingly, when we consider a more general model, this results only in a “corner” case in the space of parameters; in most cases, however, the price of anarchy converges to a constant when the number of defenders increases. Our approach is presented in [43].

Stackelberg security games have been widely deployed in recent years to schedule security resources. An assumption in most existing security game models is that one security resource assigned to a target only protects that target. However, in many important real-world security scenarios, when a resource is assigned to a target, it exhibits protection externalities:
that is, it also protects other “neighboring” targets. We investigate such Security Games with Protection Externalities (SPEs). Figure 14 illustrates protection externalities in real-world scenarios: (a) visible areas of surveillance cameras in a building; (b) reaction area of mobile security units.

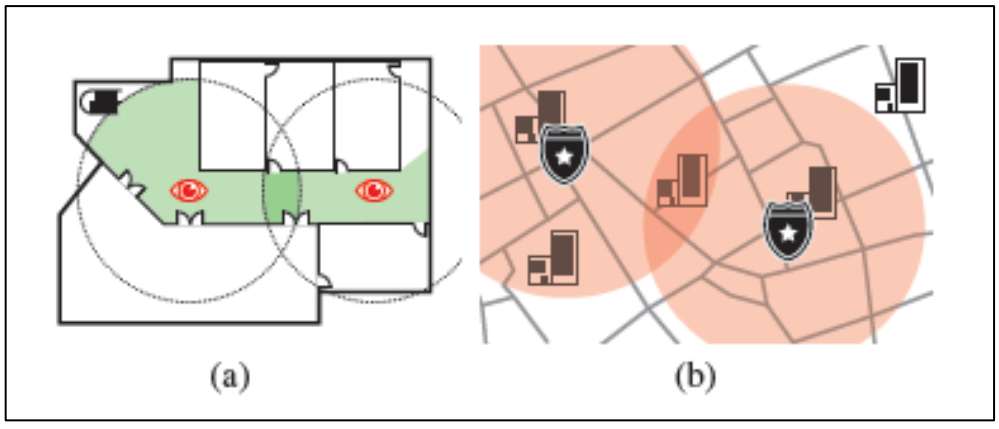

Figure 14. Protection Externalities in Real-world Scenarios

First, we demonstrate that computing a strong Stackelberg equilibrium for an SPE is NP-hard, in contrast with traditional Stackelberg security games which can be solved in polynomial time. On the positive side, we propose a novel column generation based approach—CLASPE—to solve SPEs. CLASPE features the following novelties: 1) a novel mixed-integer linear programming formulation for the slave problem; 2) an extended greedy approach with a constant-factor approximation ratio to speed up the slave problem; and 3) a linear-scale linear programming that efficiently calculates the upper bounds of target-defined subproblems for pruning. Our experimental evaluation demonstrates that CLASPE enable us to scale to realistic-sized SPE problem instances. Further details can be found in [44].

Stackelberg games have been widely used to model interactions between attackers and defenders in a broad array of security domains. One related approach involves plan interdiction, whereby a defender chooses a subset of actions to block (remove), and the attacker constructs an optimal plan in response. In previous work, this approach has been introduced in the context of Markov decision processes (MDPs). The key challenge, however, is that the state space of MDPs grows exponentially in the number of state variables. We propose a novel scalable MDP interdiction framework which makes use of factored representation of state, using a parity function basis for representing a value function over a Boolean space. We demonstrate that our approach is significantly more scalable than prior art, while resulting in near-optimal interdiction decisions [45].

Most existing models of Stackelberg security games ignore the underlying topology of the space in which targets and defense resources are located. As a result, allocation of resources is restricted to a discrete collection of exogenously defined targets. However, in many practical security settings, defense resources can be located on a continuous plane. Better defense solutions could therefore be potentially achieved by placing resources in a space outside of actual targets (e.g., between targets). To address this limitation, we propose a model called Security Game on a Plane (SGP) in which targets are distributed on a 2-dimensional plane, and security resources, to be allocated on the same plane, protect targets within a certain effective distance [46]. We investigate the algorithmic aspects of SGP. We find that computing a strong Stackelberg equilibrium of an SGP is NP-hard even for zero- sum games, and these are inapproximable in general. On the positive side, we find an exact solution technique for general SGPs based on an existing approach, and develop a PTAS (polynomial-time approximation scheme) for zero-sum SGP to more fundamentally overcome the computational obstacle. Our experiments demonstrate the value of considering SGP and effectiveness of our algorithms.

In recent years, we have seen a large number of cyber-incidents, which demonstrated how difficult it is to prevent cyber-breaches when facing determined and sophisticated attackers. In light of this, it is clear that defenders need to look beyond the first lines of defense and invest not only into prevention, but also into limiting the impact of cyber-breaches. Thus, an effective cyber-defense must combine proactive defense, which aims to block anticipated attacks, with reactive defense, which responds to and mitigates perceived attacks (e.g., isolating and shutting down compromised components). However, planning defensive actions in anticipation of and in response to strategic attacks is a challenging problem. Prior work has introduced a number of game-theoretic security models for planning defensive actions, such as Stackelberg security games, but these models do not address the overarching problem of proactive and reactive defenses in sufficient detail. To bridge this gap, we introduce a modeling approach for building high-resolution multi-stage security games. We describe several approaches for modeling proactive and

reactive defenses, consider key modeling choices and challenges, and discuss finding optimal defense policies [47]. With our study, we aim to lay conceptual foundations for developing realistic models of cyber-security that researchers and practitioners can use for effective cyber-defense.

4.3 Reliable and practical reasoning about secure computation/communication

4.3.1 Actor Networks

As computation spreads from computers to networks of computers, and migrates into cyberspace, it ceases to be globally programmable, but it remains programmable indirectly and partially: network computations cannot be controlled, but they can be steered by imposing local constraints on network nodes. The tasks of "programming" global behaviors through local constraints belong to the area of security. The ``program particles'' that assure that a system of local interactions leads towards some desired global goals are called security protocols. They are the software connectors of modern, world-wide software systems. As computation spreads beyond cyberspace, into physical and social spaces, new security tasks and problems arise. As computer networks are extended by nodes with physical sensors and controllers, including the humans, and interlaced with social networks, the engineering concepts and techniques of computer security blend with the social processes of security, that evolved since the dawn of mankind. These new connectors for computational and social software require a new ``discipline of programming'' of global behaviors through local constraints. Since the new discipline seems to be emerging from a combination of established models of security protocols with older methods of procedural programming, we use the name *procedures* for these new connectors that generalize protocols.

We develop *actor-networks* as a formal model of computation in heterogenous networks of computers, humans and their devices, where these new procedures run; and we introduce Procedure Derivation Logic *(PDL)* as a framework for reasoning about security in actor-networks. Our goal is to contribute towards a formal framework for *reliable* and *practical* reasoning about computation and communication

The diverse views of science of security have opened up several alleys towards applying the methods of science to security. We pursue a different kind of connection between science and security. We explore the idea that security is not just a suitable subject for science, but that the process of security is also similar to the process of science. This similarity arises from the fact that both science and security depend on the methods of inductive inference. Because of this dependency, a scientific theory can never be definitely proved, but can only be disproved by new evidence, and improved into a better theory. Because of the same dependency, every security claim and method has a lifetime, and always eventually needs to be improved. In this general framework of security-as-science, we explore the ways to apply the methods of scientific induction in the process of trust. The process of trust building and updating is viewed as hypothesis testing. We propose to formulate the trust hypotheses by the methods of algorithmic learning, and to build more robust trust testing and vetting methodologies on the solid foundations of statistical inference. Further details can be found in [48].

While computer programs and logical theories begin by declaring the concepts of interest, be it as data types or as predicates, network computation does not allow such global declarations, and requires concept mining and concept analysis to extract shared semantics for different network nodes. Powerful semantic analysis systems have been the drivers of nearly all paradigm shifts on the web. In categorical terms, most of them can be described as bicompletions of enriched matrices, generalizing the Dedekind-MacNeille-style completions from posets to suitably enriched categories. Yet it has been well known for more than 40 years that ordinary categories themselves

in general do not permit such completions. Armed with this new semantical view of Dedekind-MacNeille completions, and of matrix bicompletions, we take another look at this ancient mystery [49]. It turns out that simple categorical versions of the limit superior and limit inferior operations characterize a general notion of Dedekind-MacNeille completion, that seems to be appropriate for ordinary categories, and boils down to the more familiar enriched versions when the limits inferior and superior coincide. This explains away the apparent gap among the completions of ordinary categories, and broadens the path towards categorical concept mining and analysis, opened in previous work.

4.3.2 Malware

CPS are very often exposed and vulnerable to direct access by an attacker. Malware attacks pose a unique threat to systems that most often are not designed to operate in the presence of malware attacks. We leverage machine learning techniques to detect, classify, and take action on malware using existing data sets. We use a software framework for experimentation called Security-oriented Active Learning Testbed (SALT) which supports experimentation on time series data including modeling limited availability of labeling resources and feature acquisition options. We also focus on the application of active learning to the domain of CPS. Active learning addresses strategies for the allocation of finite resources such as human labelling efforts and feature extraction.

We examine the problem of aggregating the results of multiple anti-virus (AV) vendors' detectors into a single authoritative ground-truth label for every binary. To do so, we adapt a well-known generative Bayesian model that postulates the existence of a hidden ground truth upon which the AV labels depend. We use training based on Expectation Maximization for this fully unsupervised technique. We evaluate our method using 279,327 distinct binaries from VirusTotal, each of which appeared for the first time between January 2012 and June 2014. Our evaluation shows that our statistical model is consistently more accurate at predicting the future-derived ground truth than all unweighted rules of the form “k out of n" AV detections. In addition, we evaluate the scenario where partial ground truth is available for model building. We train a logistic regression predictor on the partial label information. Our results show that as few as a 100 randomly selected training instances with ground truth are enough to achieve 80% true positive rate for 0.1% false positive rate [50]. In comparison, the best unweighted threshold rule provides only 60% true positive rate at the same false positive rate.

We study a dataset of billions of program binary files that appeared on 100 million computers over the course of 12 months, discovering that 94% of these files were present on a single machine. Though malware polymorphism is one cause for the large number of singleton files, additional factors also contribute to polymorphism, given that the ratio of benign to malicious singleton _les is 80:1. The huge number of benign singletons makes it challenging to reliably identify the minority of malicious singletons. We present a large-scale study of the properties, characteristics, and distribution of benign and malicious singleton files. We leverage the in- sights from this study to build a classifier based purely on static features to identify 92% of the remaining malicious singletons at a 1**:**4% percent false positive rate, despite heavy use of obfuscation and packing techniques by most malicious singleton files that we make no attempt to de-obfuscate. Finally, we demonstrate robustness of our classifier to important classes of automated evasion attacks [51].

4.3.3 Adversarial Machine Learning

The success of machine learning, particularly in supervised settings, has led to numerous attempts to apply it in adversarial settings such as spam and malware detection. The core challenge in this class of applications is that adversaries are not static data generators, but make a deliberate effort to evade the classifiers deployed to detect them. We investigate both the problem of modeling the objectives of such adversaries, as well as the algorithmic problem of accounting for rational, objective-driven adversaries. In particular, we demonstrate severe shortcomings of feature reduction in adversarial settings using several natural adversarial objective functions, an observation that is particularly pronounced when the adversary is able to substitute across similar features (for example, replace words with synonyms or replace letters in words). We offer a simple heuristic method for making learning more robust to feature cross-substitution attacks. We then develop a more general approach based on mixed-integer linear programming with constraint generation, which implicitly trades off overfitting and feature selection in an adversarial setting using a sparse regularizer along with an evasion model. Our approach is the first method for combining an adversarial classification algorithm with a very general class of models of adversarial classifier evasion. We show that our algorithmic approach significantly outperforms state-of-the-art alternatives. Our results are reported in [52].

Cheap ubiquitous computing enables the collection of massive amounts of personal data in a wide variety of domains. Many organizations aim to share such data while obscuring features that could disclose identities or other sensitive information. Much of the data now collected exhibits weak structure (e.g., natural language text) and machine learning approaches have been developed to identify and remove sensitive entities in such data. Learning-based approaches are never perfect and relying upon them to sanitize data can leak sensitive information as a consequence. However, a small amount of risk is permissible in practice, and, thus, our goal is to balance the value of data published and the risk of an adversary discovering leaked sensitive information. We model data sanitization as a game between 1) a publisher who chooses a set of classifiers to apply to data and publishes only instances predicted to be non-sensitive and 2) an attacker who combines machine learning and manual inspection to uncover leaked sensitive entities (e.g., personal names). We introduce an iterative greedy algorithm for the publisher that provably executes no more than a linear number of iterations, and ensures a low utility for a resource-limited adversary. Moreover, using several real world natural language corpora, we illustrate that our greedy algorithm leaves virtually no automatically identifiable sensitive instances for a state-of-the-art learning algorithm, while sharing over 93% of the original data, and completes after at most 5 iterations. Our results are reported [53].

Recommendation and collaborative filtering systems are important in modern information and e-commerce applications. As these systems are becoming increasingly popular in the industry, their outputs could affect business decision making, introducing incentives for an adversarial party to compromise the availability or integrity of such systems. We introduce a data poisoning attack on collaborative filtering systems. We demonstrate how a powerful attacker with full knowledge of the learner can generate malicious data so as to maximize his/her malicious objectives, while at the same time mimicking normal user behavior to avoid being detected. While the complete knowledge assumption seems extreme, it enables a robust assessment of the vulnerability of collaborative filtering schemes to highly motivated attacks. We present efficient solutions for two popular factorization-based collaborative filtering algorithms: the alternative minimization

formulation and the nuclear norm minimization method. Finally, we test the effectiveness of our proposed algorithms on real-world data and discuss potential defensive strategies [54].

The effectiveness of supervised learning techniques has made them ubiquitous in research and practice. In high-dimensional settings, supervised learning commonly relies on dimensionality reduction to improve performance and identify the most important factors in predicting outcomes. However, the economic importance of learning has made it a natural target for adversarial manipulation of training data, which we term *poisoning attacks*. Prior approaches to dealing with robust supervised learning rely on strong assumptions about the nature of the feature matrix, such as feature independence and sub-Gaussian noise with low variance. We propose an integrated method for robust regression that relaxes these assumptions, assuming only that the feature matrix can be well approximated by a low-rank matrix. Our techniques integrate improved robust low-rank matrix approximation and robust principle component regression, and yield strong performance guarantees. Moreover, we experimentally show that our methods significantly outperform state of the art both in running time and prediction error. Our results are presented in [55].

Cheap ubiquitous computing enables the collection of massive amounts of personal data in a wide variety of domains. Many organizations aim to share such data while obscuring features that could disclose personally identifiable information. Much of this data exhibits weak structure (e.g., text), such that machine learning approaches have been developed to detect and remove identifiers from it. While learning is never perfect, and relying on such approaches to sanitize data can leak sensitive information, a small risk is often acceptable. Our goal is to balance the value of published data and the risk of an adversary discovering leaked identifiers. We model data sanitization as a game between 1) a publisher who chooses a set of classifiers to apply to data and publishes only instances predicted as non-sensitive and 2) an attacker who combines machine learning and manual inspection to uncover leaked identifying information. We introduce a fast iterative greedy algorithm for the publisher that ensures a low utility for a resource-limited adversary. Moreover, using five text data sets we illustrate that our algorithm leaves virtually no automatically identifiable sensitive instances for a state-of-the-art learning algorithm, while sharing over 93 percent of the original data, and completes after at most five iterations. Our results are reported in [56].

4.3.4 Privacy in Cyber-Physical Systems

As our ground transportation infrastructure modernizes, the large amount of data being measured, transmitted, and stored motivates an analysis of the privacy aspect of these emerging cyber-physical technologies. In this paper, we consider privacy in the routing game, where the origins and destinations of drivers are considered private. This is motivated by the fact that this spatiotemporal information can easily be used as the basis for inferences for a person's activities. More specifically, we consider the differential privacy of the mapping from the amount of flow for each origin-destination pair to the traffic flow measurements on each link of a traffic network. We use a stochastic online learning framework for the population dynamics, which is known to converge to the Nash equilibrium of the routing game. We analyze the sensitivity of this process and provide theoretical guarantees on the convergence rates as well as differential privacy values for these models. We confirm these with simulations on a small example. Our results are reported in [57].

The quantity of personal data gathered by service providers via our daily activities continues to grow at a rapid pace. The sharing, and the subsequent analysis of, such data can support a wide range of activities, but concerns around privacy often prompt an organization to transform the data

to meet certain protection models (e.g., k-anonymity or "differential privacy). These models, however, are based on simplistic adversarial frameworks, which can lead to both under- and over-protection. For instance, such models often assume that an adversary attacks a protected record exactly once. We introduce a principled approach to explicitly model the attack process as a series of steps. Specifically, we engineer a factored Markov decision process (FMDP) to optimally plan an attack from the adversary's perspective and assess the privacy risk accordingly. The FMDP captures the uncertainty in the adversary's belief (e.g., the number of identified individuals that match the de-identified data) and enables the analysis of various real world deterrence mechanisms beyond a traditional protection model, such as a penalty for committing an attack. We present an algorithm to solve the FMDP and illustrate its efficiency by simulating an attack on publicly accessible U.S. census records against a real identified resource of over 500,000 individuals in a voter registry [58]. Our results demonstrate that while traditional privacy models commonly expect an adversary to attack exactly once per record, an optimal attack in our model may involve exploiting none, one, or more individuals in the pool of candidates, depending on context.

4.3.5 Optimal Personalized Filtering

To penetrate sensitive computer networks, attackers can use spear phishing to sidestep technical security mechanisms by exploiting the privileges of careless users. In order to maximize their success probability, attackers have to target the users that constitute the weakest links of the system. The optimal selection of these target users takes into account both the damage that can be caused by a user and the probability of a malicious e-mail being delivered to and opened by a user. Since attackers select their targets in a strategic way, the optimal mitigation of these attacks requires the defender to also personalize the e-mail filters by taking into account the users' properties. In this work, we assume that a learned classifier is given and propose strategic per-user filtering thresholds for mitigating spear-phishing attacks. We formulate the problem of filtering targeted and non-targeted malicious e-mails as a Stackelberg security game. We characterize the optimal filtering strategies and show how to compute them in practice. Finally, we evaluate our results using two real-world datasets and demonstrate that the proposed thresholds lead to lower losses than nonstrategic thresholds [59].

Despite decades of effort to combat spam, unwanted and even malicious emails, such as phish which aim to deceive recipients into disclosing sensitive information, still routinely find their way into one's mailbox. To be sure, email filters manage to stop a large fraction of spam emails from ever reaching users, but spammers and phishers have mastered the art of filter evasion, or manipulating the content of email messages to avoid being filtered. We present a unique behavioral experiment designed to study email filter evasion. Our experiment is framed in somewhat broader terms: given the widespread use of machine learning methods for distinguishing spam and non-spam, we investigate how human subjects manipulate a spam template to evade a classification-based filter. We find that adding a small amount of noise to a filter significantly reduces the ability of subjects to evade it, observing that noise does not merely have a short-term impact, but also degrades evasion performance in the longer term. Moreover, we find that greater coverage of an email template by the classifier (filter) features significantly increases the difficulty of evading it. This observation suggests that aggressive feature reduction—a common practice in applied machine learning—can actually facilitate evasion. In addition to the descriptive analysis of behavior, we develop a synthetic model of human evasion behavior which closely matches observed behavior and effectively replicates experimental findings in simulation. Our results are reported in [60].

Spear-phishing attacks pose a serious threat to sensitive computer systems, since they sidestep technical security mechanisms by exploiting the carelessness of authorized users. A common way to mitigate such attacks is to use e-mail filters which block e-mails with a maliciousness score above a chosen threshold. Optimal choice of such a threshold involves a tradeoff between the risk from delivered malicious emails and the cost of blocking benign traffic. A further complicating factor is the strategic nature of an attacker, who may selectively target users offering the best value in terms of likelihood of success and resulting access privileges. Previous work on strategic threshold-selection considered a single organization choosing thresholds for all users. In reality, many organizations are potential targets of such attacks, and their incentives need not be well aligned. We therefore consider the problem of strategic threshold-selection by a collection of independent self-interested users. We characterize both Stackelberg multi-defender equilibria, corresponding to short-term strategic dynamics, as well as Nash equilibria of the simultaneous game between all users and the attacker, modeling long-term dynamics, and exhibit a polynomial-time algorithm for computing short-term (Stackelberg) equilibria. We find that while Stackelberg multi-defender equilibrium need not exist, Nash equilibrium always exists, and remarkably, both equilibria are unique and socially optimal. Our results are reported in [61].

4.3.6 Human in the Loop Visual Analytics for Exploring Document Collections

The project investigated a new approach for the visual analysis of large collections of documents by utilizing inverse computations in Machine Learning (ML) Black Box Models from semantic interactions. The results show that we are able to leverage user suggestions to provide feedback into a visual analytics system and modify the internal system model of the black box ML algorithm, enabling it to perform document clustering that is more aligned with the needs of the user. This allows analysts who are not experts in ML to better tailor ML systems toward making the better discoveries. We believe that approaches such as this will be increasingly important as a means to enable humans to "mentor" ML systems, enabling humans to keep pace with the exponential growth of document data.

4.4 Evaluation and experimentation

CPS are inherently heterogeneous not only in terms of their components but also in terms of essential design requirements. Besides functional properties, CPS are subject to system level requirements, such as safety and security. To date, security and resilience have been considered as largely disjoint aspects of CPS design. In the model-based framework, the software and system implementation task is to derive the code and models such that they satisfy functional, safety, and security as well as the resource constraint requirements imposed by the underlying platform. To address these challenges, the project developed a suite of methods evaluation and experimentation. Large-scale CPS security and resilience experimentation involve a multitude of variables that cannot be practically explored to arrive at conclusive results. To address these challenges, we developed a suite of simulation and hardware-in-the-loop testbeds and conducted comprehensive evaluation of both system and security properties.

4.4.1 Integrated Simulation Testbed for Security and Resilience

The objectives of this work are to analyze the cybersecurity risks, propose resilient monitoring and control mechanisms, and evaluate their effectiveness as well as their performance impact on system operations. We consider security and resilience as system properties emerging from the intersection of system dynamics and the computing architecture. This integrative view allows pursuing cross-domain tradeoffs and system-security co-design. Resilient dynamics generalize functional performance by augmenting design concerns to attain robustness against faults and cyber attacks. The effects of failures and intrusions are modeled as uncertainties and casted as adversarial games. We investigate how to efficiently solve these games and design efficient defense strategies against worst-case attacks. More importantly, we develop a modeling and simulation integration platform (Figure 15) that enables *evaluation of resilience of CPS in the presence of cyber attacks based on attacker-defender games using simulations of sufficient fidelity* [39].

The platform incorporates (1) realistic models of cyber and physical components and their interactions, (2) cyber attack models that focus on the impact of attacks to CPS behavior and operation, and (3) operational scenarios that can be used for evaluation of cybersecurity risks. Further, it allows the evaluation of performance impact and assessment of resilient monitoring and control algorithms. The main innovation of our approach is that research processes and results are documented as executable software models, simulations, and generated data that support cybersecurity analysis and design in a quantifiable manner. We evaluate the approach using three case studies: (1) Vulnerability analysis of transportation networks to traffic signal tampering, (2) resilient sensor selection for forecasting traffic flow, and (3) decentralized resilient traffic signal control in the presence of denial-of-service attacks. It should be noted that transportation systems are treated as any other network critical infrastructure, and hence, the proposed approach can be directly applied to other similar classes of CPS.

The platform enables in-depth experimental evaluation of security and resilience that is necessary for developing the scientific foundations and technology. Theoretical analysis is accompanied by large amounts of experimental work and empirical observations use realistic CPS models and integrated simulations of tightly coupled cyber and physical components. Additionally, the platform allows the design and execution of controlled experiments of large-scale CPS by configuring the system and attack models. The main idea is to untangle poorly understood interactions and improve understanding by simulating real-world CPS. Simulation of such complex systems can lead to new knowledge by predicting how an assemblage of heterogeneous components will behave and discover what are the implications of the assumptions imposed on the system.

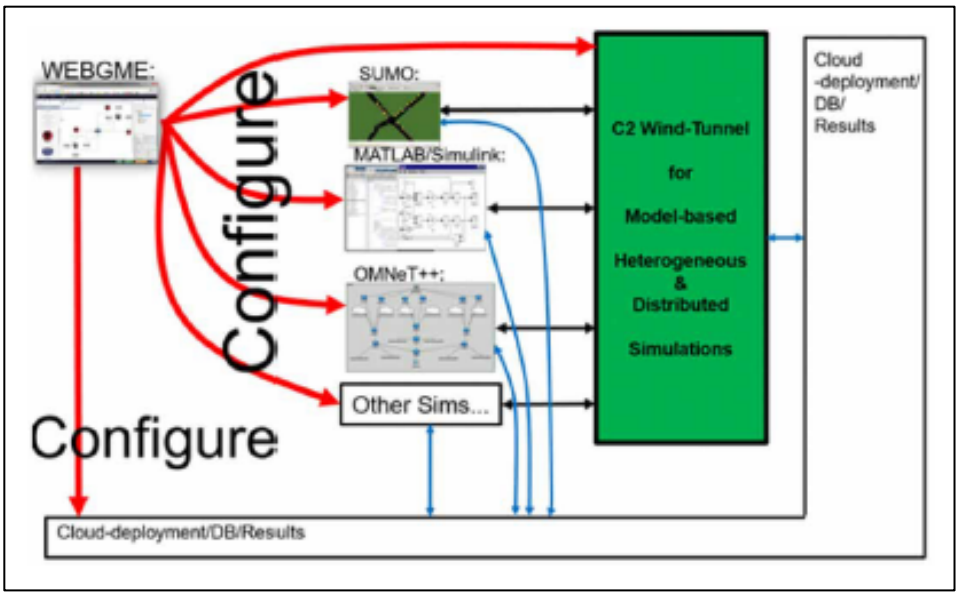

Figure 15. SURE Platform Architecture

Analyzing security and resilience of these complex ICSs is difficult as it requires evaluating many subsystems and factors in an integrated manner. Integrated simulation of physical systems and communication network can provide an underlying framework for creating a reusable and configurable testbed for such analyses. Using a model-based integration approach and the IEEE High-Level Architecture (HLA)-based distributed simulation software; we have created a testbed for integrated evaluation of large-scale cyber-physical systems (CPS). Our tested supports web-based collaborative metamodeling and modeling of CPS system and experiments and a cloud computing environment for executing integrated networked co-simulations. A modular and extensible cyber-attack library enables validating the CPS under a variety of configurable cyber-attacks, such as DDoS and integrity attacks. Hardware-in-the-loop simulation is also supported along with several hardware attacks. Further, a scenario modeling language allows modeling of alternative paths for what-if scenarios. These capabilities make our testbed well suited for analyzing security and resilience of CPS. In addition, the web-based modeling and cloud-hosted execution infrastructure enables one to exercise the entire testbed using simply a web-browser, with integrated live experimental results display. The integrated testbed is presented in [62].

As shown in Figure 15, WebGME is used to model the system and configure the simulation tools used for different aspects of the modeled CPS. The different simulation tools are integrated for timed data exchange and time-synchronized execution using the framework Command and Control Wind Tunnel (C2WT). C2WT supports integration of a variety of simulation tools such as Matlab/Simulink, OMNeT++, CPNTools, SUMO, TrainDirector, and Gridlab-D. We host the C2WT integration and execution platform in a cloud environment using OpenStack. Additionally, when integrated simulations are executed in the C2WT platform the live experiment results are populated in a streaming InfluxDB database. These results are then queried by the WebGME tool to display live charts of the experiment results as the simulations are executed. One key feature of our testbed is that it can execute many variations of the simulation experiments in parallel, limited only by the available cloud resources. This is crucial for analyzing CPS, which requires analyzing many configurations, parameters, and workflows.

Modeling of the CPS, both in WebGME and in C2WT, is based on Model-Integrated Computing (MIC) that focuses on formally representing the system components, their interactions,

and rules for composition and configuration. A metamodel is a Domain-Specific Modeling Language (DSML) designed for a particular application domain. For example, the metamodel captures the modeling of CPS systems and scenarios with a focus on security and resilience analysis, whereas in C2WT it focusses on distributed simulation integration.

One of the fundamental aspects of the testbed is the capability to analyze the CPS under a variety of attacks. Further, the testbed allows detailed scenario modeling for performing what-if analyses using Courses-of-Actions (COAs) The attack library and COAs use an integrated hardware-in-the-loop testbed to run hardware elements and deploy attacks in the hardware.

Many attacks and phenomena are not analyzable in simulations due to performance reasons or unavailability of high-fidelity simulation models, thus requiring use of real, physical hardware for performing Hardware-in-the-loop (HIL) simulation. However, the HIL platform must be connected with a distributed simulation platform, which provides scalability and time-synchronization needed for complex distributed simulations. Our HIL platform is comprised of two parts: a hardware-in-the-loop testbed and the C2WT distributed simulation environment. This allows for taking advantage of both the scalability of the C2WT with the fine-tuned ability to analyze CPS controller behavior on real emulated hardware consistent with the platforms deployed in the field.

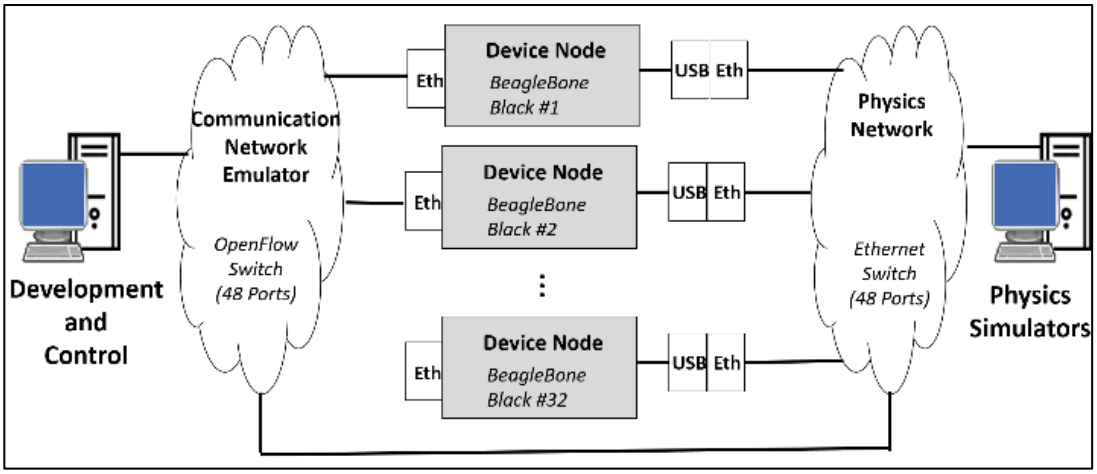

Figure 16. Hardware-In-The-Loop Platform

As illustrated in Figure 16, the HIL testbed is comprised of 5 different components. These include the development system where CPS control software is developed, the CPS nodes which consist of embedded computing boards consistent with operating platforms in the field, a software defined networking interface that enables controlling various communication parameters and protocols through the network, a Physics simulator serving as the physical plant, and a physical network connecting embedded computing nodes with the simulator interface.

The HIL testbed also has an interface (see Figure 17) to connect the emulated software in the hardware with the simulated software in C2WT. The interface protocol communication utilizes Google Protocol Buffers – a language and platform neutral extensible mechanism for serializing data – for formatting custom messages, and the ZeroMQ API for transmitting and receiving messages throughout the network. The integration interface has two components, viz. HIL Proxy and HIL gateway. The role of HIL Proxy is to serve as the interface between the embedded computing nodes on the HIL testbed and the simulation environment. As such, this proxy mechanism receives sensor information from each HIL node as well as sends custom commands

to each respective node to adjust behavior. The role of the HIL gateway is to serve as an interface between C2WT simulators and the controller code in the HIL testbed. This can include communication between controllers and sensors defined in C2WT with controllers in the HIL testbed, as well as receiving controller commands from respective HIL nodes. Since the simulator interface is defined in C2WT as a simulation, the gateway is additionally responsible for serving as an interface for HIL node controllers to interact with the physical plant simulator. Further, two message types are used, viz. HIL messages and interface messages. HIL messages correspond to internal messages in the HIL testbed such as internal controller communications or commands. However, when communication needs to be established with the C2WT environment such as obtaining sensor values or sending actuation commands to the simulator, interface messages are utilized for transmission.

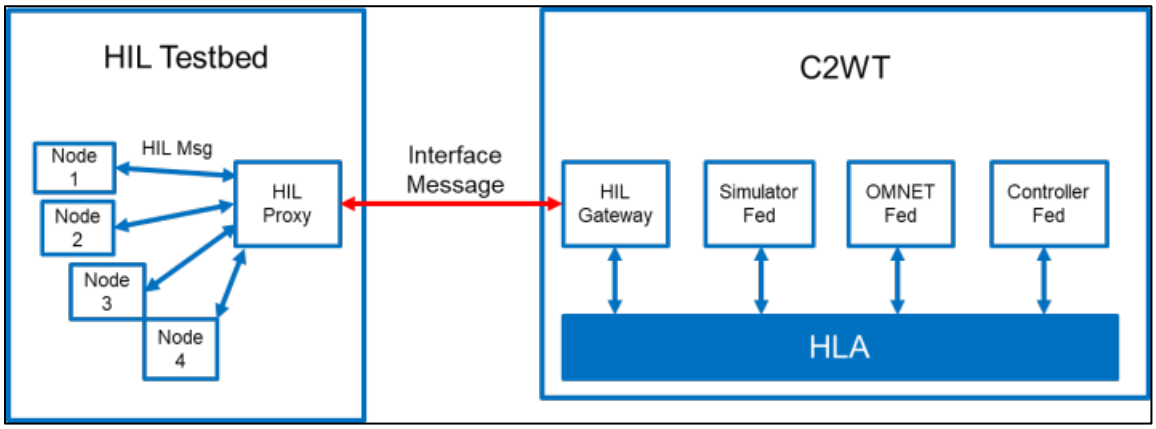

Figure 17. HIL Integration with C2WT

Because the physical and the cyber aspects of the system are tightly coupled, cyber-security is important for ensuring the system functions properly and safely. However, the effects of a cyber-attack on the whole system may be difficult to determine, analyze, and therefore detect and mitigate. Our work aims at developing a hardware-in-the-loop (HIL) testbed shown in Figure 18 which connects physics simulators with networked embedded computers for performing analysis of cyber-attack effects in networked CPS [63]. We configure the HIL CPS Testbed to emulate a road traffic controller network for the Vanderbilt area in Nashville, TN. A primary intersection servicing the area's main hospital is chosen as a high-value target for attack. The attack on the intersection is a Distributed Denial-of-Service (DDoS) attack from many fake inductive loop sensors to the traffic light controller. The execution behavior of the traffic light controller software is measured together with the road traffic at the target intersection in the simulated road network. Simulated road traffic into and out of a level-1 trauma center was increased due to the successful DDoS of one traffic light controller at a high-value intersection. DDoS on the software component controlling the intersection directly led to denial of service of the road traffic by the light, and thus denial of entry to and exit from the primary regional hospital. This work demonstrates the ability to determine how a change in software behavior due to an attack causes a change in physical system behavior. Such results can for example be used for improved threat level analysis.

We also develop a system simulation framework for the design and performance evaluation of complex wireless cyber-physical systems. We describe the simulator architecture and the specific developments that are required to simulate cyber-physical systems relying on multi-

channel, multihop mesh networks [64]. We introduce realistic and efficient physical layer models and a system simulation methodology, which provides statistically significant performance evaluation results with low computational complexity. The capabilities of the proposed framework are illustrated in the example of WirelessHART, a centralized, real-time, multi-hop mesh network designed for industrial control and monitor applications.

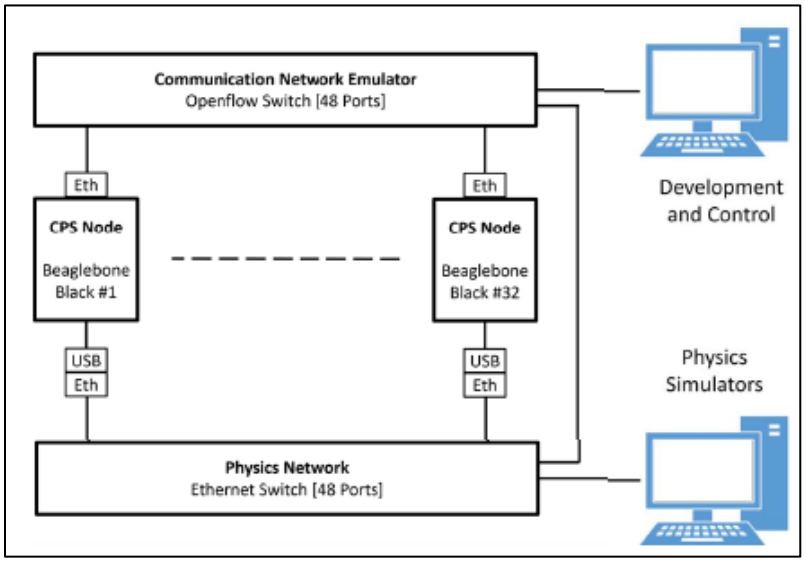

Figure 18. Architecture of HIL for Evaluation of CPS Security

Design-time analysis and verification of distributed real-time embedded systems necessitates the modeling of the time-varying performance of the network and comparing that to application requirements. Earlier work has shown how to build a system network model that abstracted away the network's physical medium and protocols which govern its access and multiplexing. We show how to apply a network medium channel access protocol, such as Time-Division Multiple Access (TDMA), to our network analysis methods and use the results to show that the abstracted model without the explicit model of the protocol is valid. Details of the approach can be found in [65].

In modern networked control applications, confidentiality and integrity are important features to address in order to prevent against attacks. Moreover, many networked control systems employ Time-Triggered (TT) Architectures that provide mechanisms enabling the exchange of precise and synchronous messages. TT systems have computation and communication constraints, and with the aim to enable secure communications in the network, it is important to evaluate the computational and communication overhead of implementing secure communication mechanisms. Our work presents an extended empirical study of the effects of adding a Hash-based Message Authentication (HMAC) to TT networked control systems [66]. The performance evaluation for the computation and communication overhead is presented for a wired and a wireless TT network. The respective theoretical analysis, experimental validation and performance evaluation is provided, as well as the communication overhead for adding more nodes in the network. As an example, an automotive application is used, and the results show that it is feasible to implement a secure communication mechanism without interfering with the existing automotive controller execution times. The empirical evaluation and respective results can be used for evaluating the performance impact of security mechanisms, and thus, for design of secure wired and wireless TT networked control systems.

4.4.2 Integrated Moving Target Defense and Control Reconfiguration for Securing CPS

With the increasingly connected nature of Cyber-Physical Systems (CPS), new attack vectors are emerging. Normally, an adversary will use memory corruption attacks to achieve manipulation of the cyber sub-system, leading to alteration of the physical dynamics. As such, the compromise of safety-critical systems opens the gates for attackers to exfiltrate sensitive data, or inappropriately control actuation. It is critical to shift the CPS security focus into a more proactive approach, aimed at creating more resilient architectures. For example, Automobiles today are extremely complex systems of systems, consisting of several hundred electronic digital components with over a million lines of code. The internal automotive network consists of a series of multiple communication buses such as CAN, LIN, FlexRay, and MOST. Due to the traditionally standalone design of vehicle architectures, the communication and controller designs prioritize functionality and cost over cybersecurity. Additionally, with the majority of software being written in legacy code, vast numbers of vulnerabilities are potentially included. With the introduction of external interfaces such as infotainment centers and telematics systems, adversaries now have remote avenues in place to access the internal vehicle network.

The two primary instances of memory corruption attacks are code injection and code reuse attacks. Code injection attacks exploit existing input vulnerabilities for injecting a custom designed instruction payload that can be executed by control flow redirection. For code injection attacks to be successful, the adversary has to rely on knowing the native instruction set architecture of the target machine. Code reuse attacks on the other hand leverage existing code by diverting control flow to legitimate code segments allowing the adversary to achieve his/her malicious goal even in the cases where directly injecting code is not possible. One of the most popular examples of this type of attack is return oriented programming (ROP) in which case existing code gadgets are chained together to form a program that can execute malicious behavior. One of the most common memory corruption vulnerabilities in legacy code leading to code injection and code reuse attacks is the buffer overflow. Buffer overflow vulnerabilities allow attackers to input data longer than designed, overflowing into adjacent areas, and if properly designed, can be leveraged to redirect control flow.

Moving Target Defense (MTD) aim to prevent legacy vulnerabilities by dynamically changing system properties. Compared to traditional defense mechanisms which focus on identifying malware, and suspicious communications, MTD focus on decreasing the reconnaissance knowledge of the adversary with the goal of minimizing the probability of successful reverse engineering, vulnerability discovery, and exploit deployment. Two MTD techniques utilized in our work are Instruction Set Randomization (ISR), and Address Space Randomization (ASR). ISR is a technique for protecting against code injection attacks by changing the binary instruction set architecture to a randomized version that is not known. ASR is a technique for mainly protecting against code reuse attacks by introducing diversity in the various segments of a program to make external memory access unpredictable. ASR can be implemented at various granularities including course grained, and fine grained, while also having the ability to be customized to protect the most critical memory segments.

In the CPS domain, even when successfully protecting against cyber-attacks, it is equally as important to maintain reliable, safe, and predictable operation of the system. With ISR and ASR deployed, code injection and code reuse attacks will be thwarted, but an invalid instruction or invalid address access exception will be generated, leading to program termination. In this sense, it is not acceptable for a safety-critical system to stop functioning, as any loss of availability can

lead to unsafe actuation causing physical damage. As such, there has to be recovery mechanisms in place to keep the system up and running at all times, even when under a cyber-attack campaign.

To address the difficulty of guaranteeing system availability, while preventing code injection and code reuse attacks, we have developed a security architecture that includes an AES 256 ISR implementation for protecting against code injection attacks, combined with a fine grained ASR implementation for protecting against the relative, and direct control flow redirection necessary for code reuse attacks. Our security architecture consists of three stages including attack protection (randomize, derandomize), detection, and recovery. The main CPS challenge addressed in this paper is protecting system integrity during cyber-attacks, while maintaining system availability with safe and reliable operation. Our work makes the following contributions:

- We develop a CPS security architecture for providing secure protections against code injection and code reuse attacks by utilizing AES 256 ISR, and function level fine grained ASR.
- We incorporate control reconfiguration into our security architecture for maintaining system availability in the event of a cyber-attack.
- We implement a hardware in the loop testbed prototype using a combination of off-the-shelf embedded computing hardware and open source simulation software for analyzing the effects of cyber-attacks and our security architecture in CPS environments consistent with deployment settings.
- We present an autonomous vehicle case study to demonstrate the effectiveness of our security architecture in limiting the physical impact of code injection, and code reuse attacks on driving safety.

The system model is shown in Figure 19. The model includes: a sensor cluster, actuator cluster, driving controller, telematics control unit (TCU), remote function actuator (RFA), and radio-frequency identification (RFID) sensor. The sensor cluster provides critical data representing the current state of the vehicle such as the speed, position on the track, and heading. The actuator cluster provides the ability to manipulate vehicular behavior such as steering and acceleration. The driving controller is responsible for performing computation based on the provided sensor cluster input, and outputting commands to the actuation cluster. Both the TCU, and RFA are responsible for providing the external interface for the vehicle. The TCU monitors the various metrics of the system, transmitting data to a remote operating station for maintenance and emergency purposes. The RFA is responsible for determining the presence of a key fob for allowing the vehicle to be turned on.

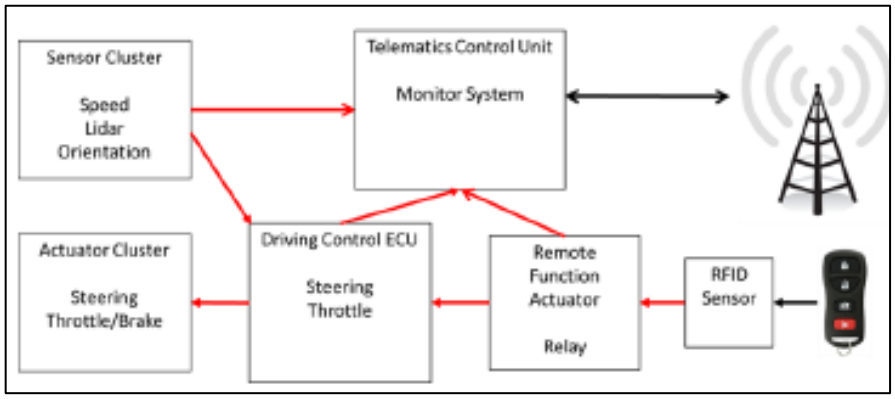

Figure 19: System Model

Figure 20 shows the security architecture. The key components are the (1) Configuration Manager (CM) that oversees, customizes, and adjusts the operation of the various operating components, (2) CPS Controllers which control the physical plant, (3) Dynamic Binary Translator (DBT) which provides a sandboxed runtime environment for each CPS controller, and (4) Operating System Kernel which handles the task scheduling and exception detection. We assume that each CPS controller in our architecture may be vulnerable to cyber-attacks by the adversary, but the remaining components are secure. Our security architecture is designed with the goal of keeping the CPS controller from becoming compromised by the attacker. Our results are reported in [67][68].

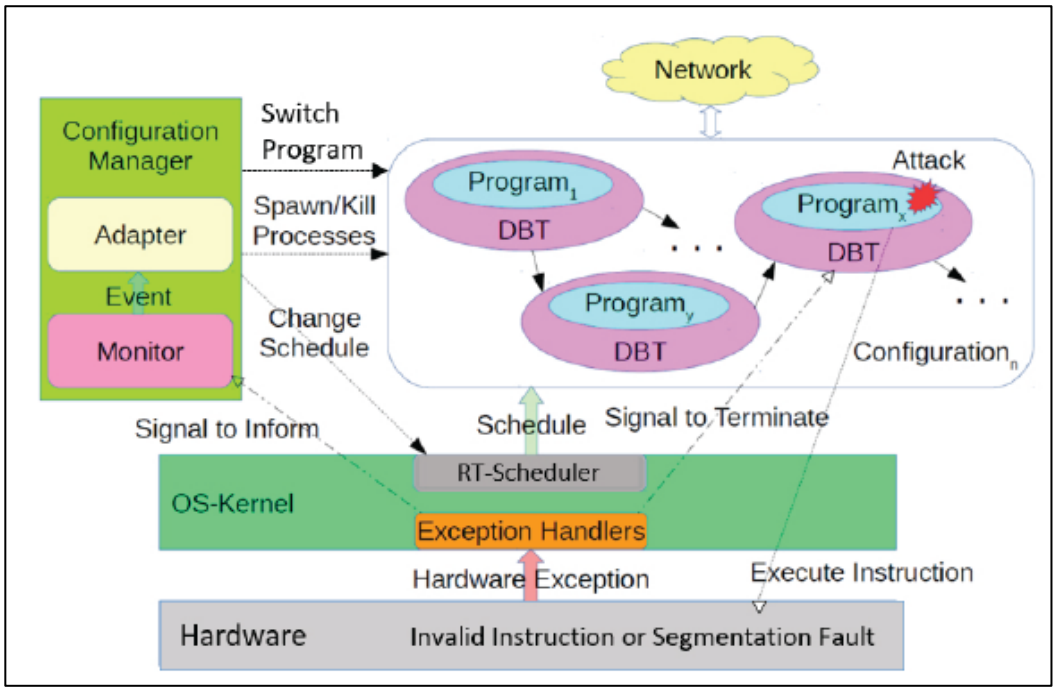

Figure 20. Security Architecture

Verifying that hardware design implementations adhere to specifications is a time intensive and sometimes intractable problem due to the massive size of the system's state space. Formal methods techniques can be used to prove certain tractable specification properties; however, they are expensive, and often require subject matter experts to develop and solve. Nonetheless, hardware verification is a critical process to ensure security and safety properties are met, and encapsulates problems associated with trust and reliability. For complex designs where coverage of the entire state space is unattainable, prioritizing regions most vulnerable to security or reliability threats would allow efficient allocation of valuable verification resources. Stackelberg security games model interactions between a defender, whose goal is to assign resources to protect a set of targets, and an attacker, who aims to inflict maximum damage on the targets after first observing the defender's strategy. In equilibrium, the defender has an optimal security deployment strategy, given the attacker's best response. We apply this Stackelberg security framework to synthesized hardware implementations using the design's network structure and logic to inform defender valuations and verification costs. The defender's strategy in equilibrium is thus

interpreted as a prioritization of the allocation of verification resources in the presence of an adversary. We demonstrate this technique on several open-source synthesized hardware designs [69].

4.5 Education and outreach

This section summarizes the activities related to the education and outreach of project.

Summer CPS Security Camps

The main activity for education and outreach is the organization of summer camps on CPS and CPS security for high school students. The goals of the program are to help students understand CPS, increase diversity and interest in CPS, improve teaching methods for delivering content in CPS curricula, and prepare students for related coursework in college. Specifically, the program used simulation-based and experimental platforms of autonomous vehicles to demonstrate how fundamental concepts from computer science and engineering are integrated to develop operational and dependable systems. The long-term goals of this initiative include establishing (or building) a pipeline that targets highly-qualified students who are interested in CPS-related engineering disciplines.

In 2017, the summer camp took place at Vanderbilt University on July 31st - August 4th. Invited participants were 15 MLK Jr Magnet School rising HS Seniors and Juniors. The curriculum outlines a hands-on workshop on learning robotics for middle and high school students. Robotics is explored through group discussions, instructor guided lessons and experimentation which presents engineering as a broad domain and attempts to bridge classroom lessons to real world applications. The course uses the Arduino based robot MiniQ 2WD which is an affordable and accessible platform for schools. MATLAB and Simulink are utilized as an intuitive learning environment and as means to implement code on robots. Videos are provided for students to help demonstrate concepts as well as videos for instructors to learn about the Simulink models before carrying out each lesson. The curriculum maps to skills found in the Common Core Standards and the Next Generation Science Standards. The detailed curriculum and all the materials used in the summer camp are available at https://cps-vo.org/group/CPSSummerCamp17.

In 2018, the CPS Summer Camps were a 5-day experience for highly-qualified students interested in the growing field of CPS Security. Two camps were organized at the Vanderbilt University Institute for Software Integrated Systems during the weeks of June 4 and June 11 with a total of 24 students from grades 7-12. Students were recruited from local area high schools. In addition, we invited one teacher from each high school to participate. The goals of the program are to help students understand CPS Security, increase diversity and interest in CPS, improve teaching methods for delivering content in CPS curricula, and prepare students for related coursework in college. The detailed curriculum and all the materials used in the summer camp are available at https://cps-vo.org/group/CPSScamp18.

In 2019, we continued offering the summer. Two 2019 summer camps were held at Vanderbilt University Institute for Software Integrated Systems in Nashville, Tennessee. In addition to the summer camps for students, a Teacher Camp was held June 3-7, 2019 at Vanderbilt University Institute for Software Integrated Systems in Nashville, Tennessee. Fifteen teachers attended the camp, learn the curriculum, and then offer the curriculum to high school students. More information can be found at https://cps-vo.org/group/CPSSteachercamp19.

The camp offers a team-based, hands-on curriculum for teaching cybersecurity, distributed programming, and robotics. The learning goals include gains in computational thinking, demonstrated ability in computer programming, increased awareness and understanding of

cybersecurity in cyber physical systems (CPS), and motivation for students to learn more. This class utilizes WiFi enabled robots and NetsBlox - a visual block-based programming environment specifically designed to teach distributed programming and computer networking Participants learned to program Roboscape robots in order to accomplish simple tasks. Roboscape is a collaborative, networked robotics environment that introduces key ideas in complex systems and cyber-physical systems (CPS) science. RoboScape relies on the institute's prior innovative work creating NetsBlox (https://netsblox.org/) a networked, visual programming environment specifically targeted at introducing students to distributed computation and computer networking.

Undergraduate and Graduate Education

The project has substantial impact on education activities that include the development of CPS-related courses and course materials in all the institutions participated in the project at both undergraduate and graduate level. In addition, leveraging the education activities in the project, Vanderbilt University has established an interdisciplinary master's degree in cyber-physical systems (PI Koutsoukos is the program director; further details can be found at https://engineering.vanderbilt.edu/academics/m_eng/CPS/index.php). In addition, the project involved post-doctoral scholars graduate students, and undergraduate students, working in the areas of resilience and security of CPS.

Dissemination and Outreach

The project generated over 70 papers published in journals, conferences, workshops, and as book chapters. The main papers are listed in the references of this report. The main web site of the project is hosted by the Cyber-Physical Systems Virtual Organization (CPS-VO) and includes a comprehensive list of publications and presentations (http://cps-vo.org/group/sure). Software developed in the project have been disseminated also using the CPS-VO and other web resources.

In addition to the publications, the participants gave many keynote and plenary presentations in conferences as well as invited presentations in various research institutions. The PIs also held positions in multiple editorial boards of journals related to CPS, organized special issues, and held leadership positions in conference organization (e.g., ICCPS and HotSoS). Further, they organized multiple workshops related to CPS. Finally, several collaborations have resulted from this project that include additional research projects between the participating institutions as well as outreach to other universities and research organizations.

Community outreach includes the support of the planning and logistics for the Computational Cybersecurity in Compromised Environment (C3E) workshop and the continued development of the Science of Security Virtual Organization collaboration portal. Further details can be found in the quarterly technical reports.

5. CONCLUSIONS

The objective in this SURE project was to develop the Science of Secure and Resilient Cyber-Physical Systems. In the area of hierarchical coordination and control, the project developed languages and tools for threat and attack modeling in CPS in order to understand the impact of cyber-attacks to system performance and operation, methods to improve resilience in CPS techniques for resilient monitoring of CPS, resilient distributed control and coordination algorithms in the presence of malicious agents, and algorithms for improving the resilience and security of CPS flow networks such transportation, water, and power networks. In the area of decentralized security, the project developed methods for secure decentralized control, investigated in depth multidefender games considering real-world scenarios, and developed a framework for high-resolution multi-stage security games. Our work investigated practical reasoning techniques for addressing security problems that affect CPs including actor-networks as a formal model of computation in heterogenous networks of computers, malware detection using active learning strategies, adversarial machine learning, privacy in CPS, optimal personalized filtering, and methods for human-in-the-loop visual analytics for exploring document collections. The project developed a suite of simulation and hardware-in-the-loop testbeds and conducted comprehensive evaluation of both system and security properties. Finally, outreach and education activities include summer camps on CPS and CPS security for high school students and community outreach supporting multiple conference and workshops and the science of security virtual organization (SOS-VO).

6. REFERENCES

[1] Mark Yampolskiy, Mark, Peter Horvath, Xenofon Koutsoukos, Yuan Xue, and Janos Sztipanovits. “CP-ADL: Cyber-Physical Attack Description Language.” *International Journal of Critical Infrastructure Protection*. Volume 8, 40-52. January 2015.

[2] Goncalo Martins, Sajal Bhatia, Xenofon Koutsoukos, Keith Stouffer, Chee Yee Tang, and Richar Candell."Towards a Systematic Threat Modeling Approach for Cyber-Physical Systems", *3rd International Symposium on Resilient Cyber Systems (ISRCS 2015)*, Philadelphia, PA, August 18-20, 2015.

[3] Bradley Potteiger, Goncalo Martins, and Xenofon Koutsoukos. “Software and Attack Centric Integrated Threat Modeling for Quantitative Risk Assessment”, *2016 Symposium and Bootcamp on the Science of Security (HotSoS'16)*, Pittsburg, PA, April 19-21, 2016.

[4] Aron Laszka, Bradley Potteiger, Yevgeniy Vorobeychik, Saurabh Amin, and Xenofon Koutsoukos. ‘Vulnerability of Transportation Networks to Traffic-Signal Tampering”, *ACM/IEEE 7th International Conference on Cyber-Physical Systems*, Vienna, Austria, April 12-14, 2016.

[5] Amin Ghafouri and Xenofon Koutsoukos. "Vulnerability of Fixed-Time Control of Signalized Intersections to Cyber-Tampering", *9th International Symposium on Resilient Control Systems*, Chicago, IL, Aug. 16-18, 2016.

[6] Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Resilient observation selection in adversarial settings." In *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 7416-7421. IEEE, 2015.

[7] Amin Ghafouri, Aron Laszka, Abhishek Dubey, and Xenofon Koutsoukos. “Optimal Detection of Fault Traffic Sensors Used in Route Planning”, *2nd International Workshop on Science of Smart City Operations and Platforms Engineering (SCOPE’17)*, Pittsburg, PA, April 21, 2017.

[8] Amin Ghafouri, Aron Laszka, Xenofon Koutsoukos. "Application-Aware Anomaly Detection of Sensor Measurements in Cyber Physical Systems", Sensors,18(8):2448, 2018.

[9] Waseem Abbas, Lina Sela Perelman, Saurabh Amin, and Xenofon Koutsoukos. “Resilient Sensor Placement for Fault Localization in Water Distribution Networks", *ACM/IEEE 8th International Conference on Cyber-Physical Systems (ICCPS 2017)*, Pittsburgh, PA, April 18 - 21, 2017.

[10] Aron Laszka, Waseem Abbas, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Synergic security for smart water networks: Redundancy, diversity, and hardening." In *Proceedings of the 3rd International Workshop on Cyber-Physical Systems for Smart Water Networks*, pp. 21-24. ACM, 2017.

[11] Aron Laszka, Waseem Abbas, Yevgeniy Vorobeychik and Xenofon Koutsoukos. "Synergistic Security for the Industrial Internet of Things: Integrating Redundancy, Diversity, and Hardening", *2018 IEEE International Conference on Industrial Internet (ICII)*, Seattle, WA, Oct. 21-23, 2018.

[12] Aron Laszka, Waseem Abbas, Yevgeniy Vorobeychik and Xenofon Koutsoukos. "Integrating Redundancy, Diversity, and Hardening to Improve Security of Industrial Internet of Things", Cyber-Physical Systems. Published online: 10 June 2019.

[13] Aron Laszka, Waseem Abbas, S. Shankar Sastry, Yevgeniy Vorobeychik, and Xenofon Koutsoukos."Optimal thresholds for intrusion detection systems", Proceedings of the

Symposium and Bootcamp on the Science of Security (HotSos '16). Pittsburgh, PA, April 19-21, 2016.

[14] Amin Ghafouri, Waseem Abbas, Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. “Optimal Thresholds for Anomaly-Based Intrusion Detection in Dynamical Environments”, *2016 Conference on Decision and Game Theory for Security (GameSec 2016),* New York, NY, Nov. 2 - 4, 2016.

[15] Amin Ghafouri, Waseem Abbas, Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. “A game-theoretic approach for selecting optimal time-dependent thresholds for anomaly detection”, *Autonomous Agents and Multi-Agent Systems*. 33(4), 430-456, July 2019.

[16] Aron Laszka, Waseem Abbas, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. “Detection and Mitigation of Attacks in Transportation Networks as a Multi-Stage Security Game”, *Computers & Security.* Vol. 87, Nov. 2019.

[17] Waseem Abbas, Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Scheduling Intrusion Detection Systems in Resource-Bounded Cyber-Physical Systems", *First ACM Workshop on Cyber-Physical Systems Security & Privacy (CPS-SPC'15),* Denver, CO, October 16, 2015. (Held in conjunction with CCS 2016.)

[18] Aron Laszka, Waseem Abbas, and Xenofon Koutsoukos. “Scheduling Battery-Powered Sensor Networks for Minimizing Detection Delays", *IEEE Communication Letters.* 21(4), 789-792, April 2017.

[19] Waseem Abbas, Sajal Bhatia, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Guarding Networks Through Heterogeneous Mobile Guards", *2015 American Control Conference (ACC 2015),* Chicago, IL, July 1-3, 2015.

[20] Waseem Abbas, Aron Laszka, and Xenofon Koutsoukos. "Resilient wireless sensor networks for cyber-physical systems." *Cyber-Physical System Design with Sensor Networking Technologies; Zeadally, S., Jabeur, N., Eds* (2016): 239-267.

[21] Waseem Abbas and Xenofon Koutsoukos. "Efficient Complete Coverage Through Heterogeneous Sensing Nodes", *IEEE Wireless Communication Letters.* Vol. 4, No. 1, 14-17, February 2015.

[22] Qie Hu, Young Hwan Chang, and Claire J. Tomlin. "Secure estimation for unmanned aerial vehicles against adversarial cyber attacks." *30th Congress of the International Council of the Aeronautical Sciences*, September 25-30, 2016.

[23] Nika Haghtalab, Aron Laszka, Ariel D. Procaccia, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Monitoring Stealthy Diffusion", *IEEE International Conference on Data Mining (ICDM 2015),* Atlantic City, NJ, November 14 - 17, 2015.

[24] Nika Haghtalab, Aron Laszka, Ariel D. Procaccia, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Monitoring Stealthy Diffusion", *Knowledge and Information Systems.* 52(3), 657-685, September 2017.

[25] Waseem Abbas, Aron Laszka, and Xenofon Koutsoukos. “Graph-Theoretic Approach for Increasing Participation in Social Sensing”, *The 2nd International Workshop on Social Sensing (SocialSens 2017),* Pittsburg, PA, April 21, 2017.

[26] Waseem Abbas, Aron Laszka and Xenofon Koutsoukos, "Improving Network Connectivity and Robustness Using Trusted Nodes with Application to Resilient Consensus," in *IEEE Transactions on Control of Network Systems.* 54(3), 2036-2048, Dec. 2018.

[27] Waseem Abbas, Aron Laszka, Yevgeniy Vorobeychik and Xenofon Koutsoukos. “Improving Network Connectivity Using Trusted Nodes and Edges", *The 2017 American Control Conference (ACC 2017)*, Seattle, WA, May 24-26, 2017.
[28] Heath LeBlanc and Xenofon Koutsoukos. “Resilient Consensus and Synchronization of Networked Multi-Agent Systems”, *IEEE Transactions on Control of Networked Systems*. 5(3), 1219 - 1231, Sept. 2018.
[29] Jiani Li and Xenofon Koutsoukos. “Resilient Distributed Diffusion for Multi-Task Estimation”, *International Conference on Distributed Computing in Sensor Systems (DCOSS 2018)*, New York, NY, June 18 - 20, 2018.
[30] Jiani Li, Waseem Abbas, and Xenofon Koutsoukos. “Resilient Distributed Diffusion in Networks with Adversaries”, *IEEE Transactions on Signal and Information Processing over Networks*. Accepted for publication.
[31] Gabor Karsai, Xenofon Koutsoukos, Himanshu Neema, Peter Volgyesi, and Janos Sztipanovits. "Transportation Networks." In *Cyber Resilience of Systems and Networks*, pp. 425-446. Springer, Cham, 2019.
[32] Mathieu Dahan and Saurabh Amin. "Network flow routing under strategic link disruptions." In *2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 353-360. IEEE, 2015.
[33] Waseem Abbas, Lina S. Perelman, Saurabh Amin, and Xenofon Koutsoukos. "An Efficient Approach to Fault Identification in Urban Water Networks Using Multi-Level Sensing", *2nd ACM International Conference on Embedded Systems for Energy Efficient Buildings (ACM BuildSys 2015)*, Seoul, North Korea, November 4-5, 2015.
[34] Lina Sela Perelman, Waseem Abbas, Xenofon Koutsoukos, and Saurabh Amin. “Sensor placement for fault location identification in water networks: a minimum test cover approach”, *Automatica.* Volume 72, 166-176, October 2016.
[35] Waseem Abbas, Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Scheduling Resource-Bounded Monitoring Devices for Event Detection and Isolation in Networks", *IEEE Transactions on Network Science and Engineering*, 5(1), 65-78, Jan. - Mar. 2018.
[36] Ayan Mukhopadhyay, Yevgeniy Vorobeychik, Abhishek Dubey, and Gautam Biswas. "Prioritized allocation of emergency responders based on a continuous-time incident prediction model." In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems*, pp. 168-177. International Foundation for Autonomous Agents and Multiagent Systems, 2017.
[37] Devendra Shelar, Jairo Giraldo, and Saurabh Amin. "A distributed strategy for electricity distribution network control in the face of der compromises." In *2015 54th IEEE conference on decision and control (CDC)*, pp. 6934-6941. IEEE, 2015.
[38] Jian Lou and Yevgeniy Vorobeychik. “Decentralization and security in dynamic traffic light control”. In *Symposium and Bootcamp on Science of Security*, 2016.
[39] Xenofon Koutsoukos, Gabor Karsai, Aron Laszka, Himanshu Neema, Bradley Potteiger, Peter Volgyesi, Yevgeniy Vorobeychik, and Janos Sztipanovits. "SURE: A Modeling and Simulation Integration Platform for Evaluation of SecUre and REsilient Cyber-Physical Systems", *Proceedings of the IEEE*, 106(1), 93-112, January 2018.
[40] Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Integrity Assurance in Resource-Bounded Systems through Stochastic Message Authentication", *Symposium and Bootcamp on the Science of Security (HotSoS 2015)*, UIUC, April 21-22, 2015.

[41] Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. “A game-theoretic approach for integrity assurance in resource-bounded systems", *International Journal of Information Security.* 17(2), 221 - 242, April 2018.
[42] J. Lou, A. M. Smith and Y. Vorobeychik, "Multidefender Security Games," in *IEEE Intelligent Systems*, vol. 32, no. 1, pp. 50-60, Jan.-Feb. 2017.
[43] Jian Lou and Yevgeniy Vorobeychik. "Equilibrium analysis of multi-defender security games." In *Twenty-Fourth International Joint Conference on Artificial Intelligence*. 2015.
[44] Jiarui Gan, Bo An, and Yevgeniy Vorobeychik. "Security games with protection externalities." In *Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015.
[45] Swetasudha Panda, and Yevgeniy Vorobeychik. "Near-optimal interdiction of factored MDPs." In *Conference on Uncertainty in Artificial Intelligence*. 2017.
[46] Jiarui Gan, Bo An, Yevgeniy Vorobeychik, and Brian Gauch. “Security games on a plane”. In *AAAI Conference on Artificial Intelligence,* 2017.
[47] Aron Laszka, Xenofon Koutsoukos, and Yevgeniy Vorobeychik. "Towards High-Resolution Multi-Stage Security Games." In *Proactive and Dynamic Network Defense,* pp. 139-161. Springer, Cham, 2019.
[48] Dusko Pavlovic. "Towards a science of trust." In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, p. 3. ACM, 2015.
[49] Toshiki Kataoka and Dusko Pavlovic, “Towards Concept Analysis in Categories: Limit Inferior as Algebra, Limit Superior as Coalgebra.” In Proceedings of the 6^{th} *Conference on Algebra and Coalgebra in Computer Science* (CALCO), Nijmegen, Netherlands, 2015.
[50] Alex Kantchelian, Michael Carl Tschantz, Sadia Afroz, Brad Miller, Vaishaal Shankar, Rekha Bachwani, Anthony D. Joseph, and J. Doug Tygar. "Better malware ground truth: Techniques for weighting anti-virus vendor labels." In *Proceedings of the 8th ACM Workshop on Artificial Intelligence and Security*, pp. 45-56. ACM, 2015.
[51] Bo Li, Kevin Roundy, Chris Gates, and Yevgeniy Vorobeychik. "Large-scale identification of malicious singleton files." In *Proceedings of the Seventh ACM on Conference on Data and Application Security and Privacy*, pp. 227-238. ACM, 2017.
[52] Bo Li and Yevgeniy Vorobeychik. "Feature cross-substitution in adversarial classification." In *Advances in neural information processing systems*, pp. 2087-2095. 2014.
[53] Bo Li, Yevgeniy Vorobeychik, Muqun Li, and Bradley Malin. "Iterative classification for sanitizing large-scale datasets." In *2015 IEEE International Conference on Data Mining*, pp. 841-846. IEEE, 2015.
[54] Bo Li, Yining Wang, Aarti Singh, and Yevgeniy Vorobeychik. "Data poisoning attacks on factorization-based collaborative filtering." In *Advances in neural information processing systems*, pp. 1885-1893. 2016.
[55] Chang Liu, Bo Li, Yevgeniy Vorobeychik, and Alina Oprea. "Robust linear regression against training data poisoning." In *Proceedings of the 10th ACM Workshop on Artificial Intelligence and Security*, pp. 91-102. ACM, 2017.
[56] Bo Li, Yevgeniy Vorobeychik, Muqun Li, and Bradley Malin. "Scalable iterative classification for sanitizing large-scale datasets." *IEEE Transactions on knowledge and data engineering.* 29, no. 3 (2016): 698-711.
[57] Roy Dong, Walid Krichene, Alexandre M. Bayen, and S. Shankar Sastry. "Differential privacy of populations in routing games." In *2015 54th IEEE Conference on Decision and Control (CDC)*, pp. 2798-2803. IEEE, 2015.

[58] Weiyi Xia, Murat Kantarcioglu, Zhiyu Wan, Raymond Heatherly, Yevgeniy Vorobeychik, and Bradley Malin. "Process-driven data privacy." In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pp. 1021-1030. ACM, 2015.

[59] Aron Laszka, Yevgeniy Vorobeychik, and Xenofon Koutsoukos. "Optimal personalized filtering against spear-phishing attacks." In *Twenty-Ninth AAAI conference on Artificial Intelligence*. 2015.

[60] Liyiming Ke, Bo Li, and Yevgeniy Vorobeychik. "Behavioral experiments in email filter evasion." In *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.

[61] Aron Laszka, Jian Lou, and Yevgeniy Vorobeychik. "Multi-defender strategic filtering against spear-phishing attacks." In *Thirtieth AAAI Conference on Artificial Intelligence*. 2016.

[62] Himanshu Neema, Bradley Potteiger, Xenofon Koutsoukos, Gabor Karsai, Peter Volgyesi, and Janos Sztipanovits. Integrated Simulation Testbed for Security and Resilience of CPS", The *13 ACM/SIGAPP Symposium on Applied Computing (SAC 2018)*. Pau, France, April 9-13, 2018.

[63] Bradley Potteiger, William Emfinger, Himanshu Neema, Xenofon Koutsoukos, Chee Yee Tang and Keith Stouffer. "Evaluating the effects of cyber-attacks on cyber physical systems using a hardware-in-the-loop simulation testbed", *2017 Resilience Week (RWS)*, pp. 177-183, Wilmington, DE, September 18-22, 2017.

[64] Peter Horvath, Mark Yampolskiy, and Xenofon Koutsoukos. "Efficient evaluation of wireless real-time control networks." *Sensors* 15, no. 2 (2015): 4134-4153.

[65] W. Emfinger and G. Karsai, "Modeling Network Medium Access Protocols for Network Quality of Service Analysis," *2015 IEEE 18th International Symposium on Real-Time Distributed Computing*, Auckland, 2015, pp. 292-295.

[66] Goncalo Martins, Arul Moondra, Abhishek Dubey, Anirban Bhattacharjee, and Xenofon Koutsoukos. "Computation and Communication Evaluation of an Authentication Mechanism for Time-Triggered Networked Control Systems", *Sensors, Special Issue on Real-Time and Cyber-Physical Systems*, 16(8), 1166, 2016.

[67] Bradley Potteiger, Zhenkai Zhang, and Xenofon Koutsoukos. “Integrated Instruction Set Randomization and Control Reconfiguration for Securing Cyber-Physical Systems”. *Symposium and Bootcamp on the Science of Security (HotSoS 2018)*, Raleigh, NC, April 10 -11, 2018.

[68] Bradley Potteiger, Zhenkai Zhang, and Xenofon Koutsoukos. “Integrated Moving Target Defense and Control Reconfiguration for Securing Cyber-Physical Systems”, *Microprocessors and Microsystems, Special Issue on Cyber-Physical Systems: Design and Applications.* Under review.

[69] Andrew M. Smith, Jackson R. Mayo, Vivian Kammler, Robert C. Armstrong, and Yevgeniy Vorobeychik. "Using computational game theory to guide verification and security in hardware designs." In *2017 IEEE International Symposium on Hardware Oriented Security and Trust (HOST)*, pp. 110-115. IEEE, 2017.

LIST OF SYMBOLS, ABBREVIATIONS, AND ACRONYMS

AAAD	Application-Aware Anomaly Detection
ADS	anomaly detection systems
ARC-P	adversarial resilient consensus protocol
ASR	address space randomization
AV	anti-virus
C3E	computational cybersecurity in compromised environments
CM	configuration manager
COAs	courses-of-actions
CP-ADL	cyber-physical attack description language
C2WT	command and control wind tunnel
CPS	cyber-physical systems
CPS-VO	cyber-physical systems virtual organization
CVSS	common vulnerability scoring system
DDOS	distributed denial-of-service
DBT	dynamic binary translator
DER	distributed energy resources
DLMS	diffusion least-mean squares
DSML	domain-specific modeling language
FMDP	factored Markov decision process
GME	generic modeling environment
HLA	high-level architecture
HMAC	hash-based message authentication
IDS	intrusion detection systems
II	industrial internet
IoT	internet of things
ISR	instruction set randomization
LTI	linear, time-invariant systems
MIC	model-integrated computing
MDPs	Markov decision processes
ML	machine learning
MSD	mean-square-deviation
MSC	minimum set cover
MTC	minimum test cover
MTD	moving target defense
PDL	procedure derivation logic
PTAS	polynomial-time approximation scheme
RFA	remote function actuator
RFID	radio-frequency identification
ROP	return oriented programming
SALT	security-oriented active learning testbed
SGP	security game on a plane
SoS-VO	science of security virtual organization
SPEs	security games with protection externalities

STRIDE	spoofing, tampering, repudiation, information disclosure, denial of service, elevation of privilege)
SURE	secure and resilient cyber-physical systems
TCU	telematics control unit
TDMA	time-division multiple access
TT	time-triggered
UAV	unmanned aerial vehicles
WSN	wireless sensor networks