Fall 2021

Robustness against attacks and uncertainties in smart cyber-physical systems

Prithwiraj Roy

ROBUSTNESS AGAINST ATTACKS AND UNCERTAINTIES IN SMART CYBER-PHYSICAL SYSTEMS

by

PRITHWIRAJ ROY

A DISSERTATION

Presented to the Graduate Faculty of the

MISSOURI UNIVERSITY OF SCIENCE AND TECHNOLOGY

In Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

in

COMPUTER SCIENCE

2021

Approved by:

Sajal K. Das, Advisor
Shameek Bhattacharjee
Venkata Sriram Siddhardh Nadendla
Tony T. Luo
Nan Cen

ABSTRACT

Cyber-Physical Systems (CPS) are sensing, processing, and communicating platforms, embedded with physical devices that provide real-time monitoring and control. Security challenges in CPS necessitate solutions that are robust against attacks and uncertainties and provide a seamless operation, especially when used in real-time applications to monitor and secure critical infrastructures. CPS mainly consists of a physical component for sensing or monitoring and a cyber component for processing and communicating. The quality of interactions between physical and cyber systems has direct impacts on the system's performance and reliability.

CPS plays a major role in smart services and applications within a smart living environment, such as smart cities, smart energy management systems, traffic control, critical infrastructure protection, and many defense-related systems. Such smart CPS are integrating sensing, communication, computation, and control aim to achieve stability, high performance, robustness, and efficiency. This thesis concentrates on three aspects of CPS robustness and security. First, we investigate the security of smart grid systems against adversarial attacks and how reliable automation of smart grids depends on decisions based on situational awareness extracted via real-time system monitoring. Second, we take a crowdsourcing vehicular network environment to identify potential security concerns, specifically that impacts the quantification of aggregate truthfulness of events (quality of information or QoI). Finally, we generalize the CPS system to create a large-scale network to investigate how information propagates in the CPS network and possible ways to control and mitigate the information spread. In all these cases, we identify the attack strategies that can be employed by an intelligent adversarial entity to disrupt the operation of the application and how different measures can be developed to make the system more robust and resilient against attacks and uncertainties.

ACKNOWLEDGMENTS

I express my deepest gratitude to my advisor, Dr. Sajal K. Das, and co-advisor, Dr. Shameek Bhattacharjee for their advice, support, and guidance throughout my doctoral studies at Missouri University of Science and Technology. I am thankful to my parents, Ajit Roy and Rita Roy as they supported me throughout this journey. I must also mention my colleague, collaborator, and friend Venkata Praveen Kumar Madhavarapu who helped me immensely.

I am very grateful to Dr. Venkata Sriram Siddhardh Nadendla for his continuous encouragement and support. I would also like to thank Dr. Tony Luo and Dr. Nan Cen for being my Ph.D. committee members and giving their insights that helped me significantly improve the contents and presentation of my dissertation.

I am really grateful to my fellow CReWMaN member Satyaki Roy for his support and guidance. I would also like to thank Dr. Federico Coro, Dr. Ashish Gupta, and Arindam Khanda for their valuable feedbacks.

Finally, I would like to thank my dear friend Aniruddha Maiti for encouraging me to pursue doctoral studies. I should also mention my close friends Sourav, Arghya, Atanu, Pujanjan, Suman, Partha for their continuous support and encouragement whenever I reached out to them.

I would like to acknowledge the National Science Foundation for supporting this work with NSF grants CNS-1545050, CNS-1818942, and SaTC-2030624.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Figure Page

LIST OF TABLES

1. INTRODUCTION

A smart city is a model framework, predominantly composed of Information and Communication Technologies (ICT) to develop, deploy, and promote sustainable development practices to address growing urbanization challenges. To address this, smart cities are using ICT to develop energy-efficient applications coupled with automated decision making, to support various public services in urban spaces. The core part of a ICT framework is essentially an intelligent network of connected objects and machines that transmit data using wired or wireless technology and the cloud. Citizens engage with smart city ecosystems in various ways using smartphones, wearable devices, and connected cars and homes. Pairing devices and data with a city's physical infrastructure and services can reduce costs and improve sustainability and robustness of the system. Communities can improve energy distribution, decrease traffic congestion, and even improve living quality with help from interconnected networks.

Cyber-physical systems (CPS) comprise the backbone of critical infrastructures of a smart city such as power grids, transportation systems, medical devices, etc. Such CPS systems, generally termed as smart CPS is considered as the next generation of systems that integrate sensing, communication, computation, and control in order to achieve stability, high performance, robustness, and efficiency as it relates to physical systems. This coupling between the cyber and physical layer is manifested from the nano-world to large-scale wide-area systems of systems [1].

Although cyber-physical systems are being widely incorporated into various critical infrastructures, however, given the lack of countermeasures, security breaches could have catastrophic consequences. For example, in an electric grid network, if communication channels between the end devices (smart meters) to the central Energy Management System (EMS) are compromised, then the whole power grid may become unstable, possibly creating a large-scale cascaded blackout. Similar to the way the internet has transformed how humans interact and communicate with one another, CPS will transform how humans access and control the physical world around them.

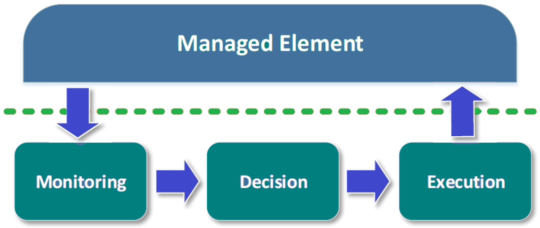


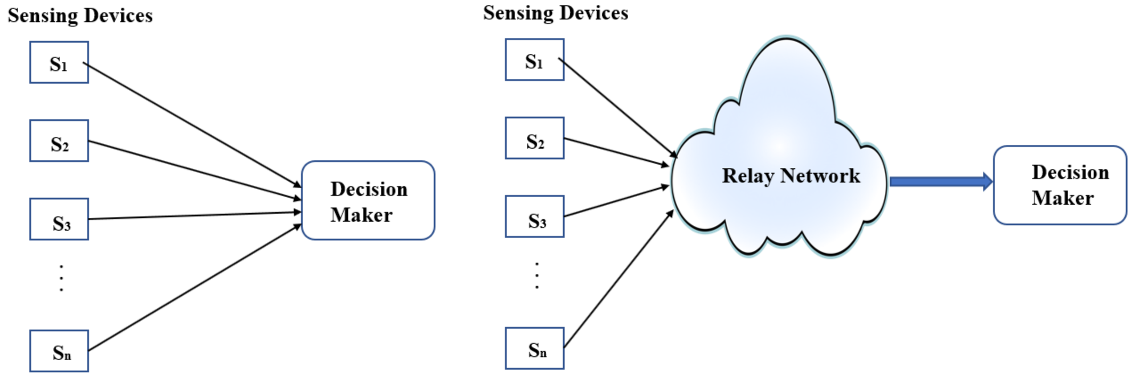
Figure 1.1. The Monitor-Control-Actuation Loop of a CPS.

Figure 1.1 [2] illustrates the monitoring and control loop of an typical CPS system. The entire system can be categorized into three phases: monitoring, decision making, and execution. In addition to security issues, CPS privacy is another serious concern. Cyber-physical systems are often distributed across wide geographic locations and they collect large volumes of different types of data for analysis and decision making. Based on the analysis and decision control actions are executed. Breaches in the data accumulation process could potentially lead to wide-scale data theft/leakage, much of which is private or sensitive information. An intelligent adversary can target to attack different stages of the system's operation, including data collection, data transmission, data operation, and data storage.

1.1. DATA COLLECTION BY SENSING DEVICES

In a typical CPS system, information is collected by sensors and then transmitted via a wired or wireless communication channel to the central server where the data is processed and some decision is taken based on the collected data. Smart applications require to sense the physical environment in real-time and process the generated data to perform decision making. In a smart CPS system, the sensing devices are not necessarily always physical sensors, in recent times, mobile crowdsensing (MCS) [3] has emerged as an enabler of cost-effective sensing infrastructure. MCS allows users possessing smart devices like smartphones, tablets, smart wearables, etc. to collect rich sensory data about the surrounding environment. Such sensing can either be opportunistic which indicates sensing through smartphone sensors or participatory sensing where sensory data is collected by humans [4]. The transmission between the sensing devices and the server can happen in two ways as shown in Figure 1.2.

1.1.1. Single Hop Transmission. In these types of systems the sensing devices, humans send the data directly to the decision-maker in a single hop. This transmission can happen either by wired or wireless technologies. In an unreliable environment that assumes the presence of adversarial attacks and uncertainty, the data captured by the sensors can not always be trusted. The attacks on the sensing devices can be carried out by a malicious adversary or even the device itself can malfunction and sends incorrect information. In this type of single hop transmission channel, it is critical to identify the presence of such compromised/faulty devices to prevent the decision maker from taking bad decisions.



(a) Single-Hop Transmission

(b) Multi-Hop Transmission

Figure 1.2. Single-Hop and Multi-Hop Transmission.

1.1.2. Multi-Hop Transmission. In these types of systems the sensing devices or the humans (*participatory sensing*) send the data via a relay network to the decision maker. For example in a collaborative data collection framework such as bioMCS [5], the decision maker can be located far away from the physical sensing devices and the devices communicate with each other to relay the collected information to the decision maker in multiple hops. In this type of transmission settings where the collected data reaches the decision maker indirectly using multiple hops, the security challenges become more varied compared to single hop transmission settings. Along with the physical attack or faulty devices, the relay network itself can introduce security issues.

1.1.3. CPS Security Objectives.

- Confidentiality: Ensures the prevention of disclosure of information to an unauthorized entity. For example, healthcare CPS requires personal patient information to be transmitted confidentially to a specific destination. In case of any breach in the network, that needs to be identified quickly and preventive measures to be incorporated.
- Integrity: States that the data or resource cannot be modified without proper authorization. When data integrity is compromised the receiver may receive falsified information and can believe it to be true. Any presence of data modification by an adversary needs to be detected in real-time and counter security measures to be deployed.
- Availability: States that the data or services must be available when it is needed to ensure an seamless and uninterrupted service.

1.1.4. CPS Security Practices. Any CPS system such as a smart grid network, crowdsourcing system, or complex social network is prone to adversarial threats. With the improvements of the defense mechanisms, the adversarial strategies are also evolving and more sophisticated methods are being incorporated by the adversarial entities. It is almost impossible to make any system to be 100% resilient to adversarial threats. There have been a plethora of works that discuss and contribute to the security and robustness of smart CPS systems. We define robustness as the ability of a network to carry out information flow under the presence of attack and uncertainties in the communication channel. The design of defense mechanisms is distributed into three broad categories.

- Detection: The first step of the secure and resilient defense mechanism is the detection of attack and anomalies in the system. Early detection of such attacks/anomalies is very important as it can prevent physical damage and financial loss to the system.

- Mitigation: As mentioned earlier the attacker and the attack strategies are continuously evolving, thus designing a full-proof defense mechanism is almost impossible. So there is a need to analyze the impact of an attack and mitigation strategy. Moreover, in some CPS systems such as crowdsourcing, it is certainly very difficult to control user participation. Maintaining a trust score for every individual device is both time and resource-consuming. Even in the presence of such a scoring mechanism, it is very easy for the attacker to change its identity and create a new account. Thus in such cases, the presence of attackers or malicious users can be assumed, and then a new robust and resilient defense mechanism needs to be developed.

- Control: The third and last step of the CPS security practices is the controlling of the spread of the attack. As the CPS system is deeply interconnected, the adversary with the access of a certain vulnerable point might be able to influence other segments of the network creating large-scale damage. So in that case, if an attack, presence of malicious information is detected, restricting the said attack is very necessary so that a large portion of the system can be disconnected from the infected segment and thus can be protected.

1.2. SMART GRID ARCHITECTURE

The electrical power delivery system has often been considered the greatest and most complex network ever built. It consists of wires, cables, towers, transformers, monitoring devices, and circuit breakers — all connected together in some fashion, as shown in Figure 1.3.

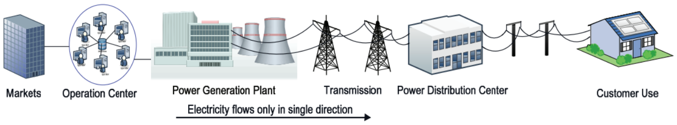


Figure 1.3. Electric Power System.

1.2.1. Supervisory Control and Data Acquisition (SCADA). Historically, the electric power grid operators and planners had limited information for the system status and behavior of the grid. The only available information was measurements from decentralized SCADA (supervisory control and data acquisition) systems, mostly recorded at several-second intervals, and they did not include the physical state variables of the a.c network like the complex voltages at every node, and time-shifted voltage information. Thus the primary focus of the system was designed for the most extreme conditions, specifically, peak loads and faults – and then try to ensure that the grid operated within that expected range. Despite the good design, operation, and maintenance efforts, over 90% of customers' electric outages occur due to problems on the distribution system rather than from transmission or generation level problems [6]. Moreover, with the growth of distributed energy resources (example: rooftop photo-voltaic cells), two-way electricity flows and new customer devices such as electric vehicles necessitate better situational awareness and insight into distribution system conditions and performance to make the grid more robust, more efficient, more distributed, reconfigurable, more interactive, with faster protection and control.

To meet these requirements, smart grid integrates modern advanced sensor technology, measurement technology, communication technology, information technology, computing technology, and control technology into it, where information and electricity flow bi-directionally [7] and the smart grid can: (1) Enable active participation by customers; (2) Accommodate all generation and storage options; (3) Enable new products, services, and markets; (4) Optimize asset utilization and operate efficiently; (5) Anticipate and respond to system disturbances; (6) Operate resiliently against attacks and natural disasters [8]. Its conceptual model is shown in Figure 1.4.

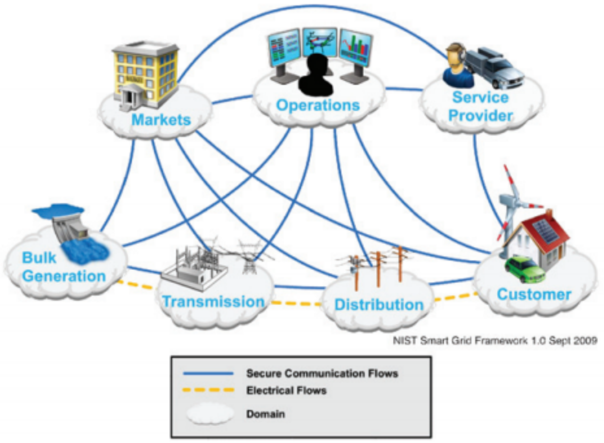


Figure 1.4. Smart Grid Conceptual Model by NIST.

1.2.2. Phasor Measurement Unit (PMU). Phasor Measurement Unit (PMU) is one of the most important grid-monitoring devices which provide dynamic visibility of the state of the power grid. PMUs capture time-synchronized real-time measurements of the magnitude and phase angle of bus voltage and line currents, the frequency, and the rate of change of frequency at the substations where they are installed. The data are sampled at rates of 20, 30, 50, 60, or 120 measurements per second and time-stamped accurately based on a common time source of the Global Positioning System (GPS). At the PMU, the synchrophasor data is assimilated in the form of the data packet and immediately sent to the Phasor Data Concentrator (PDC). The PDC receives data from all its respective PMUs and based on the associated time-tags sorts and merges the PMU data into single datasets and forwards it to the super PDC or control center or the EMS as shown in Figure 1.5.

This PMU generated data, also known as synchrophasor data are used along with the SCADA information for various power grid applications like state estimation, optimal power flow, real-time congestion control, etc. This data provides real-time close monitoring of the grid network that can assist grid operators to identify the exact type, time, and

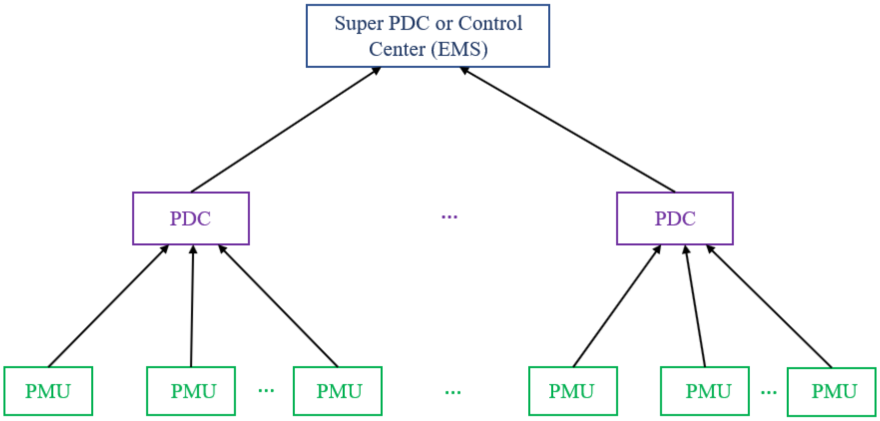


Figure 1.5. Synchrophasor System Architechture.

location of a fault or disturbance. With μPMUs installed at multiple locations throughout the electrical distribution network (e.g. the substation, end of the feeder, and any key distributed generation facilities), the PMU monitoring device is capable of supporting the analysis and operation of single or multiple feeders originating from the same substation, or even contribute to the wide-area state estimation of transmission and distribution-level grid network. The precision of the PMU measurements, time synchronization, and the ability to cross-reference locations can bring more insight to more distributed generation and storage on customer premises, more customer-initiated demand response, electric vehicles, and other changing customer load characteristics. The deployment locations of the PMUs are very critical for monitoring the grid network, a conceptual PMU network is shown in Figure 1.6 [9]. This PMU data is extremely critical since the control center bases their control actuation decisions either directly on these measurements or the output of various applications using these measurements.

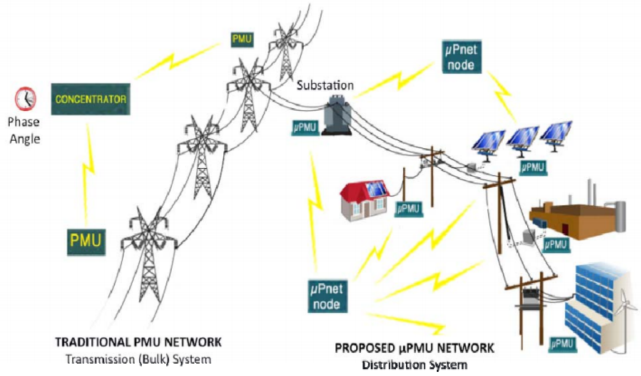


Figure 1.6. PMU Deployment Concept in an Electric Network.

1.3. CROWDSENSING SYSTEM

Today's smartphones are ubiquitous devices equipped with a plethora of embedded multimodal sensors, integrating wireless communication technologies such as 4G/WiFi, and also possess complex processing capabilities. For example, the smartphone cameras can act as video and image sensors [10], and the microphone can be used as an acoustic sensor [11, 12, 13]. Apart from that, the embedded global positioning system (GPS) receiver in the smartphones can gather accurate location information and the gyroscopes, accelerometers can extract contextual information about individual users [14, 15]. Moreover, additional sensing devices such as temperature, air quality, and humidity sensors can be connected to the smart devices via bluetooth or wired connection. These technological features, combined with the advanced sensing capability of humans, have garnered a significant amount of research from both industry and academia and generated myriad applications based on the emerging mobile crowdsensing paradigm [16].

Crowdsensing, sometimes referred to as mobile crowdsensing, is a technique where a large group of individuals having mobile devices capable of sensing and computing (such as smartphones, tablet computers, wearables) collectively share data and extract information to measure, map, analyze, estimate or infer (predict) any processes of common interest [wikipedia]. Mobile crowdsensing (MCS) has emerged as a novel paradigm for large-scale data collection and collective knowledge formation. MCS enables users equipped with energy-constrained smart devices to participate in sensing and reporting assigned tasks.

A typical MCS paradigm involves an MCS server that accumulates voluntary contributions that are supplied by either autonomous sensing agents (IoT devices) or by humans via an App. This new trend leverages the proliferation of modern sensing-capable devices along with the smartphones in order to offer a better understanding and analysis of people's activities and surroundings. MCS users are expected to voluntarily contribute sensed data to a central server via a communication channel. The benefit of using an MCS paradigm is that precise and fine-grained information collection is possible without a dedicated infrastructure to achieve seamless communication as well as efficient resource and energy management. However, this data is often tagged with spatiotemporal information which, if misused, could potentially reveal sensitive user-specific information such as their whereabouts and their health condition. Furthermore, as there is no incentive for sharing the data, there is no control over who can contribute. Most Mobile Crowdsensing applications deploy rating feedback mechanisms to help quantify the aggregate truthfulness of events (quality of information (QoI)) to improve decision-making accuracy. In case of adversarial presence in the crowdsensing environment, it is of utmost importance to devise a robust technique to validate the authenticity of the contribution made to the MCS server.

1.3.1. Quality of Information (QoI) in MCS.

Most Mobile Crowdsensing applications deploy rating feedback mechanisms to quantify the aggregate truthfulness of events (quality of information (QoI)). Based on the QoI score and the specific requirement of the application the authenticity of the event is measured. Most of the time the feedback is accumulated from voluntary contributions that are supplied by either autonomous sensing agents (IoT devices) or by humans via an App. Typical QoI scoring methods in MCS include 3 major phases: (1) accumulation of ratings/feedbacks/labels on the published event, which 'serves as a body of evidence'; (2) quantification of an event's truthfulness via a

QoI scoring model that uses the received feedbacks. (3) classification decision on whether the event is truthful or not by comparing the QoI score with a hard or soft threshold. QoI score is directly related to the rating population size, inherent error probabilities of the rating providers, and most importantly in the presence of an adversary, the fraction of the population controlled by the adversary.

1.3.2. Information Spread in Complex MCS Network. In recent times, MCS has emerged as a paradigm of cost-effective sensing infrastructure. For the participatory sensing (sensing by humans) applications, the human users are actively engaged in sensing tasks and the data from multiple users are aggregated by the MCS platforms to build a body of knowledge to support decision making. However, as the participation of MCS users is voluntary and mostly without incentives, often users are reluctant to contribute due to the incurred cost in terms of data subscription plan (if cellular connectivity is used for data transfer) and/or energy spent from device batteries. In recent times, A few energy-efficient data transfer mechanisms are proposed to keep MCS-based data acquisition sustainable for smart city applications. In [5], a centralized, energy-efficient and robust data collection framework, called *bioMCS* is proposed based on the topological properties of a biological network called *transcriptional regulatory network (TRN)*. bioMCS uses collaborative sensing among users in close proximity by leveraging energy-efficient device-to-device communication. However, this collaborative sensing is heavily dependent on the functioning of the cluster head.

Although this method is shown to be an energy-efficient and robust framework that leverages collaborative sensing to achieve high data delivery, It does not discuss how the data/information is actually spread throughout the network. Information spread control in complex networks has gained attention in recent years in a broad range of applications, from large scale IoT networks, smart grid networks, healthcare and medicine (e.g., drug design [17] or curbing epidemics [18]) to social networks [19] (e.g. moderating fake information that can lead to polarization). Selective removal of certain subgraphs called *motifs* based on the spread function value is one of the most powerful approaches to curb the overall influence spread in any complex network. The presence of motifs has been reported in many applications including large scale IoT networks, social networks, biological networks, ecological networks, and communication networks [20, 21]. The high degree of evolutionary conservation of motifs suggests that they play a key role in information dissemination, making their detection and analysis the first step towards investigating their contribution to spreading dynamics within complex networks. However, the design of an effective control mechanism to curb influence spread poses important challenges: (i) high computational complexity, especially in large complex networks, (ii) ethical issues, such as harmful bias and lack of accountability in social networks, and (iii) long-term control impact, e.g., potential reduction in network operations due to the presence of malicious information in the network which can only be observed in hindsight.

1.4. SUMMARY OF CONTRIBUTION

As discussed in the Introduction, the defense mechanism of any CPS system depends on the type of transmission (single hop or multi-hop) used in the system to transfer the data to the decision maker. Each of these brings unique security challenges.

Moreover, the defense mechanisms can be broadly categorized into three classes such as *Detection, Mitigation,* and *control.* In this thesis we have particularly put our focus on these three aspects and proposed defense mechanisms for each of these. Now all these aspects are time dependant but the detection of attack/anomaly is the most time critical task. So for that reason, we have used a smart grid application to construct our defense mechanism. Secondly, for the mitigation step, we have assumed that the CPS system is prone to attack and considered the presence of malicious users. To implement that we have considered a crowdsourcing application where devices are users contribute voluntarily.

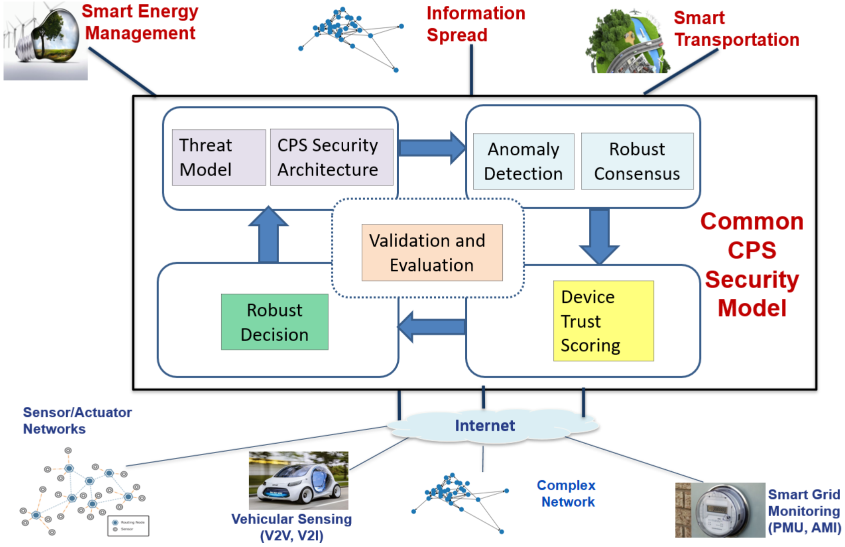


Figure 1.7. Common CPS Security Model.

Finally, in the control phase, we envisioned the CPS system as a large complex network and investigated the possible ways to control the spread of an attack. Figure 1.7 depicts all three aspects of the CPS security model.

1.4.1. PMU Security. We first discuss multiple attack strategies for data falsification attacks in distribution layer PMUs. Then, we propose a process variable selection that reduces the dimensionality of the dataset to design a light weight anomaly detection model. We use the ratio of harmonic means to arithmetic means of the active power derived from the synchrophasor data sent from PMUs as a data-driven ‘invariant’ for anomaly detection. Specifically, we find the appropriate spatial and temporal considerations of the PMU network, such that an ‘invariant’ is highly stable under no attacks but shows unique changes under various kinds of data falsification attacks. Then, we propose a two-tier threshold based detection criterion involving stateless and stateful residuals of the anomaly detection metric, that better improves the false alarm versus detection sensitivity trade-off. The two-tier detector uses the sum of long term residuals from the median absolute deviation of the ratio based metric observed over the training phase. We further propose the use of Cauchy/Lagrangian loss function to fine tune the detection threshold. Finally, we validate our work by using real two PMU datasets.

The main benefits of our approach are to provide a practical framework for compromised PMU identification that (i) real time, light weight, semi- supervised, (ii) enables quick identification, and (iii) simultaneously works for a variety of data falsification attack types.

1.4.2. Crowdsensing Security. Here we establish the use of a moving target subsampling technique as a method to ensure active resilience against bad mouthing attacks when rating sample sizes are smaller, honest rating labelers have errors in their judgment and uncertainty, and adversaries have knowledge of QoI scoring models. Specifically, we first establish some conditions under which linear QoI models (e.g., Josang, Beta Trust Models) and nonlinear models (QnQ) fail. Then, we describe the sub-population sampling method under various adversarial scenarios, and available information present to the MCS provider. We show that when consensus is lacking and the attack scale is unknown, a sub-sampled strategy for quantifying QoI is a better approach that improves the probability of evasion of bad mouthing attacks, and still infer an event as truthful, regardless of whether the adversary controls the majority or minority of the rating feedback population. We then provide a game theoretic formulation where we analyze strategic behaviors of MCS and adversaries by taking economics of security attack and defense into consideration to show improved resilience to bad mouthing attacks and boost in QoI scores, compared to traditional methods.

1.4.3. Information Security. Influence spread control in complex networks has some specific challenges: (i) high computational complexity, especially in large complex networks, (ii) ethical issues, such as harmful bias and lack of accountability in social networks, and (iii) long-term control impact, e.g., potential reduction in network operations due to the presence of malicious information in the network which can only be observed in hindsight.

Here, we first focus on the first challenge, namely, designing an efficient information spread control scheme to curb the adversarial influence on the network. Selective removal of certain subgraphs called *motifs* based on the spread function value is one of the most powerful approaches to curb the overall influence spread in any complex network. In particular, we propose a novel scoring mechanism that leverages the presence of *motifs* (or specific subgraphs) that are recurrent patterns occurring in complex networks in higher numbers than in randomized networks [22]. We first prove that any general spread function preserves both monotonicity and submodularity properties even under motif removal operations. Next, we propose a scoring mechanism as a novel spread function that quantifies the relative importance of a given motif within the overall influence spread dynamics on the complex network. We design a novel algorithm that eliminates motifs with high spread scores to curb influence spread. We evaluate the performance of our proposed spread control algorithm using simulation experiments in the context of 3-node motifs called feed forward loops (FFLs) in both real and synthetic network topologies. We demonstrate that high-scoring motifs intercept a high number of short paths from the pre-assigned source and sinks, because of which their elimination results in a significant effect on curbing the influence spread. Furthermore, we empirically evaluate the run-time and cost versus performance trade-off of the proposed algorithm.

2. LITERATURE REVIEW

2.1. ATTACK DETECTION IN DISTRIBUTION LEVEL PMUS

Real-time use of PMUs is a growing field of research, however, the practical implementation of PMU technology has been scattered. The proposed approach in this paper is to develop a statistical based semi-supervised learning framework for modeling the multivariate stream of raw PMU data that captures frequency, voltage, current measurements, and phase angles that exhibit nontrivial dependencies in real-time.

Most of the prior research on power grid data analytics (including SCADA and PMU measurements) is mainly focused on either placement of PMU in strategic locations or detecting events on the grid transmission level.

In [23, 24], the authors designed satistical testing with sliding windows that detect anomalies in a single variable. [23] propose to use the compensation theorem in circuit theory to generate an equivalent circuit to identify a specific event. [24] propose an approach for clustering sets of events to reveal unique features that distinguish different events from one another. However, both the designs are built on a single data stream, thus, can result in computational overhead. Similarly, the authors of [25] used a rule-based mechanism to detect perturbations in each data streams independently by identifying optimal placements of PMUs. In [26], the authors focus on detection and classification of smart meter anomalies using an unsupervised machine learning techniques using neural network based models. However, given the latency critical and high sampling nature of PMU, it is difficult to extend the proposed method in real-time.

In [27], a mechanism based on continuous monitoring of phase-wise equivalent transmission line impedance was proposed, for detecting data falsification on the voltage data from transmission system PMUs. However, they require two PMUs deployed at both ends of the transmission line and one of them needs to be honest. More importantly, we found that the PMU data streams at the transmission level were inherently stable making anomaly detection a less challenging problem. In [28] a Support Vector Machine (SVM) was used for detection, against a mirroring spoof attack strategy on the voltage data at distribution level PMUs. However, only falsification of voltage stream was considered which is relatively stable and makes anomaly detection less challenging.

The [29] proposed a decision tree-based anomaly detection scheme to differentiate between normal tripping and malicious tripping by training on specific attack samples. However, it is not feasible to generate 100% of all the possible legitimate line tripping cases for training in [29]. In [30] a smart Time Synchronization Attack (TSA) based on GPS spoofing was shown to be equivalent to modifying the phase angle measurement from PMUs. However, they have not discussed any defense mechanism.

In[31] a density-based local outlier factor (LOF) analysis was used to detect the anomalies among the data, to describe spatio-temporal outliers among all the synchrophasor measurements from the grid. However, this method might not be able to detect attacks in real time, and in their proposed method the authors have only considered an attack on voltage magnitude.

A critical analysis of all previous works on the detection of PMU data falsification revealed that current data falsification for PMU streams was not investigated. Furthermore, we found that, unlike transmission level PMUs, the distribution level PMU's current synchrophasor data shows high dynamic variations in benign conditions, making anomaly detection challenging. Finally, all previous defenses are stream specific in the sense that they only work for either voltage or phase falsification. Since each PMU contains 4 streams and has 3 phases, a stream specific defense will require 12 different defense models that need complex cross-coordination.

2.2. ATTACK RESILIENCE IN MCS

Typical trustworthy MCS includes 5 major phases: (1) accumulation of ratings/ feedbacks/ labels on the published event, that ‘serves as a body of evidence’; (2) quantification of the event’s truthfulness score via a Quality of Information (QoI) model that uses the body of evidence (i.e., rating feedbacks). (3) classification decision on whether the event is truthful or not by comparing the QoI score with a hard or soft classification threshold. (4) a user reputation model that rewards the reporters who were involved in the reporting concluded truthful events while penalizing those reporting events concluded as not truthful. (5) Such rewarding are aggregated over time across the whole network, to calculate the final reporting reputation of a user, and incentives are disbursed based on the reputation.

Normally, the QoI is calculated by modeling evidence obtained using (i) ground truth, (ii) similarity based outlier detection, (iii) spatio-temporal provenance, (iv) prior reputation context, and (v) the rating feedback mechanism. However, the ground truth is not always immediately available, and often acquiring the same is not guaranteed or feasible. Additionally, obtaining ground truth often requires deployment of dedicated infrastructure thus obviating the main benefit of crowdsensing.

The QoI’s is typically computed by Majority Voting [32], Josang’s Belief [33], Dempster Shafer [34], or nonlinear models (QnQ [35], Gompertz [36]. The user reputation model is computed by Dempster Shafer, Beta/Dirichlet distribution via various aggregation operations for multi-source fusion and time averaging that utilize outcomes of the QoI phase. Similarity based outlier detections such as [37], [36], [38] checks the similarity among user’s contributions in terms of event type, location, and time stamp and awards higher QoI to events having higher similarity. Spatio-temporal provenance based schemes [39] rely on the existence of prior and reliable reputation scores of all the users. Research works also investigated the design of incentive mechanisms in order to influence users behaviors so that users will produce high-quality data. Such incentive mechanisms mostly include auctions [40], [41], lotteries using the Tullock contests [42], trust and reputation systems [43]. Other such incentive mechanisms include bargaining games [44], contracts [45], or market mechanisms [46]. For example, [40] proposed Thanos that incorporates QoI into an incentive mechanism based on reverse combinatorial auctions to achieve near-optimal social welfare.

[47] proposed a cross validation (CV) approach to quantify the data quality by using a *validating crowd* to ratify the *contributing crowd* in terms of the sensor data contributed by the latter, and uses the validation result to reshape data into a more credible posterior belief of the ground truth. This approach leverages the "side information" possessed by people, which includes (diversely) people's domain knowledge, professional expertise, etc.

All QoI models are based on the assumption that the presence of a crowd automatically means the presence of a substantial amount of rating feedbacks that is enough to keep the relative proportion of compromised feedbacks to the total number of feedbacks low enough for QoI models to successfully infer the QoI that is unbiased. Nonetheless, this assumption is often not practical or useful for the following reasons. First, MCS the main adversary is a rival business, who's aim is to not let a new business competitor grow. Initially, all newly launched crowdsensing systems have a lower customer base, and hence rating feedback labelers are lower to begin with. The rival business with a practically small attack budget can poison the QoI inference and prevent good reporters from gaining reputation. Eventually, demotivated good reporters will cease to be active or exist. Consequently, such action can be accompanied by introducing rogue reporters by the rival business to introduce fake or spam reports which will get accepted easily due to less competition from good users.

Second, even when the crowdsensing systems may have a high user base, the geographical spread of this user base may not be spatio-temporally uniform. For example, downtown area has less crowd during night, parts of the city can be inherently sparsely populated than other areas. Event QoIs have a strong spatial relationship in general and this creates further danger in having negative QoI inferences. Hence, this makes the problem a very important security issue. Furthermore, most of these solutions do not mathematically incorporate the error probability of rating feedbacks from honest raters. When the error probability is combined with the presence of an attack, even for a minority malicious rating population, the QoI with existing models can produce sub-optimal results.

2.3. INFORMATION PROPAGATION IN COMPLEX NETWORKS

The Topological and functional properties of motifs are important for analyzing large-scale complex networks. While 4-node motifs were found to be responsible for nutrient metabolism and bio-synthesis in biological networks [22, 48], 3-node motifs are shown to affect information propagation by acting as filters, pulse generators, and response accelerators [49]. This motivated researchers to investigate motifs from the perspective of information processing. For example, the influence of motifs on information dissemination in neuronal networks is investigated in [50]. Motifs are also hypothesized to regulate information flow within constrained time windows [21]. Hence, complementary research efforts were motivated to decode the impact of broader organizational patterns of motifs. For example, the authors in [51] proposed an approach called *motif generalization* to group similar motifs into families; while the organizational pattern of motifs was analyzed in [52]. Finally, the notion of higher-order motifs is explored in [53], which shows motif aggregation to form *motif-of-motif* structures.

Topological robustness due to the presence of motifs is well-studied in biological networks such as Transcriptional Regulatory Network (TRN). The authors in [21] proposed a network centrality measure, called *motif-based centrality*, to quantify the importance of a node in terms of its motif participation. This metric was utilized to design efficient fault-tolerant topologies and robust routing protocols in various communication networks, such as wireless sensor networks [54, 55, 56], delay tolerant networks [57], and edge computing frameworks [58]. These networks optimize communication performance goals, such as data delivery rates, communication delay, and energy efficiency. Since motifs play a key role in the information spread, their removal due to failures or attacks makes the performance of complex networks (e.g., wireless sensor networks or smart grid networks) highly vulnerable. Thus, it is imperative to evaluate the cost and effects of failures or attacks that knock off network motifs.

3. REAL TIME STREAM MINING BASED ATTACK DETECTION IN DISTRIBUTION LEVEL PMUS FOR SMART GRIDS

CPS systems are widely used in a smart city environment for sensing the physical environment in real-time and processing the generated data to inform decision making. One of the most important such applications is the smart grid management system. Reliable automation of smart grids depends on decisions based on situational awareness extracted via real time system monitoring and accurate state estimation. The Phasor Measurement Units (PMU) at distribution and transmission layers of the smart grid provide high velocity real time information on voltage and current magnitudes and angles in a three phase electrical grid. Naturally, the authenticity of the PMU data is of utmost operational importance. Data falsification attacks on PMU data can cause the Energy Management Systems (EMS) to take wrong decisions, potentially having drastic consequences on the power grid's operation. The need for automated data falsification attack detection and isolation is key for EMS protection from PMU data falsification. Here, we propose an automated distributed stream mining approach to time series anomaly based attack detection that identifies attacks while distinguishing from legitimate changes in PMU data trends. Specifically, we provide a real time learning invariant that reduces the multi-dimensional nature of the PMU data streams for quick big data summarization using a Pythagorean means of the *active power* from a cluster of PMUs. Thereafter, we propose a methodology that learns thresholds of the invariant automatically, to prove the predictive power of distinguishing between small attacks versus legitimate changes. Extensive simulation results using real PMU data are provided to verify the accuracy of the proposed method.

3.1. BACKGROUND

Traditionally, power grid operators had limited information about dynamically varying system states in the grid. Many major faults in the grid are usually preceded by ephemeral warning signs (e.g., voltage sags) that Supervisory Control And Data Acquisition (SCADA) measurements (with data resolution of several seconds) could not capture as shown in [6]. To alleviate this problem, PMUs are deployed to capture fine grained high resolution time series data. These PMUs form the crucial endpoint device for the PMU Infrastructure, one of the key cornerstones of the modern smart grid design. Furthermore, with the increasing market penetration of Distributed Energy Resources (DERs) (e.g. solar panels), two-way electricity flows, and novel loads (such as electric vehicles), the grid requires real time grid monitoring, making the integrity of PMU data streams of strategic importance. The PMUs record time-synchronized measurements of voltage, current, phase angle and frequency (collectively known as synchrophasor data) and sends it to an aggregator called Phasor Data Concentrator (PDC) using *single hop transmission* protocol IEEE C37.118-2. The PDC, in turn, relays such data to a control center, allowing grid operators to localize and infer the type, time and location of a fault or disturbance as well as support critical control-actuation operations such as state estimation, maintain optimal power flow, based on the measured PMU data streams. The architecture of a typical PMU-PDC infrastructure is shown in Figure 3.1.

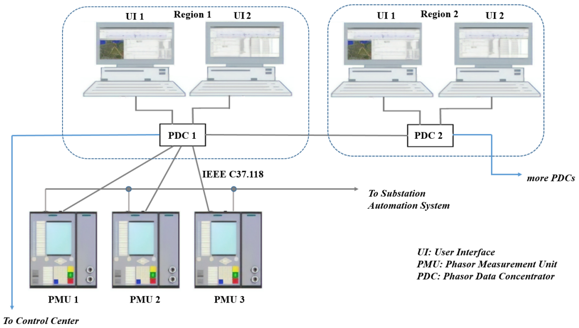


Figure 3.1. Architecture of a PMU Infrastructure.

However, in recent years, power distribution systems have faced cyber-attacks, threatening their security, reliability of operations. The report of US National Research Council highlights potential multi-state blackouts as a result of coordinated False Data Injection (FDI) attacks on power systems[59]. Such an attack on the Ukrainian power grid resulted in the loss of service for approximately 225,000 customers in three different territories which lasted for several hours [60]. Stuxnet worm has directly affected more than 100,000 industrial components [61]. However, the widely accepted IEEE C37.118-2 protocol for synchrophasor communication is highly vulnerable to cyber-attacks [62, 63]. In fact, most synchrophasor data transmission happen on non-reliable and insecure IP networks. Heavy encryption is not possible due to the latency critical nature of PMU data applications, thus increasing the chances of FDI attacks. This motivates the need for anomaly based intrusion detection in PMUs. While some existing research [27, 30] offer solutions, they have the following limitations: [27] focus on transmission layer PMUs, where data is very stable, thus making anomaly detection easy. The [28] considers the problem of only voltage data falsification, which is stable and hence easy to detect, ignoring current data falsification.

In this work, we first discuss multiple attack strategies for data falsification attacks in PMUs. Then, we propose a process variable selection that reduces the dimensionality of the anomaly detection problem. Then, we use a ratio of harmonic means to arithmetic means of the active power derived from the synchrophasor data sent from PMUs as a data-driven ‘invariant’ for anomaly detection. Specifically, we find the appropriate spatial and temporal considerations of the PMU network, such that an ‘invariant’ is highly stable under no attacks but shows unique changes under various kinds of data falsification attacks. Then, we propose a two-tier threshold based detection criterion involving stateless and stateful residuals of the anomaly detection metric, that better improve the false alarm versus detection sensitivity trade-off. The two-tier detector uses the sum of long term residuals from the median absolute deviation of the ratio based metric observed over the training phase. Finally, we validate our work by using real PMU datasets collected from Lawrence Berkley National Lab across 12 days.

The main benefits of our approach are to provide a practical framework for compromised PMU identification that (i) real time, light weight, semi- supervised, (ii) enables quick identification, and (iii) simultaneously works for a variety of data falsification attack types.

3.2. SYSTEM AND THREAT MODELS

3.2.1. PMU System Architecture. Here we first describe the PMU infrastructure network architecture. Most PMUs measure time-stamped voltages and current magnitudes and their phase angles denoted by $V_t(j)$, $I_t(j)$, $\theta_t^V(j)$, $\theta_t^I(j)$ respectively, where t is the time stamp and j is the j-th phase. These PMUs are deployed at strategic points of the transmission and distribution layers of the smart grid. Each PMU sends its data to a regional decentralized data aggregator known as PDC. The corresponding PDC in turn relays the aggregated data from multiple PMUs to a Local Controller Center (LCC). Various local controller centers communicate with each other forming a wide network for synchronizing local and global PMU data. In this work, we are specifically interested in a *decentralized anomaly detection that runs on a PDC or a LCC and facilitates early attack detection from a bunch of PMUs that are geographically proximate in terms of the PMU network.*

3.2.2. Dataset Description. We have used two different dataset for our analysis. The first dataset is collected at Lawrence Berkeley National Laboratory (LBNL) in Berkeley, CA and the second dataset is collected at Ecole Polytechnique Federale De Lausanne (EPFL) campus in Switzerland.

3.2.2.1. LBNL data. We use a dataset collected from the Power Standards Lab (PSL) at LBNL in Berkeley, CA, which developed high-precision μ-PMUs for showing how steps in our framework related to a real PMU system. The LBNL dataset contains three μ-PMUs that are deployed at multiple utility and LBNL campus locations on a 12 kV distribution grid. The μ-PMU devices are named as: Grizzly, A6, and Bank514 in the dataset. Each μ-PMU device produces 12 streams of 120 Hz high-precision values with timestamps accurate to 100 ns (the limit of GPS). The 12 streams of data include both magnitude and phase angle for both voltage and current for all three phases on a true distribution network[64].

3.2.2.2. EPFL data. This dataset is from the monitoring infrastructure of the smart-grid pilot project in the EPFL campus [65]. It consists of voltage, current, and frequency data from five PMUs deployed in a 20 kV active distribution network and the data has been recorded as 50 fps. The PMU devices are named as: PMU2, PMU3, PMU4, PMU5, PMU6. The 12 streams of data include both magnitude and phase angle for both voltage and current for all three phases on a true distribution network.

3.2.3. Threat Model. This section describes three features characterizing the threat model (e.g.,attack types, falsification margins, and falsification distributions) that can be employed by organized adversaries.

3.2.3.1. Threat model scope. PMU being a comparatively new research area, real malicious data samples from PMUs are hard to find. Therefore we generated the malicious samples by applying the three aspects of adversarial strategy over the real data. We have ensured that the falsification strategies used, do not favor or suit our proposed defense mechanism.

In simple electrical terms, the term load is equivalent to the current magnitude in each phase. Typically, in any phase, there could be two possibilities of load change. Either there could be an increase or decrease in current, both creating an imbalance in the power grid. An increase in the phase current will cause the phase voltage to drop. If the current increases too much, then the phase is shed or the load is switched to other phases. Imbalance can also occur if the current drops in any phase, making the system inefficient in terms of utilization. *This creates a motivation to falsify current measurements.*

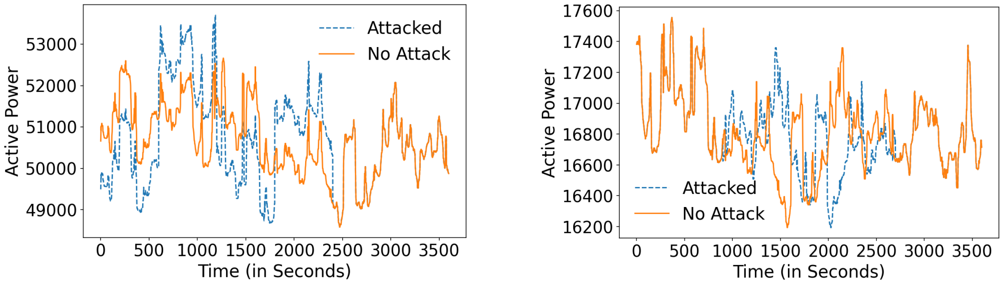
3.2.3.2. Attack types. Attacks can be categorized in different types based on how data is changed across multiple PMUs. Organized adversaries can falsify data from single or multiple compromised PMU(s) simultaneously. Based on the objective and intent of the adversary, any of the four streams (*Voltage Magnitude, Voltage Angle, Current Magnitude, Current Angle*) of each phase can be falsified.

We assume the adversary falsify the 'current magnitude'. Let $I_t^i(act)$ be the actual current magnitude of i-th PMU at time t, while I_t^i be it's reported value. Under no attacks, the actual and reported value $I_t^i = I_t^i(act)$, while under attacks the reported value I_t^i can be biased by the following ways:

- Deductive: In this case I_t^i from the i-th compromised PMU at time t is changed to $I_t^i(act) - I_{\delta_t}$, where $I_{\delta_{min}} \leq I_{\delta_t} \leq I_{\delta_{max}}$, for $I_{\delta_{min}} > 0$ is the false bias. Deductive attacks disrupt the efficiency of the grid by reducing the power utilization.

- Additive: An additive attack can be launched by a rival utility to make the control center believe in a sudden increase in load which might lead to load shedding in that particular phase. Therefore, for additive falsification, the modified attack sample is $I_t^i = I_t^i(act) + I_{\delta_t}$ from a compromised PMU.

- Alternating Attack: The adversary alternates between additive and deductive falsification for equal time duration over the time domain with the same average bias value of I_{δ_t}. In such a case, the effect of additive and deductive falsification will cancel each other's effect over a particular time period making it hard over most device specific statistical anomaly detectors to detect such attacks. Figure 3.2a demonstrates the impact on the power if a PMU is compromised with such an attack.



(a) (b)

Figure 3.2. Attack on PMU2. (a) Alternating Attack; (b) Mirroring Attack.

- Mirroring Attack: Here attacker captures I_t^i for some period and then replaces the actual current measurements with the mirror image of captured I_t^i. Using mirroring attack the adversary can disrupt the system without violating the upper and lower value of the attacked PMU active power. Figure 3.2b shows the impact of mirroring attack on the active power of EPFL dataset.

3.2.3.3. FDI margin. We consider $I_{\delta_{avg}}$ as the *average margin of false data* for each compromised PMU. The strategic value of $I_{\delta_{avg}}$ is selected by an adversary as some value that ensures some minimum damage to the system. We keep this as an uncontrolled variable to test detection sensitivity since there could be various applications of PMU data. We consider that the attack is uniformly distributed $I_{\delta_t} \in [I_{\delta_{min}}, I_{\delta_{max}}]$ that does not change the resultant shape of the load distribution drastically, making it a smarter and less obvious attack.

3.2.3.4. Attack strategies. We consider three types of attack strategies:

- Step Strategy: In this case the adversary modifies all samples to higher (additive) or lower (deductive) values by $I_{\delta_{avg}}$ in the attack period, Δ_a.

$$I_t^i = \begin{cases} I_t^i(act), & \text{if } t \notin \Delta_a \\ I_t^i(act) + I_{\delta_t}, & \text{if } t \in \Delta_a \end{cases}$$

- Ramp Strategy: A ramp attack involves gradual modification of the actual measurements by the adversary. Here adversary gradually increases the I_{δ_t} in each time slots to reach $I_{\delta_{max}}$ and then again gradually decreases I_{δ_t} [27]. Based on the adversary's intend, this attack can also be both of additive and deductive in nature.
$$I_t^i = \begin{cases} I_t^i(act), & \text{if } t \notin \Delta_a \\ I_t^i(act) + \lambda_r.t, & \text{if } t \in \Delta_a/2 \\ I_t^i(act) - \lambda_r.t, & \text{if } \Delta_a/2 \in t \in \Delta_a \end{cases}$$
- Random Strategy: This attack involves the addition/subtraction of positive values generated by a uniform random function to the actual measurements. The upper (a) and lower (b) bounds for selection are provided to the function as an input
$$I_t^i = \begin{cases} I_t^i(act), & \text{if } t \notin \Delta_a \\ I_t^i(act) \pm random(a,b), & \text{if } t \in \Delta_a \end{cases}$$

3.3. PROPOSED FRAMEWORK

The proposed framework is divided into four steps: (1) Propose a derived process variable (active power from synchrophasor measurements) that will form the basis for the anomaly detection process; (2) Design the invariant metric by optimizing spatial and temporal granularities of the process variable; (3) Design of a stateless and a stateful detection thresholds that identify the normal region of invariants under no attacks from the training set, such that false alarms are not drastically sacrificed for detection sensitivity improvement; (4) Determine the detection criteria parameters, based on learning from the training and cross validation steps, and apply it on the testing set, such that the predictive accuracy of distinguishing between legitimate changes versus malicious attacks is improved.

3.3.1. Choosing Process Variable for Anomaly Detection. Given the high velocity of the data, quick lightweight analytical tools are required for big data summarization to ensure the security and integrity of the dataset. However, due to 12 streams of data per PMU, the variety of data is extremely large. With multiple data streams per PMU, the anomaly monitoring of all these streams separately increases the computational cost and latency in anomaly detection analytics.

Hence, we propose the active power calculated from synchrophasor data streams per PMU, as the process variable over which the data driven invariant is designed. The active power ($P(j)$) per phase from PMU measurements are calculated using the following standard power equations:

$$P(j) = V(j)I(j)\cos\theta(j). \tag{3.1}$$

Here $j \in \{1, 2, 3\}$ denote the phases and $V(j), I(j), \theta(j)$ are voltage magnitude, current magnitude, and angle difference between voltage and current phases respectively, for the j-th phase. This reduces the complexity of the monitoring each stream separately unlike existing works.

Another advantage is that any deliberate falsification of the voltage or current (both in terms of magnitude and phase) will impact the active power, and hence we can potentially detect an attack on any of the data streams from PMUs. Therefore, for our anomaly detection, we propose to use the phase wise monitoring of the active power $P(j)$ as a starting point. To clean the raw dataset [64] we have also applied 95% Winsorization before proceeding with our model.

3.3.2. Achieving an Invariant for Anomaly Detection Metric. For real time anomaly detection in CPS, it has been established that a metric which is invariant under normal operating conditions (without any attack) is ideal for attack detection. However, unlike tightly controlled industrial CPS applications, the distribution level synchrophasor data is affected by randomness and renewable power outputs and consumption patterns, causing traditional statistical invariants to have high randomness. As shown in Figure 3.3(a) the arithmetic mean of the time series is not stationary. Here day 1, 2, 3, 6 are weekdays and day 4, 5 are weekends. Prior works such as [66] propose the use of derived smoothing statistics of the arithmetic mean (such as ARMA, EWMA, CUSUM control charts) for time series anomaly detection. However, Figure 3.3(a) shows that time series of PMUs active power fluctuates greatly over time windows, making it difficult to distinguish legitimate changes from a malicious one. Any moving average or smoothing technique either loses sensitivity for a small margin of attacks (since the moving average does not reflect the changes beyond already existing deviations or has large false alarms.

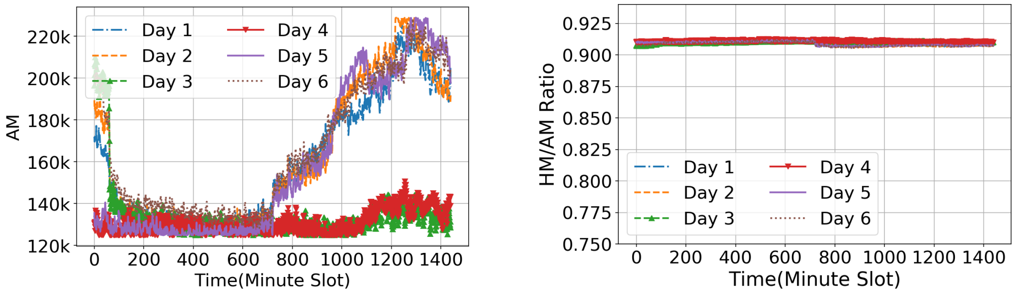
Let $\boldsymbol{P_t} = [P_t^1, ..., P_t^N]$ denote the active power from N PMUs at time slot t. Recently, in [67], we have shown that the ratio of harmonic mean and arithmetic mean of positively correlated variables exhibit invariance in their time series even when the individual means show non-stationarity. Additionally, [67] showed that the data perturbations in any variable cause the ratio to lose its invariance. However, this stability is guaranteed for appropriately correlated variables only. Hence, our primary goal is to investigate how to apply this on active power from PMUs. To this aim, we need to find the appropriate spatial and temporal granularity that maximizes the correlation between active powers on a given phase across different PMUs, which ensures invariance in the detection metric. Figure 3.3(b) depicts the stability of the invariant metric under optimal spatio-temporal optimization.

We propose to use the Harmonic to Arithmetic mean ratio as our invariant. Let the harmonic mean (HM_t) and arithmetic mean (AM_t) of P_t at time slot t be defined as:

$$HM_t = N(\sum_{i=1}^{N} P_t^i)^{-1} \quad \text{and} \quad AM_t = \frac{1}{N}\sum_{i=1}^{N} P_t^i. \tag{3.2}$$

We calculate HM_t and AM_t for slot t over a time window T of length n slots. Then we calculate the average HM_t to AM_t ratio, $Q^r(T)$, at the end of each window as follows:

$$Q^r(T) = \frac{\sum_{t=1}^{n} HM_t}{\sum_{t=1}^{n} AM_t} \tag{3.3}$$



(a) (b)

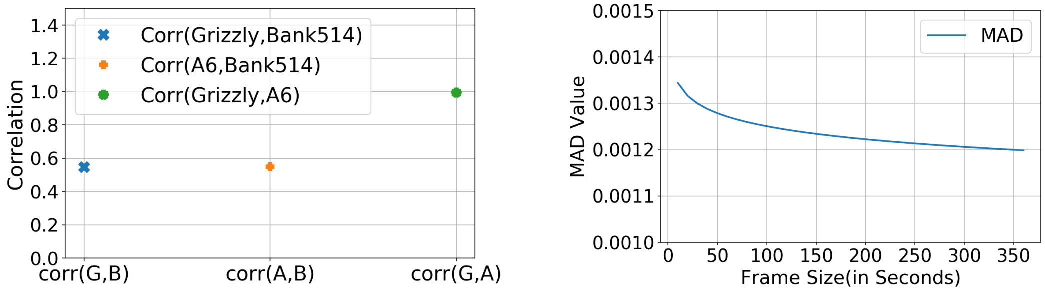
Figure 3.3. Illustrations of AM and HM/AM. (a) AM; (b) HM/AM.

where $0 \le Q^r(T) \le 1$, as $HM_t \le AM_t$.

3.3.2.1. Optimizing spatial granularity. Intuitively, a group of PMUs connected to the same feeder or serving proximate geographical areas should exhibit some interdependence in the synchrophasor data streams. We use the pairwise Pearson correlation coefficient to identify clusters that show some level of positive correlation. The higher the desired level of invariance, the higher is the required level of positive correlation. We calculate hourly Pearson's correlation among all pairs of PMUs in the training set to find groups having a maximum correlation. In the LBNL dataset, the mean of hourly correlations between Grizzly, A6 is 0.98; between Grizzly and Bank514 is 0.54; between A6 and Bank514 is 0.55 as shown in Figure 3.4(a). It is evident from the mean correlations that Grizzly and A6 are connected to the same feeder and thus can be considered in a single cluster. The average correlation identifies PMUs to be clustered under one instance of the anomaly detection technique.

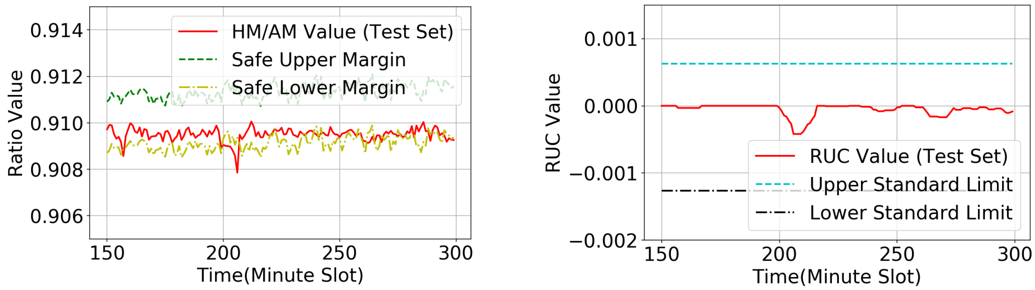
3.3.2.2. Optimizing temporal granularity. Now we focus on choosing the appropriate time granularity over which the ratio metric is calculated. The time granularity should be such that the invariance in the ratio metric is maximized (i.e., minimize the measure of dispersion in the ratio statistic). Therefore, we solve the following search problem:

$$T = \underset{T^*}{\operatorname{argmin}} MAD(Q^r(T^*)). \tag{3.4}$$



(a) (b)

Figure 3.4. PMU Clustering and MAD Over Time Window. (a) Correlation among the PMUs; (b) MAD over time windows.



(a) (b)

Figure 3.5. Stateless and Stateful Residuals for $\epsilon = 0.85$. (a) Stateless Residual; (b) Stateful Residual.

In the previous equation, $MAD(Q^r(T^*))$ is the median absolute deviation (MAD) of the resulting ratio time series with candidate time granularity $T^* \leq 360$ seconds. We choose T^* that minimizes the MAD of the ratio time series (shown in Figure 3.4(b)).

3.3.3. Stateless and Stateful Residual based Threshold Design. Intuitively, The anomaly detection needs to identify a proximate spatial region around the ratio time series that specifies the behavior of the invariant under no attacks. Usually, a threshold is calculated by tracking the difference between the actual time series value and its smoothed value over time. However, a simple threshold based approach, cannot decrease both false alarms and missed detections simultaneously [66]. Hence, we put forward a two-tier approach with stateless and stateful residuals.

3.3.3.1. Stateless residuals. The stateless residual is an instantaneous residual per time window T. Our method computes the mean μ_r and median absolute deviation, m_Q, from the probability distribution of ratio values $Q^r(T)$ for each PMU cluster (shown in Figure 3.5(a)).

Unlike our previous work [67], we propose the use of Median Absolute Deviation (MAD) as a scale parameter for designing the stateless residual rather than the standard deviation (SD), because MAD is more robust to outliers. Thus, MAD can automatically adjust the resultant safe margin under errors and outliers in the training. The MAD is robust than SD since it is based on a squared error from the mean, so a finite number of outliers can influence SD easily compared to MAD, thus reducing sensitivity to small attack strengths.

Stateless residual is parameterized as $\kappa = \epsilon m_Q$ where $\epsilon \in (0,4]$, such that $\kappa \in (0, 4m_Q]$ and m_Q is the MAD. Intuitively, larger κ values produce wider safe margins, thus reducing false alarms but increasing misdetection and vice-versa. Hence, a trade-off is necessary for selecting a threshold that will automatically generalize into lowering false alarms while not sacrificing the detection sensitivity, which is taken care of by the stateful residual as shown in Figure 3.5(b).

Our framework calculates a parameterized 'stateless residual' with two values; $\Gamma_l(T)$, and $\Gamma_h(T)$ around the observed instantaneous ratio values $Q^r(T)$, on every time window on the training dataset as shown in the Equation 3.3.

$$\begin{aligned} \Gamma_h(T) &= Q^r(T) + \epsilon m_Q; \\ \Gamma_l(T) &= Q^r(T) - \epsilon m_Q. \end{aligned} \tag{3.5}$$

$$\nabla(T) = \begin{cases} Q^r(T) - \Gamma_h(T), & \text{if } Q^r(T) > \Gamma_h(T); \\ Q^r(T) - \Gamma_l(T), & \text{if } Q^r(T) < \Gamma_l(T); \\ 0, & \text{otherwise.} \end{cases} \tag{3.6}$$

The instantaneous stateless residual $\nabla(T)$ which is the 'signed residual distance' between the observed ratio and the stateless residuals is derived using Equation 3.6.

The value of $\nabla(T)$ could be positive (or negative) depending on whether the ratio sample observed is above (or below) the upper (or lower) safe margin $\Gamma_h(T)$ (or $\Gamma_l(T)$). Thus, $\nabla(T)$ is zero when the ratio observed is within $[\Gamma_h(T)\ ,\Gamma_l(T)]$.

3.3.3.2. Stateful residuals. Our framework now maintains the sum of residuals between the ratio value and the $\Gamma_h(T)$ and $\Gamma_l(T)$ over a sliding frame of past K time windows. We denote this sum as $RUC(T)$. To calculate this metric. Now, the framework calculates $RUC(T)$ over a sliding frame of past K time windows as:

$$RUC(T) = \sum_{j=T-K}^{T} \nabla(j). \tag{3.7}$$

Algorithm 1 Calculate τ_{max}

```
input : list of $\tau$: $[\tau]$
output : $\tau_{max}$;
for $T$, $[\tau]$ do
    if $RUC(T) > 0$ then
        if $RUC(T) < \tau$ then
            $C_{max} : \frac{|\tau - RUC(T)|}{w}$
        if $RUC(T) > \tau$ then
            $P_{max} : w|RUC(T) - \tau|$
$\tau_{max} = \underset{\tau}{\text{argmin}}(|sum(C_{max}) - sum(P_{max})|)$
```

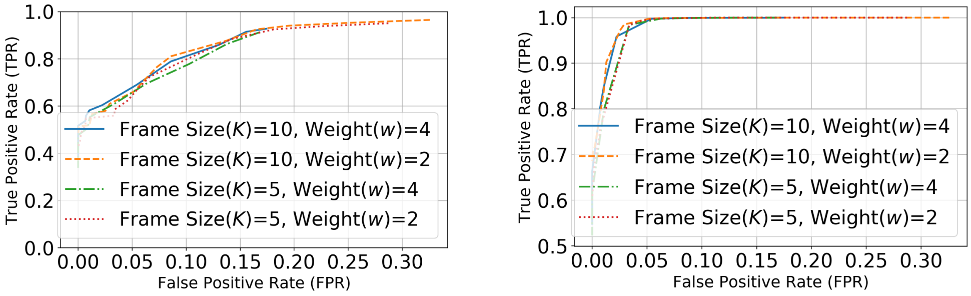
3.3.4. Optimizing Standard Limits of $RUC(T)$. We need to calculate an upper and a lower threshold from the RUC values that prevent underfitting and overfitting and improves detection performance in the test set. The procedure for calculating the upper and lower thresholds is similar. Algorithm 1 and 2 shows the method for calculating τ_{max} and τ_{min} respectively.

For this, we define a cost C, and penalty P, as the loss functions. The cost and penalty function represents the loss due to missed detection and false alarms respectively. One key consideration in time series attack detection is to minimize false alarms, since the actual probability of being under attack is much lesser. Therefore, seemingly low false alarm rates, do not necessarily indicate a good usable attack detector. Therefore, we need to give more importance to the false alarms. Hence, the loss due to false alarm (penalty P) gets more weight, compared to the loss due to missed detection (cost C) as is evident in Algorithm 1. In the end, we choose a threshold τ_{max} (and τ_{min}) which minimizes the absolute difference between total cost and penalty values for the positive RUC samples (and negative RUC samples).

Algorithm 2 Calculate τ_{min}

```
input : list of $\tau$: $[\tau]$
output : $\tau_{min}$;
for $T$, $[\tau]$ do
    if $RUC(T) < 0$ then
        if $RUC(T) > \tau$ then
            $C_{min} : \frac{|\tau - RUC(T)|}{w}$
        if $RUC(T) < \tau$ then
            $P_{min} : w|RUC(T) - \tau|$
$\tau_{min} = \underset{\tau}{\operatorname{argmin}}(|sum(C_{min}) - sum(P_{min})|)$
```

The frame size K and weight w of C and P can be determined optimally, by using a small cross validation set with a few attack samples and test what values of K and w are best. We plot the Receiver Operating Characteristic (ROC) curve for the cross validation set (See Figs. 3.6(a) and 3.6(b)) for various values of K and w, and choose that combination that gives the steepest ROC curve.



(a) (b)

Figure 3.6. Parameter Selection from Cross Validation. (a) ROC(CV) Additive Attack; (b) ROC(CV) Deductive Attack.

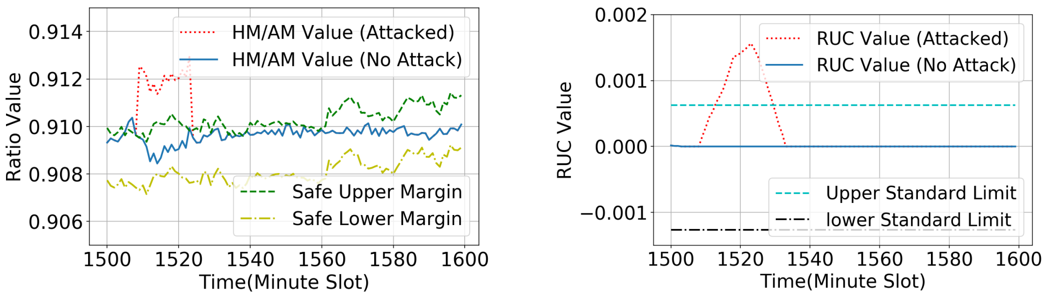
3.3.5. Detection Criterion in Test Set. The main idea behind attack detection is that RUC in the test set ($RUC(T^C)$) should not deviate from the standard limit obtained from the training set. We first calculate the stateless residuals for each time window of the testing set T^C such that $\Gamma_h(T^C) = Q^r(T^h) + \kappa_{opt}$ and $\Gamma_l(T^C) = Q^r(T^h) - \kappa_{opt}$. κ_{opt} is derived from the set of κ that produces optimal standard limit. The historical value of the ratio on that time window $Q^r(T^h)$, where T^c is the current time window and T^h is the corresponding time window in the training set, $\Gamma_{high}(T^c)$ and $\Gamma_{low}(T^c)$ are the safe margins at T^c time window of the test set.

From $\Gamma_h(T^C)$ and $\Gamma_l(T^C)$, we calculate the $RUC(T^C)$ using Equation 3.8. Then we check whether $RUC(T^C)$ violates the standard limit range identified during training set.

$$RUC(T^c) : \begin{cases} \in [\tau_{min}, \tau_{max}], \text{No Anomaly;} \\ \notin [\tau_{min}, \tau_{max}], \text{Anomaly.} \end{cases} \tag{3.8}$$

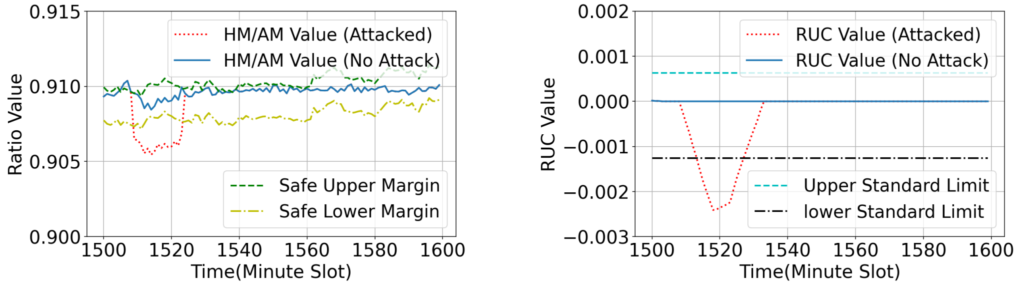
3.4. EXPERIMENTAL RESULTS

Using the LBNL PMU dataset (see Sec. 4.4), we conducted extensive experiments for different falsification margins and attack strategies. For our experimental results, the first seven days are the training set and the next two days of data is used for cross validation



(a) (b)

Figure 3.7. Anomaly Detection for Additive Step Attack. (a) Tier 1; (b) Tier 2.

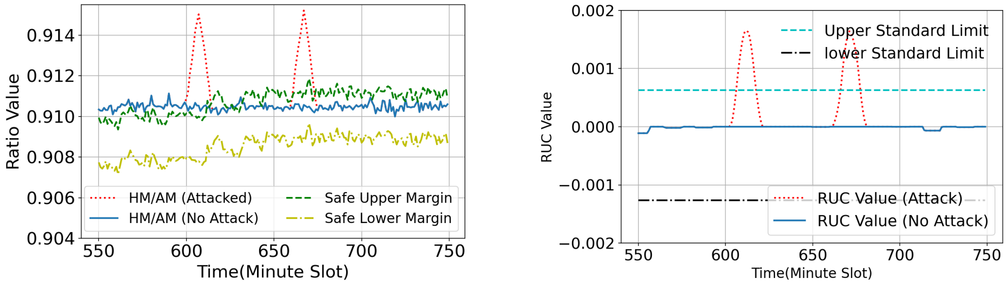


(a) (b)

Figure 3.8. Anomaly Detection for Deductive Step Attack. (a) Tier 1; (b) Tier 2.

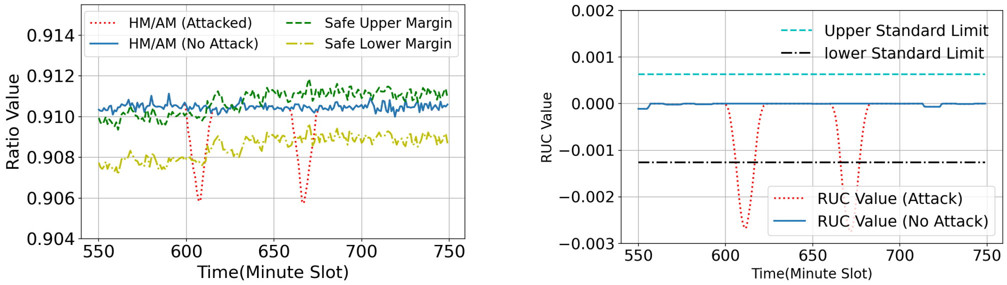
and the remaining data are testing set. We divide this section into two parts: (1) Snapshot Results: that show how our method works under several attack strategies and types (2) Performance Evaluation: that shows the sensitivity versus the false alarm across varying attack margins.

1) Snapshot Results: We randomly selected a period from the test set and introduced an *Step* attack on the current magnitude from A6 PMU with δ_{avg} 1 p.u (which is approximately 0.16 amps). Tier 1 detection scheme to infer the presence of an additive attack is shown in Figure 3.7(a) and subsequently tier 2 is applied to confirm the presence of the attack as shown in Figure 3.7(b). Similarly, for deductive type of attack, Tier 1 and Tier 2 detection schemes are shown in Figure 3.8(a) and Figure 3.8(b) respectively.



(a) (b)

Figure 3.9. Anomaly Detection for Additive Ramp Attack. (a) Tier 1; (b) Tier 2.

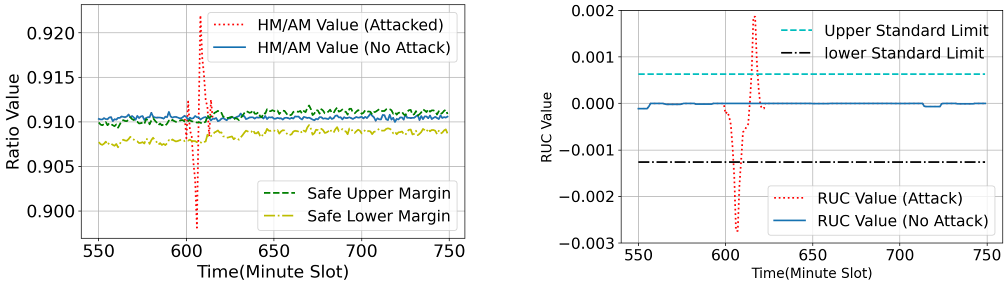


(a) (b)

Figure 3.10. Anomaly Detection for Deductive Ramp Attack. (a) Tier 1; (b) Tier 2.

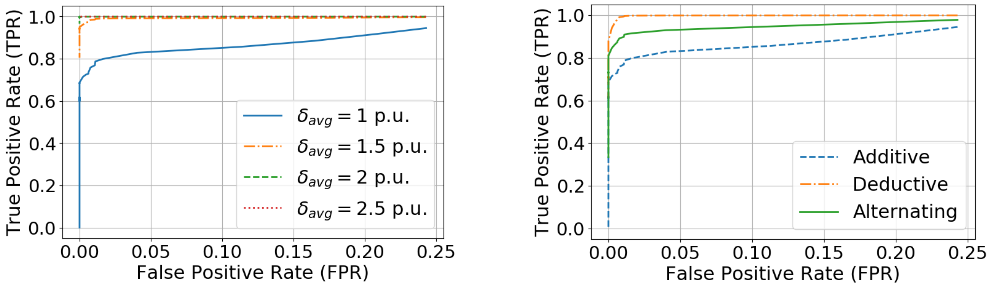
We have also verified our model against *Ramp* and *Mirroring* attack strategy. We have randomly selected periods of 15 minutes to implement the ramp attack for both additive and deductive types with a δ_{avg} of 1.5 p.u. As shown in Figure 3.9 and Figure 3.10, our model is able to detect the attacks for both the cases. Finally, Figure 3.11 shows that our model is able to detect *Mirroring* attack as well in both Tier 1 and Tier 2.

2) Performance Evaluation: For performance evaluation, we generate the ROC curve that characterizes the trade-off between the probability of attack detection vs. the probability of false alarm. we vary the δ_{avg} from 1 p.u to 2.5 p.u ($\approx$ 0.4 amps) using a step strategy to show the ROC for the additive attacks in Figure 3.12(a). A comparative analysis of ROCs of additive, deductive, and alternating attacks for an attack margin of 1 p.u. is



(a) (b)

Figure 3.11. Anomaly Detection for Mirroring Attack. (a) Tier 1; (b) Tier 2.



(a) (b)

Figure 3.12. Performance Analysis using ROCs. (a) ROC for Additive Attack; (b) ROCs for $\delta_{avg} = 1$ p.u.

shown in Figure 3.12(b). A report on accuracy (A), false positive (FP), and false negative (FN) for different attack margins in case of deductive and alternating attacks is given in Table 3.1.

Table 3.1. Experimental Results.

Attack On	Attack Type	Margin(p.u.)	A(%)	FP(%)	FN(%)
Curr. Mag.	Deductive	1	99	1	1
Curr. Mag.	Deductive	1.5	100	1	0
Curr. Mag.	Alternating	1	91	1	9
Curr. Mag.	Alternating	1.5	99	1	1

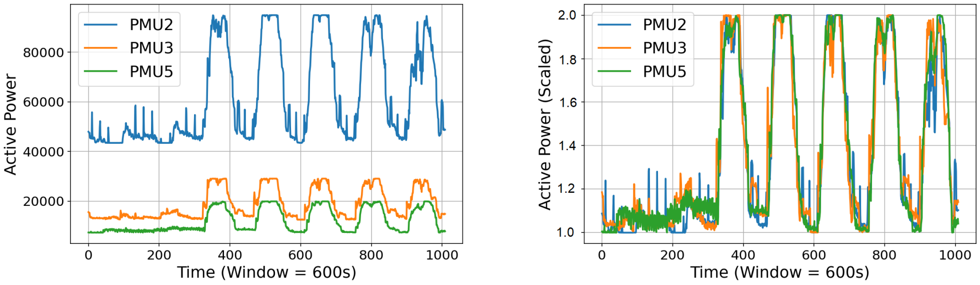
3.5. IMPROVEMENTS OVER PROPOSED METHOD

In our previous method, we have proposed a semi-supervised attack detection model that is deployed on the PDC level which receives the power distribution line information directly from the PMUs in a single hop. If the underlying distribution network has a lot of uncertainties and disturbances the proposed method can generate higher false alarms. However, in a real-time critical infrastructure system such as the smart grid, it is essential to keep the false alarms as low as possible. So we extend our previous work by incorporating noise reduction, data normalization, and adaptive loss function to make our invariant more stable under normal conditions while achieving low false alarms without compromising on the detection sensitivity.

3.5.1. Noise Removal Using Winsorization. We apply standard winsorization of 90% on the raw dataset to remove the instanteneous disturbances from the dataset. We use the same technique on both the datasets. Please note, This is done only for training and cross validation period. we did not winsorize the testing data.

3.5.2. Data Set Characterization and Normalization. PMUs are deployed at different critical locations of the grid network. Due to this reason, the active power information varies broadly from PMU to PMU as illustrated in Figure 3.13a. From the principle of machine learning models variables that are measured at different scales do not contribute equally to the model fitting and the learned model function and thus creating a bias. To mitigate this potential problem feature-wise normalization such as *MinMax* Scaling is usually used prior to model fitting. We apply the same normalization method to our selected data sources. We also added a constant value (= 1) to prevent any sample < 1 as per Equation. 3.9. For notation simplicity, we used P instead of P_{scaled} to indicate active power for our future analysis. Figure 3.13b shows the scaled active power information for 1 week from our training set.

$$P_{scaled} = \frac{P - min(P)}{max(P) - min(P)} + 1. \tag{3.9}$$



(a) (b)

Figure 3.13. MinMax Scaling. (a) Unscaled data; (b) Scaled data.

3.5.3. Optimizing Standard Limits Using Loss Function. We need to calculate an upper and a lower threshold from the RUC values that prevent underfitting and overfitting and improves detection performance. We first adopted Cauchy/Lagrangian loss function to gather the potential set of standard limits. The Cauchy loss is defined as follows:

$$L_{cauchy}(s) = \beta^2 \log(1 + \frac{s^2}{\beta^2}). \tag{3.10}$$

Here β is the tuning parameter and s is the difference between RUC value and the standard limit. We calculate both upper standard limit (τ_{max} and lower standard limit (τ_{min}), for different values of the optimization parameter β using the following algorithms. First, for each potential τ_{max}, We calculate the total loss over $RUC(T)$ by assigning different weights depending on whether the instantaneous RUC value is higher or lower than the τ_{max}. Then we find the optimal τ_{max} that minimizes the loss for a particular β. We use different range of β to calculate the respective optimal τ_{max} and τ_{min}.

Finally we use a cross validation set with minimum δ_{avg} that we target to detect and calculate the *False Alarms* (FA) and *Mis-Detection* (MD). Thus we use the following optimization to select the optimal τ_{max} and the corresponding β.

$$\tau_{max} = \underset{\tau_{max}^{\beta}}{\operatorname{argmin}}(d_1 FA + d_2 MD) \tag{3.11}$$

Algorithm 3 Calculate new τ_{max}

```
input : $RUC(T), [\tau], \beta, w_1, w_2$
output : $\tau^{\beta}_{min}$;
for $\tau \in [\tau]$ do
    $cost = 0$
    for $r \in RUC(T)$ do
        if $r > 0$ then
            $s = r - \tau$
            if $s >= 0$ then
                $cost : \log(1 + (\frac{s*w_2}{\beta})^2)$

            if $s < 0$ then
                $cost : \log(1 + (\frac{s*w_1}{\beta})^2)$


    $TotalCost : \beta^2 * \sum cost$

$\tau^{\beta}_{max} = \operatorname*{argmin}_{\tau}(TotalCost)$
```

Algorithm 4 Calculate new τ_{min}

```
input : $RUC(T), [\tau], \beta, w_1, w_2$
output : $\tau^{\beta}_{max}$;
for $\tau \in [\tau]$ do
    $cost = 0$
    for $r \in RUC(T)$ do
        if $r <= 0$ then
            $s = r - \tau$
            if $s >= 0$ then
                $cost : \log(1 + (\frac{s*w_2}{\beta})^2)$

            if $s > 0$ then
                $cost : \log(1 + (\frac{s*w_1}{\beta})^2)$


    $TotalCost : \beta^2 * \sum cost$

$\tau^{\beta}_{min} = \operatorname*{argmin}_{\tau}(TotalCost)$
```

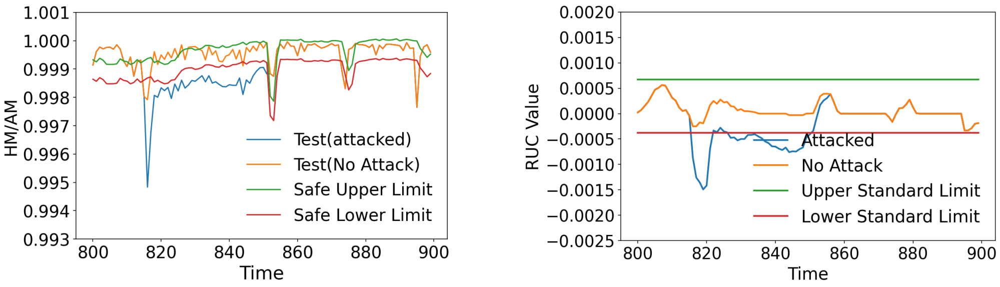
For this, we define two different weights d_1, and d_2 as loss parameters, d_1 represents the loss due to false alarms and d_2 represents the loss due to missed detection. One key consideration in time series attack detection is to minimize false alarms, since the actual probability of being under attack is much lesser. Therefore, seemingly low false alarm rates, do not necessarily indicate a good usable attack detector. Therefore, we need to give more importance to the false alarms. Hence, the loss due to false alarm (d_1) gets more weight, compared to the loss due to missed detection (d_2). In the end, we choose a threshold τ_{max} (and τ_{min}) which minimizes the total loss for the positive RUC samples (and negative RUC samples).

Table 3.2. Parameter Description and Value.

Parameter	Symbol	Value
Attack Margin	δ_{avg}	0.16-1.5
Attack Period	−	2 hours
Optimized Time Window	T	600
Sliding Frame	K	5

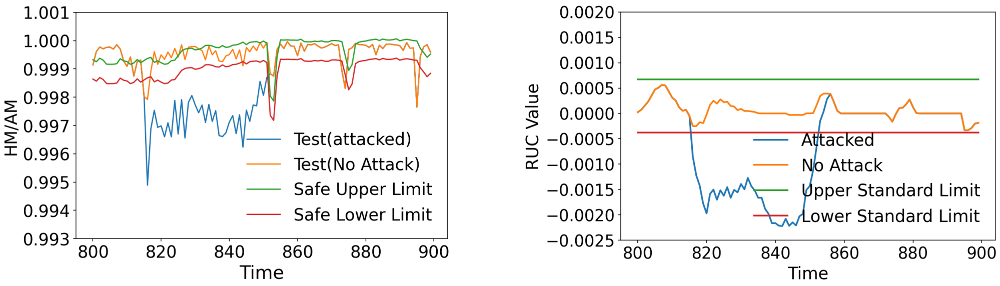
3.5.4. Snapshot Results. Using both the EPFL dataset and LBNL dataset, we conducted extensive experiments for different falsification margins and attack strategies. For the EPFL data, the first four months are considered as the training set and the next two months are considered as cross validation and test data. To show how our method works under several attack strategies and types, we randomly selected a period from the test set and introduced an attack on the current magnitude from PMU2 with δ_{avg}. Tier 1 detection scheme is used to infer the presence of an attack and subsequently tier 2 is applied to confirm the presence of the attack. The parameter values considered for the experiments are shown in Table 3.2.

Figure 3.14(a) and 3.14(b) depicts the tier 1 and tier 2 detection for an additive type of attacks. Similarly, Tier 1 and Tier 2 detection of deductive attack is shown in Figure 3.15(a) and Figure 3.15(b) respectively. In both the cases we have selected a period



(a) (b)

Figure 3.14. Anomaly Detection for Additive Step Attack for δ_{avg} = 15 p.u. (a) Tier 1; (b) Tier 2.

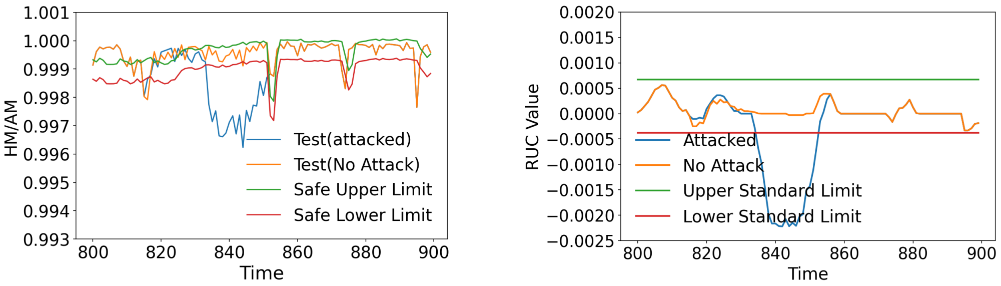


(a) (b)

Figure 3.15. Anomaly Detection for Deductive Step Attack for δ_{avg} = 15 p.u. (a) Tier 1; (b) Tier 2.

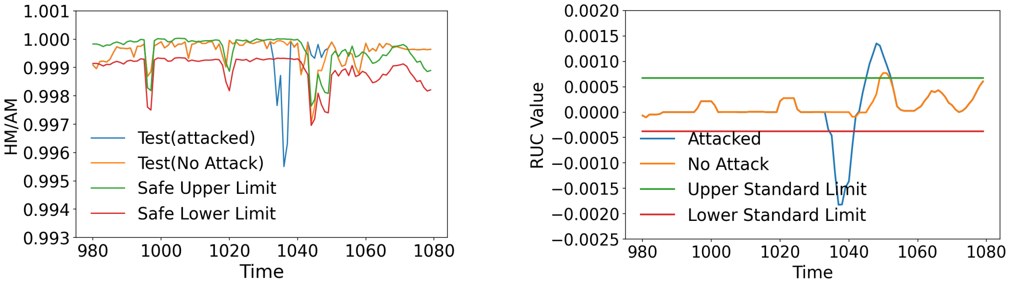
of 2 hours with an δ_{avg} = .5 amps. As discussed in Sec. III, the attacker may choose to use a combination of both additive and deductive attack for equal time duration (*alternating attack*) to increase the chances of evasion. Our proposed method is able to detect such attack as shown in Figure 3.16(a) and Figure 3.16(a). The detection of the mirroring attacks are shown in Figure 3.17(a) and Figure 3.17(a). To implement such attacks we have selected a random period of two hours of a PMU and used the mirror image of the captured time series for the next two hours.

3.5.5. Performance Evaluation. For performance evaluation, we generate the ROC curve that characterizes the trade-off between the probability of attack detection vs. the probability of false alarm. To remove attack period selection bias we introduced continuous attack for the whole testing period with different attack margins (.25-1.5 amps)



(a) (b)

Figure 3.16. Anomaly Detection for Alternating Attack for δ_{avg} = 15 p.u. (a) Tier 1; (b) Tier 2.

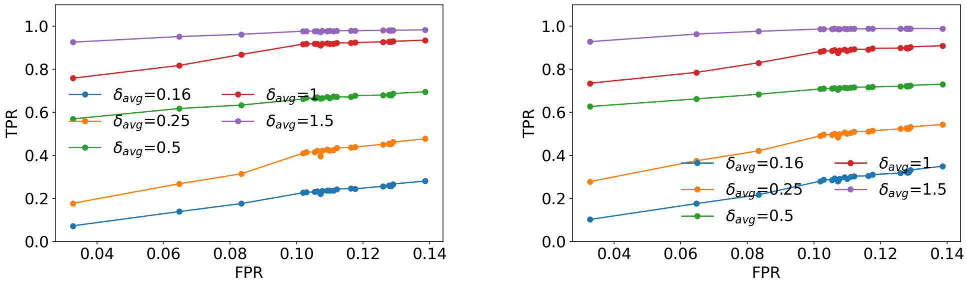


(a) (b)

Figure 3.17. Detection for Mirroring Attack. (a) Tier 1; (b) Tier 2.

and measured the % of samples points outside the standard limit boundaries for different scalar factors (σ). We introduced attack on each PMU and have taken the average of detection rate and false alarms to remove the bias of selected PMU. Figure 3.18(a) and 3.18(b) depicts the ROC for EPFL dataset of additive and deductive attacks respectively. Similarly, Figure 3.19(a) and 3.19(b) depicts the ROC for LBNL dataset of additive and deductive attacks respectively.

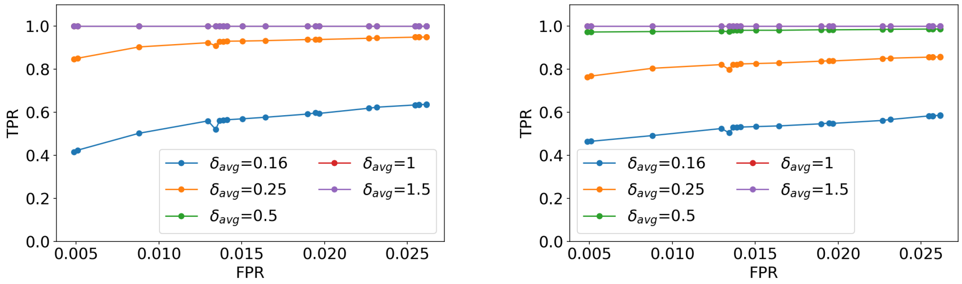
3.5.5.1. Comparison between old and new method. We have compared our old and new methods and found improved performance. We have simulated the same attack strategy and attack margin on the EPFL dataset and compared the ROC obtained for both additive and deductive types of attacks. Figure 3.20(a) and 3.20(b) shows the comparison of the old and new method for additive and deductive attacks respectively.



(a)

(b)

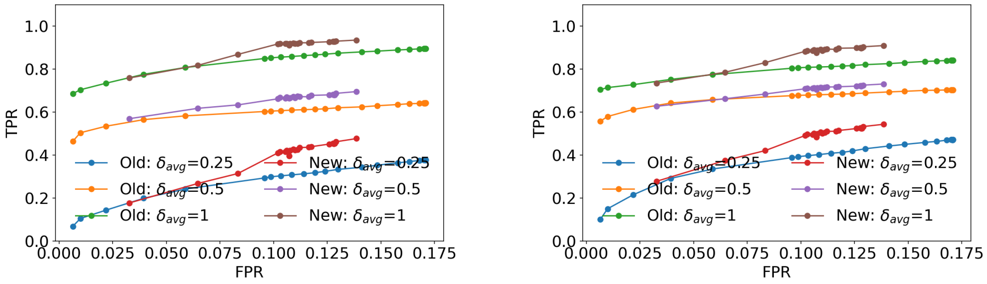
Figure 3.18. Performance Analysis on the EPFL Dataset. (a) ROC for Additive Attack; (b) ROC for Deductive Attack.



(a)

(b)

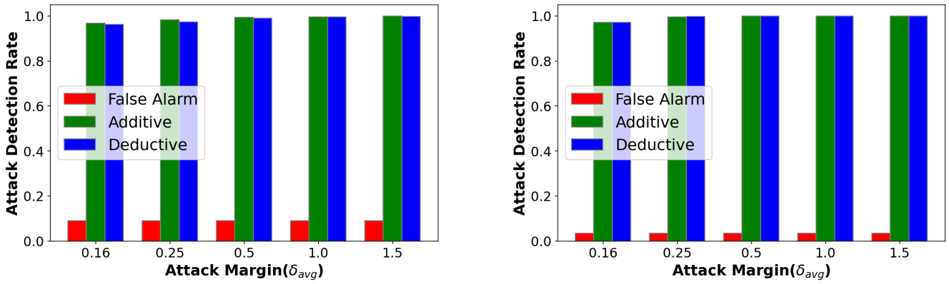
Figure 3.19. Performance Analysis on the LBNL Dataset. (a) ROC for Additive Attack; (b) ROC for Deductive Attack.



(a)

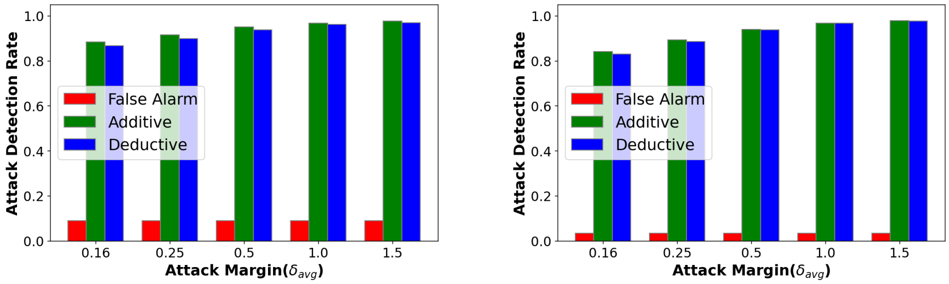
(b)

Figure 3.20. Comparison Between Old and New Method on the EPFL dataset. (a) For Additive Step Attack; (b) For Deductive Step Attack.



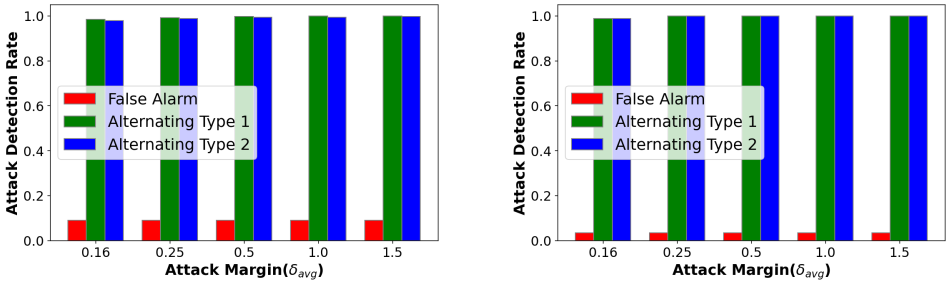
(a) (b)

Figure 3.21. Detection Analysis (Step Attack Strategy). (a) on the EPFL Data; (b) on the LBNL Data.



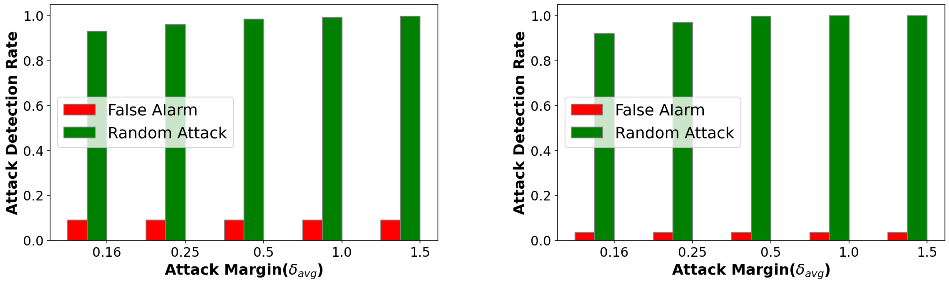
(a) (b)

Figure 3.22. Detection Analysis (Ramp Attack Strategy). (a) on the EPFL Data; (b) on the LBNL Data.



(a) (b)

Figure 3.23. Detection Analysis (Alternating Switching Attack Strategy). (a) on the EPFL Data; (b) on the LBNL Data.



(a)

(b)

Figure 3.24. Detection Analysis (Random Attack Strategy). (a) on the EPFL Data; (b) on the LBNL Data.

3.5.5.2. Detection rate and false alarm analysis. For any anomaly detection detection mechanism it is essential to investigate the false alarms raised. To analyze this we first calculated the false alarm raised in the presence of no attack and then we randomly introduced multiple short term attacks (of length 2 hours) on a different PMUs and checked the detection accuracy. Figure 3.21(a) and Figure 3.21(b) shows the true detection averaged over all the PMUs for EPFL and LBNL dataset respectively. To calculate the attack detection rate and false alarms we fraction of datapoints outside the standard limits in the all attack period combined. It is evident from the plots that our method achieves detection accuracy >= .95 for both the deductive and additive attack types on both the datasets. We investigated Ramp and Alternating attacks in similar strategy. Figure 3.22 and Figure 3.23 shows the performance of our model under *Ramp* and *Alternating Switching* types of attacks.

3.6. INFERENCES

Here, we presented a real time anomaly based attack detection for current magnitude falsification in PMU data streams. We showed that harmonic to arithmetic mean ratios can be used as an effective invariant that is stable without attacks but shows changes during attacks. We showed that even if the attacker has knowledge about the underlying time series we are still able to identify anomalies with a low false alarm rate in real time. Also, unlike many existing bad data detection methodologies, it does not require the topology of the grid network. We also proposed improvements over the generic threshold detection model by incorporating the Cauchy/Lagrangian loss function. We evaluated our model on two different datasets and showed that we are able to achieve a very low false alarm rate without compromising on the detection sensitivity.

4. RESILIENCE AGAINST BAD MOUTHING IN MOBILE CROWDSENSING APPLICATIONS VIA ACTIVE QOI BOOSTING

In the previous section, we discussed a static CPS system where physical devices are deployed in a fixed location and directly forwarding the sensed information to the decision maker in a single hop. We showed that we can achieve high stability under normal conditions (i.e. in the absence of any attacks) by considering the ratio of harmonic and arithmetic mean. Then we showed that our proposed model is able to detect attacks in the system with high accuracy and low false positive rate. Now we want to introduce mobility to the sensing devices. With the introduction of mobility, we have some additional challenges such as dynamic clustering. As the devices are moving, the clustering will change continuously and the location information of the devices will become critical for the security analysis of the system. One classic example of such a system is Mobile Crowdsensing (MCS) where sensing devices transfer the data to the MCS server using *single hop transmission*. Most Mobile Crowdsensing applications deploy rating feedback mechanisms to help quantify the aggregate truthfulness of events (quality of information (QoI)) to improve decision making accuracy. In this work, we first show that factors such as sparseness, inherent error probabilities of rating feedback labelers, and prior knowledge of the QoI models, can be used by strategic adversaries to hijack the feedback labeling mechanism itself and cause QoI score degradation under *bad mouthing attacks*. Thereafter, we propose a randomized rating sub-sampling technique inspired from the philosophy of moving target defense to mitigate the degradation in the QoI scores of truthful events, instead of traditional use of taking all the feedbacks. We offer a game theoretic strategy under various knowledge levels of an adversary and the MCS in regards to picking an optimal subsample size for bad mouthing attacks and QoI calculations respectively, by using a vehicular crowdsensing as a proof-of-concept.

4.1. MOBILE CROWDSENSING SYSTEM (MCS)

The widespread availability of sophisticated mobile (e.g., smartphones, tablets, smartwatches) and IoT devices (smart vehicles, roadside units) and rapid advances in pervasive sensing have fueled the development of the novel paradigm of Mobile Crowdsensing Systems (MCS). A typical MCS paradigm involves an MCS server that accumulates voluntary contributions that are supplied by either autonomous sensing agents (IoT devices) or by humans via an App.

Based on the nature of the MCS application, either a summary statistic of various contributions (*'events'*) or individual contributions are published by the MCS service provider. Such published events guide intelligent human choices or automated decision making that improve quality of life and civic-well being in smart cities. The benefit of using an MCS paradigm is that precise and fine-grained information collection is possible without dedicated infrastructure.

Real life examples of MCS applications include a Vehicular Crowd-Sensing applications such as Google Waze, where contributions are in the form of 'reports' from humans indicating the presence of special road and traffic events (viz., jam, accident, weather hazard, crime scene, speed trap, gasoline prices). Based on such reports, Google Waze infers the most likely event and publishes it on the Waze App that helps in making intelligent changes in traffic route selection. Additionally, most mobile app stores and online social business networks such as Yelp, run in a similar way.

Typically, to evaluate the truthfulness of contributions (known as quality of information (QoI)), there is an additional provision of a feedback monitoring mechanism that allows other users/agents known as labelers or raters, who provide a positive, negative, or uncertain rating labels to the reports/events. However, this feedback procedure motivates a possibility of orchestrated false ratings/feedbacks/labels (known as feedback weaponizing attacks) from a well-organized malicious adversary that biases the QoI scoring mechanism, in a way that negatively affects reputation systems, incentive assignment, and publish decisions. The three established feedback weaponizing attacks are ballot stuffing, bad mouthing, and obfuscation stuffing [35].

Several methods in the past have used variants of Josang's Belief Model [33], Beta Distribution [68], Dempster Shafer Belief [34] for QoI scoring. To date, the research on Quality of Information score, provides active resilience to ballot stuffing and obfuscation stuffing attacks [35] only, but does not actively prevent bad mouthing attacks.

4.2. MOTIVATION

Now we present some unique challenges not adequately addressed in previous works, which motivate our work. These challenges include sparsity in rating feedback population, error probabilities in the feedback apparatus, and targeted feedback weaponizing attacks by adversaries with prior knowledge of the QoI mechanisms in MCS.

First, we argue the sparsity challenge in the feedback apparatus of MCS. Previous QoI models assume that the presence of an MCS automatically implies the availability of a substantial amount of rating population that keeps the relative proportion of compromised feedbacks to the total number of feedbacks low enough for QoI scores to remain unbiased [35]. Nonetheless, this assumption may not be always practical since (i) newly launched MCS have a lower customer base, and hence rating feedback labelers are lower to begin with. A rival business with a practically small attack budget can poison the QoI inference and prevent good reporters from gaining reputation and forcing them to get less incentives. (ii) Second, even when the crowdsensing systems may have a high user base, the geographical spread of this user base may not be spatio-temporally uniform. For example, a downtown area has less crowd during nights, parts of the city can be inherently sparsely populated than other areas. This aspect is further elaborated in Sec. 4.3.5.

Second, none of the QoI scoring models mathematically incorporate the error probability of rating feedback from an honest rater combined with the possibility of feedback weaponizing attacks. When you combine the error probability as well as the presence of attack, we found that even by compromising a minority of the rating population, the QoI with existing models provides biased results on true events. This is elaborated further in Sec. 4.5.3.1.

Third, feedback weaponizing attacks can be better crafted when a rational and intelligent adversary has the knowledge of the QoI scoring models, and he uses that knowledge to craft attacks that creates biased QoIs that do not accurately reflect the veracity of a concerned event. Furthermore, adversary may have knowledge of the defense mechanism through several sources (insider leaks, patent studies etc.)

4.3. RATING FEEDBACK SYSTEMS

Typical QoI scoring methods in MCS includes 3 major phases: (1) accumulation of ratings/feedbacks/labels on the published event, that 'serves as a body of evidence'; (2) quantification of an event's truthfulness score via a Quality of Information (QoI) model that uses the rating feedbacks. (3) classification decision on whether the event is truthful or not by comparing the QoI score with a hard or soft threshold.

4.3.1. Rating Feedback Systems and QoI. The rating feedback systems specify a discrete state space of possibilities a particular rater gives about the authenticity of an event. This state space usually contains three categories: positive feedback (α), negative feedback (β), and uncertain feedback (μ), and goes by different names in different commercial MCS. An example of binary state space is Useful, Not Useful in Waze, and in systems like Yelp have three possibilities for rating a review given by another user. The number of ratings per event can be written as η_α, η_β, and η_μ and a event is denoted as $E : \langle N, \eta_\alpha, \eta_\beta, \eta_\mu \rangle$ where N is total population of ratings, $N = \eta_\alpha + \eta_\beta + \eta_\mu$. The quality of information is quantified by the two popular techniques over decades:

4.3.2. Beta Trust QoI. Beta Trust QoI views the state space as binary. The original Beta Reputation Systems [68] can be applied to both binary and ternary evidence state spaces. The QoI is quantified as:

$$QoI_{beta} = \frac{\eta_\alpha + 1}{N + 2} \qquad 0 < QoI_{beta} < 1 \tag{4.1}$$

where QoI_{beta} is the event truthfulness, η_α is the total number of positive feedbacks, N is the total feedbacks received.

4.3.3. Josang's Belief Model based QoI. Josang's Belief Model explicitly handles uncertainty in the evidence, by specifying expected truthfulness (E) by the following linear score:

$$QoI_{jo} = b + (a).u \quad \text{where} \quad b = \left(\frac{\eta_\alpha + 1}{N + 3}\right); u = \left(\frac{\eta_\mu + 1}{N + 3}\right) \tag{4.2}$$

The parameters b and u are the degrees of belief and uncertainty respectively and a is relative atomicity parameter that decides the extent to which uncertainty should contribute to truthfulness (benefit of doubt). The value of a is 0.5 if there is no prior information available, such that $0 < QoI_{jo} < 1$ acts as a linear predictor of the estimated truthfulness of an event.

4.3.4. QnQ Belief Model. In QnQ Belief model [35], the QoI is represented as:

$$QoI_{QnQ} = w_b.b + w_u.u \tag{4.3}$$

where b and u are the same as Josang's Belief Model, and w_b, w_u are known as *Belief Coefficient* and *Uncertainty Coefficient* respectively with the limits: $0 < \{w_b, w_u\} < 1$, such that $0 < QoI_{QnQ} < 1$. The equations for w_b and w_u are specified by non-linear functions such as generalized richard's curve and Kohlsrausch relaxation functions. The mathematical details of w_b and w_u, can be found in [69]. The framework provides active resilience to ballot and obfuscation stuffing.

We found that while ballot stuffing and obfuscation attacks are actively prevented by QnQ, the bad mouthing attacks are passively by virtue of the assumption of large crowd. This means that when the rating sample contains a higher fraction of compromised raters (happens when the total ratings received are sparse), the attacks in the presence of high negative ratings to true events cause high QoI degradation for true events, since the weights are characterized slow growth sigmoidal nature. This causes low QoI scores and poisons the reputation system when adversary controls a high proportion of the rating population. In short QnQ works well under rse samples but only for ballot stuffing and obfuscation stuffing attacks but not bad mouthing attacks.

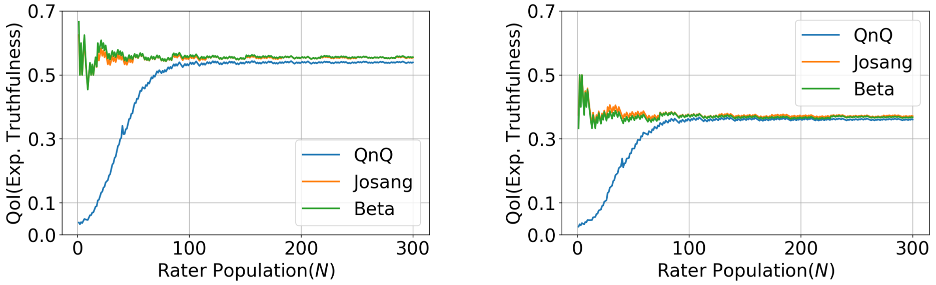
4.3.5. Relationship between Voting and QoI Scoring. Majority voting is central to dependable decision making in cooperative distributed systems. Now, we show that majority voting is deeply related to how QoI scores are interpreted for event inference and publish-subscribe decisions.

A binary rating state space can be viewed as binary votes equivalent to voting for an event. Suppose the MCS receives two events whose evidences are specified by $E1 : \langle 20, 30 \rangle$ and $E2 : \langle 21, 20 \rangle$, such that the tuple $\langle \eta_\alpha, \eta_\beta \rangle$ indicates number of positive and negative ratings received. The decision by majority voting would indicate that the event is false/not useful for E1, while it will indicate a true/legitimate event for E2. However, majority voting is a hard decision rule and lacks intelligence in the sense that it cannot quantify confidence on the event's likelihood of being actually true or false. For e.g, if $E3 : \langle 98, 2 \rangle$, the answer will still be a true event and there will be no difference between E2 and E3 although E3 quality of being actually true is higher than E2.

In contrast, QoI scoring models such as beta trust [68], have its roots in Artificial Intelligence, and enables the similar outcomes but via a soft decision rule that also allows embedding the notion of confidence into event decisions. Let us take the same example of E1 and E2. The E1's QoI score via a Beta trust model is $\frac{20+1}{20+30+2} = 0.4038$, while E2's QoI is 0.5116.

The MCS uses the QoI for various purposes: (i) to decide between a true versus a false event (a binary classification problem) (ii) use QoI scores to proportionally update reputation of those users who reported that event (iii) update reputation is used to decide user incentive which depends on the QoI scores. The first purpose (deciding between true versus false event) is a logistic regression problem, where a logit link function translates the linear predictor (QoI) to the output decision class [70]. A negative score probabilistically indicates the event is likely false and a positive score means the event is likely true [35, 71], *because mid-point of 0.5 is assumed as a neutral decision boundary between true versus false event inferences*. Therefore, for E1, E2, and E3, we get $log(0.40/(1-0.40)) = -0.16$ and $log(0.51/(1-0.51)) = 0.020$, and $log(0.97/(1-0.97)) = 1.5$ respectively. Therefore, E1 and E2 will be inferred as a false and true event respectively. We can see that the final inference is the same in both voting and QoI models. However, QoI scoring with AI methods unlike voting allows to distinguish between two events that are positive in their score (E2 and E3), but one event with a greater score the decision maker has more confidence that it is indeed true, which helps in proportional update of reputation and incentive. This makes the AI based approach to QoI more intelligent than the voting, although decision output wise, they are similar.

For ternary rating state space, [71] showed that one accommodates uncertainty by partial splitting of uncertain ratings in the ratio of the observed positive and negative ratings and adding it to the positive feedback, if chances of any one component getting compromised is uniform. (instead of using 0.5 as relative atomicity as in Josang's Model). This gives modified effective positive and negative ratings; which can be applied to Beta



(a)

(b)

Figure 4.1. QoI under Bad Mouthing. (a) Majority honest; (b) Majority compromised.

QoI [71]. This simplifies the mathematical tractability. For example, if there are 10 positive, 8 negative and 4 uncertain ratings, the effective positive ratings are obtained by partially splitting 4 uncertain ratings in the ratio of the observed η_α and η_β such that the modified positive votes are:

$$\eta''_\alpha = \eta_\alpha + \left(\frac{\eta_\alpha}{\eta_\alpha + \eta_\beta}\right).\eta_\mu; \quad \eta''_\beta = N - \eta''_\alpha; \quad QoI = \frac{\eta''_\alpha + 1}{\eta''_\beta + \eta''_\alpha + 2} \tag{4.4}$$

Intuitively, if the majority of the rating population is honest, it implies the Beta trust model, Josang belief model delivers a QoI score more than 0.5 (converse may not be true) (As depicted in Figure 4.1(a) where 60% users are honest). However, in the presence of bad mouthing attacks, if majority of the rater population is compromised (See Figure 4.1(b) with 60% compromised raters), then the QoI scores of all the previous models degrade below 0.5, failing into label the event true.

4.4. SYSTEM AND THREAT MODELS

In this section, we present the abstraction of the crowdsensing system model and the threat of bad mouthing of events in sparse feedback based crowdsensing samples, and some novel considerations that have not been incorporated in a wholistic manner in previous works.

4.4.1. System Model. We assume a network of U users/devices subscribed to a crowd-sensing application. Users have three types of roles for a given event E: (a) reporter (b) rater (c) passive consumers who do not participate in either reporting/rating.

- Events: Any event has an event boundary and a time period of viability. The users inside this boundary and within the time period are liable to rate on this event. The reporters cannot rate the event reported by themselves. Similarly, a rater can rate a single event only once.

- Reporters: Reporters are the set of users who report to the CS server indicating an 'event' of interest. The CS server publicly publishes, either the individual reports or the statistical aggregate of all reports, as the 'event' of interest, to all other users of that application during the cold-start phase. The same event can be reported by many reporters within a time epoch.

- Raters: Raters are the set of users (humans) or machines (drones) who provide a feedback/rating/label on the published events which indicate the crowd-sensed perception of the relative goodness of the published events from the CS server. We denote the total number of ratings received for a given event as N, and the number of positive, negative, and uncertain ratings are η_α, η_β, η_μ respectively.

- Rating Probabilities of Honest Users: The probability p_b is the probability that an honest user accurately rates a legitimate event as true. Any rater when not controlled by an adversary has an error probability of p_e with which it mistakenly rates a true event as false. For example, in weather conditions with low visibility can produce errors in judgment. The probability that an honest user rates a true event as uncertain is p_u. These probabilities might depend on the inherent level of environmental uncertainty that might cause them to rate a true event as uncertain. So from the above, for an honest user $p_b + p_u + p_e = 1$. Therefore, the probability that an honest user does not rate a true event with positive feedback is $1 - p_b$. *This is a robustness issue rather than a security issue, but it affects QoI negatively.*

4.4.2. Threat Model for MCS. Let us discuss several aspects that specify the threat model.

- Bad Mouthing Attacks: The rogue raters may be controlled by an organized adversary launching Bad Mouthing Attacks to give false ratings to true events, causing legitimate events to get low QoI scores. This not only prevents the published legitimate event to be deleted but also undermines inputs from honest reporters, degrading their reputation, and also suppressing events from such honest reporters in future [35].

- Attack Scale: The attack 'budget' B is the maximum number of unique raters adversary can afford to compromise. The Adversary uses knowledge of defense mechanism and MCS to decide whether to apply his full budget or not. If purpose is served without using the full budget, then it is a rational choice for the adversary. The 'attack scale' is the fraction of compromised ratings to the total number of ratings (from the perspective of the defender). Hence, if the rating population is low, an attacker with a low budget can dominate the rating population with biased ratings.

- Adversarial Capabilities and Budgets: Among N users that have provided ratings to a particular truthful event j, K users are not compromised by the malicious adversary, while $N - K$ users are compromised by the adversary. In many MCS, the rating feedback mass N for an event/item is often sparse for two reasons. First, when an application is launched initially, the user base is not very high which reduces the chances of getting a high feedback mass. Second, the lack of motivation or incentives to provide ratings. We assume an MCS system where the number of raters can be less due to the following reasons: (i) this is a newly launched MCS application with a low initial user base; (ii) the application runs in a town with a lower population size;

(iii) the application is running in a part of the city which has a sparse population on specific times. When N is small, the adversary's attack budget $N - K$ can dominate the proportion of the rating sample N leading to altered event QoI values. This is because malicious users are motivated by their attack rationale and coordinated in their rating behavior.

- Adversary assumptions: (1) The adversary knows that the defender may use a traditional QoI method or sub-sampling QoI method. (2) We assume that the adversary knows that there is a true event and therefore a compromised rater's p_e does not affect the output of the compromised rater. (3) Adversary has enough budget such that he can compromise majority of the rating users if required. This makes sense due to the sparse sampling problem in MCS. However, once it has compromised a rater, it exhausts a resource and it cannot change it. (4) Adversary can possess knowledge about the rater population size N at given time/place and the probability of accuracy, p_b, of the honest rating users. (5) An intelligent rational adversary would always try to maximize his gain and will not have any strategy that causes net loss. (6) We study the problem only when there is a split vote of at least 30% or more for either true or false event. If it is more than that then it is a dogmatic system and a mitigation strategy does not make sense.

4.5. PROPOSED APPROACH

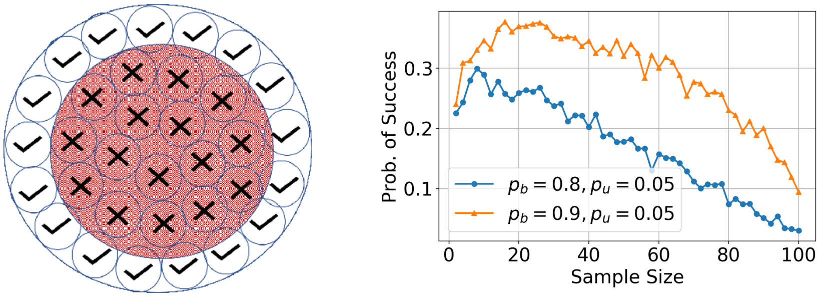
The proposed approach is divided into four parts. First, we show how and why an MTD approach is required for the problem. Second, we show that the strategic situation between the MCS and the adversary can be modeled into a two person game theoretic formulation with many alternative strategies. Third, we provide a mathematical analysis of the randomized subsampling approach and the adversary's cost benefit analysis that helps to prune the strategy space of the attacker defender game. Fourth, we show the final game with the reduced strategy space and solve for best rational strategies for both MCS and the adversary.

4.5.1. Randomized Rater Sub-Sampling as MTD. The philosophy of moving target defense (MTD) [72] mandates the creation of constantly shifting environments by a defender to introduce asymmetric uncertainty between defenders and attackers.The Department of Homeland Security, USA refers to the idea of moving targets in the form of strategies that are diverse and continually shift and change over time to limit opportunities for attack, increase the cost of attacks and/or increase system resiliency. The guidelines in [72], specify dynamic system randomization at run time is one of the overarching ways for a defender to constantly change the effective attack surface.

Most classical defense methods focus on host/resource/user level detection methods. They suffer from scalability and agility concerns in MCS because: (i) network sizes always vary rapidly as a function of time, (ii) new users quickly join or leave, and (iii) behaviors keep evolving. Therefore, in MCS, it is rather prudent to think about mitigation of or resilience against attacks rather than focusing on detection of attacks and attacking devices. This is complicated by how easy it is to bias the QoI under sparse rating samples and change the inference on the event completely, if the classical QoI approaches are used.

Hence, due to the above reasons, we concluded that an MTD approach is a better alternative rather than focusing on detecting whether an individual feedback rater is providing a false rating or not. Motivated from the above, we propose a way to bring ideas of moving target defense into QoI scoring for a more resilient MCS.

DEFINITION 1: Randomized Subsampling: *Our main idea of dynamic system randomization is we hypothesize that instead of using all the ratings received for calculating a QoI score, the defender uses a subsample of a strategic size from the set of ratings received, and then calculates the QoI of the event calculated from rating counts with that subsample.*



(a) (b)

Figure 4.2. Subsampling Intuition and Optimal Sample Sizes. (a) Intuition; (b) Optimal Sample Sizes.

Our hypothesis is given credence by the following: Figure 4.2(a), offers a visual intuition of why randomized subsampling can mitigate the effects of feedback weaponizing attacks. We have already established in Sec 4.3.5 that if the majority of the ratings involved in the calculation of QoI Score is not compromised, the QoI score progressively moves towards the correct event status (because of being above 0.5), with higher truthfulness in the presence of bad mouthing attacks. There are three key considerations from Figure 4.2(a). (1) The big outer transparent circle represents the full sample N (total number of ratings received); (ii) the medium inner red shaded circle represents a set of compromised ratings which is a subset of N but greater than 50% of N. (iii) the smallest circles with a tick or a cross represents a random subsample of a certain size (say n) from the set N (represented by the smaller circles in Figure 4.2(a). The tick corresponds to samples where majority of the ratings in that sample correspond to non-compromised rating set, while crosses correspond to the opposite case.

We can conclude visually that the red circle size is more than half the outer circle. In this case, a full sample majority voting will always result in an outcome that corresponds to what the inner red circle indicates, which will be misleading and therefore considered a failure for the defender MCS. In this case, the defender has zero chances of success in predicting a QoI that should be greater than 0.5.

However, if the defender picks a randomized subsample of a certain size n (represented by the smaller circles in Figure 4.2(a) from set N; there are $\binom{N}{n}$ such possibilities). Without knowing which ones are compromised, we can see that success probability is non-zero. and thus better than using the whole sample. Regardless of the number of selections with ticks, it is better than zero chance of success using full sample.

To give credence to the above hypothesis, we conducted a numerical simulation to mimic the following scenario: A true event receives N=100 ratings with 48 of them controlled by adversary giving a negative rating and $p_b = 0.9, p_u = 0.05$ for the remaining honest raters. We then perform an exhaustive search of the subsample search space to figure the fraction of times (out of 1000 rounds of iteration) a subsample of a given size contains a majority of positive ratings (which is the accurate rating for the concerned event) in that selection. *This fraction is termed as the probability of success*, because if a majority of true ratings in a subsample guarantees a $QoI > 0.5$ required for inferring the event correctly.

Figure 4.2(b), shows the y-axis as the probability of having more 50% of positive ratings versus different possible sample sizes. We can see that the subsample size 8 maximizes this probability as the above mentioned settings. For another instance with $N = 100, K = 60, p_b = 0.8, p_u = 0.05$ the optimal sample size is 16. It also shows that even if the adversary compromised just short of majority of the rating sample, the classical use of using the full rating sample for QoI calculation causes a probability of success = 0 (rightmost point on the x-axis). In this case, 30% of the time, the MCS was successful in obtaining a QoI above 0.5.

4.5.2. Attacker Defender Game. Here we first discuss that a game theoretic approach is possible in the strategic situation above.

4.5.2.1. Defender's perspective (MCS). In the previous sections, we provided an intuition that using a subsample of a certain size taken as a strategy by a defender will provide better outcomes if the majority is compromised. However, this size could be anything from 1 to $N - 1$.

However, the defender knows that adversary possesses knowledge about its option of a randomized subsampling approach. Based on this knowledge, defender expects that adversary would re-adjust its strategy accordingly to compromise a minority of ratings (if and as long as its rational). Thus, there is a possibility that adversary can might hope to compromise an effective minority of the population if the defender chooses a sub-sampling approach. Hence, choosing the full sample N could also be a viable strategy.

4.5.2.2. Adversary's perspective (MCS). The Adversary expects the defender to play a sub-sampling strategy if it compromises a majority of the ratings. However, it also knows that the $1-p_b$ is a feature that helps its cause. Thus, to overturn an event decision, the actual number required to compromise will be lesser than 51% of the total ratings received.

On the other hand, the adversary knows that subsampling will provide better outcomes for defender, it might contemplate to switch to strategy with minority of ratings compromised to confuse the defender. Therefore, the adversary theoretically has a option to compromise 1 to $N-1$ raters, unless subject to any upper bound on its attack budget or any strategy where is gain is lesser than its investment.

The above presents a huge search space for finding equilibrium strategies in the attacker defender game. To be rational, the defender should select an optimal sub-sample size if any, that will maximize the probability of success out of all possible subsample or full sample sizes that could be potential strategies. The adversary will try to maximize its net benefit and would avoid any strategy that leads to a loss.

Therefore, our next effort will to prune the search space of strategies for defender and adversaries before we formalize the game formulation.

4.5.3. Pruning the Defender MCS Strategy Space. In this section, we will provide an analysis of success probabilities with the randomized subsampling technique with the use of hypergeometric distribution. The security status of the rater can be classified into one of the two mutually exclusive categories; compromised or non-compromised. Each rater is counted as a success if not compromised (or failure if compromised). Let there be K number of non-compromised and $N - K$ number of compromised raters in the received rating sample of size N. Let X be a random variable, denoting the number of non-compromised ratings in a chosen subsample from the population. Given that a subsample of size $n \leq N$ is drawn from such a population of size N, the probability of observing exactly k number of non-compromised raters from a population that originally contains exactly K non-compromised raters, can be specified by the pmf of a Hyper-Geometric distribution:

$$P(X = k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}} \tag{4.5}$$

The Equation 4.5 assumes that each non-compromised rater is never wrong in their judgment and produces a correct rating at all times. However, in the MCS model, each honest rater is either a human or sensor/drones that are prone to errors (p_e) and uncertainty (p_u). Hence, we need to modify the hypergeometric pmf, to find the effective true success probability of observing exactly k number of successes. Note that for MCS, success indicates those ratings that contribute to trust, and not the security status of the raters.

4.5.3.1. Embedding error of rating labelers. To simplify, first we just assume a binary rating system ($p_u = 0$), and then extend the analysis for ternary rating space in Sec 4.5.3.4. Given exactly k honest raters are picked in a sub-sample, the chances that there will be at-least l or more successes (positive ratings) out of the k honest candidates, can be modeled by a binomial distribution:

$$p(Y > l | X = k) = \sum_{j=l}^{k} \binom{k}{j} (p_a)^j (1 - p_a)^{k-j} \tag{4.6}$$

where l is the lower bound required for successes under the normal fusion rule. For example, in case of majority voting $l = \frac{n+1}{2}$ (if n is even), and p_a is the chance that an honest rater provides a rating that eventually contributes to the increase in QoI. The interpretation of how p_a is calculated changes with whether one is using a Beta Trust for QoI scoring and will be elaborated separately in Section 4.5.3.4.

As l is the lower bound required for successes, based on whether subsample size is even or odd (n), l needs to be changed. If sample size is even, $l = \frac{n+1}{2}$ to ensure number of successes are majority in the selected subsample. In contrast, if we take odd samples then $l = \frac{n}{2}$ to obtain the majority of selected subsample. In this analysis, we choose a discussion for even samples only, but it works seamlessly for odd samples with the only change in equations being $l = \frac{n}{2}$.

4.5.3.2. PMF of success under attacks. Here we will derive the probability mass function for achieving a success where dominant rating attacks and errors coexist. Intuitively, the probability mass function will have a positive value, if there are at least $\frac{n+1}{2}$ honest raters are in the sample of size n. However, in theory, if the number of dishonest raters is very low, then the minimum of honest raters in a larger sample size can be greater than $\frac{n+1}{2}$. The following is an illustrative example that explains this:

- Lower Bound on Honest Raters: Considering two scenarios for a rating population of size $N = 100$. In scenario I, the number of honest raters $K = 40$ is in minority, while the rest of the $N - K = 60$ is compromised. Scenario I, represents a case of organized and orchestrated attack. In contrast, in the scenario II, the number of honest users $K = 90$ is in the strong majority, and $N - K = 10$ represents individual selfish users or a small group selfish users. Now consider the candidate subsample size $n = 70$. For scenario II, the $(n + 1)/2 = 36$. However, given that there are 90 honest raters, even in the worst case, if all dishonest raters fall my sample of size $n = 70$, the minimum possible honest users in this selected sub-sample is $n - (N - K) = 60$. Notice, however, that 60 is greater than 36. Therefore, the lower bound of Scenario II is 60 and not 36. For scenario I, the lower bound on the honest users is $(n + 1)/2$. Hence, the lower bound of the number of honest users in a sample of size n required for a success is given by the following:

$$k_{min} = max(\frac{n+1}{2}, n - (N - K)) \tag{4.7}$$

- Upper Bound on the Number of Honest Raters: The maximum possible number of honest raters that can be picked depends on whether the total number of successes K is larger or smaller than the candidate sample size. Hence, the maximum possible upper bound for the number of honest raters:

$$k_{max} = min(n, K) \tag{4.8}$$

- Probability of Success under attacks: The probability mass function of the resultant success in having a majority of the rating labels chosen as authentic labels is given by the following:

$$P(n) = \frac{\sum_{i=k_{min}}^{k_{max}} \binom{K}{i}\binom{N-K}{n-i} \sum_{j=(n+1)/2}^{i} \binom{i}{j}(p_a)^j(1-p_a)^{(i-j)}}{\binom{N}{n}} \tag{4.9}$$

4.5.3.3. Optimal sampling size for subsampling. Intuitively, the MCS server's policy should be to pick that random sub-sample of size n_{opt} that maximizes $P(n)$, and this would be the MCS's most rational strategy. Mathematically, we can write it as the following:

$$n_{opt} = \underset{n}{\operatorname{argmin}}(P(n)) \qquad P_{opt} = P(n = n_{opt}) \tag{4.10}$$

The above equation gives the optimal sub-sample size n_{opt} for a given $N - K$ (no. compromised raters) such that the probability of success (in terms of getting more than 50% of the correct rating we desire in that sample), is maximized. Hence, among all possible subsampling strategies, the defender will pick n_{opt} since it maximizes its success.

Now since we have already shown that the majority of ratings belonging to the true class is directly related to getting the desired QoI scores and accurate event truth inference, we would like to analyze how Beta Trust and Josang's Belief Models perform under the evidence restricted to this optimal subsample size.

4.5.3.4. Interpretation of p_a in ternary state space. Both positive and a portion of uncertain feedbacks (the benefit of doubt) contribute to the QoI score. We envision the positive and any portion of the uncertain feedbacks that contributes to the QoI score as a 'success' since it contributes to increase in the QoI score, when there is a true event (from Section 4.3.5). Hence, p_a needs to be expressed in terms of p_b and p_u. Note that, the compromised $N-K$ raters always provide a negative feedback. Therefore, the probability of having an honest user giving a rating that contributes to the QoI score is $p_a^{beta} = p_b + p_u\frac{K}{N}$. Therefore, the *theoretical result* on the probability of success is given by the following:

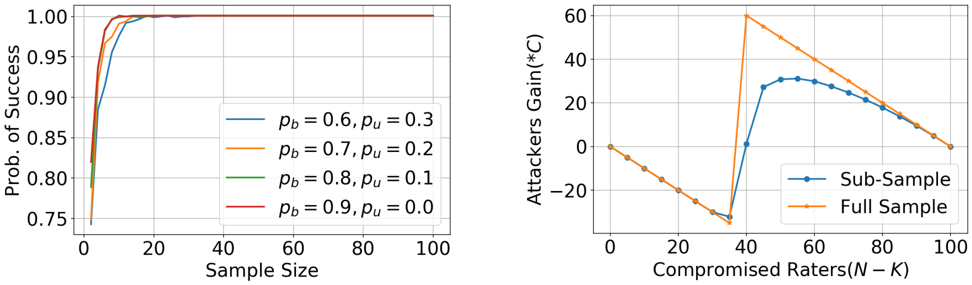
$$P^{beta}(n) = \frac{\sum_{i=k_{min}}^{k_{max}} \binom{K}{i}\binom{N-K}{n-i} \sum_{j=(n+1)/2}^{i} \binom{i}{j}(p_a^{beta})^j(1-p_a^{beta})^{(i-j)}}{\binom{N}{n}}. \tag{4.11}$$

$$n_{opt}^{beta} = \underset{n}{\text{argmin}}(P^{beta}(n)) \qquad P_{opt} = P^{beta}(n = n_{opt}) \tag{4.12}$$

where P_{opt} *is the optimal probability of success* and n_{opt} *is the optimal sample size* that produces P_{opt}. In the experiment section, we will compare the theoretical result with the experiment to prove the correctness of the equation. A similar derivation can be done under Josang's Belief Model.

4.5.3.5. Pruning of adversary strategy space. In a real world MCS, the adversary can choose not to attack the system all the time so there may be the absence of attacks, but p_e may still cause false ratings and degrade QoI. To examine whether this limits the game formulation, we assessed the effect of optimal sampling under no attacks, using a numerical simulation shown in Figure 4.3(a). We observed that optimal sample size converges to N (using Equation 4.11), if no attack is present in the system (regardless the p_b).

Now if adversary chooses to attack, we can prune its strategy space accordingly: If attacker incurs a uniform cost, C, for compromising a single rater, the net payoff G of the adversary is defined as follows: With an investment of $(N-K)$ compromised raters, the adversary can overturn a decision worth the same as if it compromised all raters N.



(a)

(b)

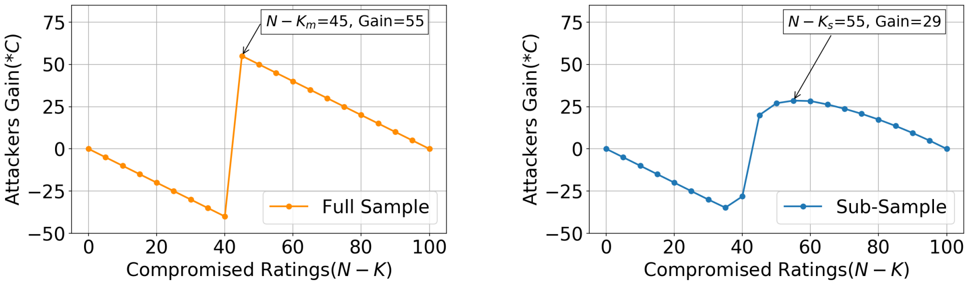
Figure 4.3. Pruning of Adversarial Strategy. (a) No attack as a strategy; (b) Attacker's Gain $(p_b = 0.7\ and\ p_u = 0.25)$.

Therefore, investing in $(N - K)C$, it gains a return equivalent of $N.C$ but this happens only with a probability of $(1 - P_{opt})$, given defender's rational choice of optimal sub-sample size. So the net payoff G is computed as the difference between its return and investment. Hence,

DEFINITION 2: The net payoff of the adversary $G = (1 - P_{opt})NC - (N - K)C$. where $(N - K)C$ quantifies the adversaries' investment, given $(N - K)$ is the compromised rater count.

The Figure 4.3(b) shows adversary payoff for each possible value of compromised raters $(N - K)$ when the defender plays its optimal sub sample size for a given $N - K$ (blue line) and full sample size (orange line).

From Figure 4.3(b), we make a key observation: when the $(N - K)$ is lesser, the attacker's net payoff under sub-sampling is larger than a full sampling approach. However, the G in all such cases is negative, which indicates a net loss for the adversary. Thus, we conclude that a rational adversary who wants to make a net profit, will at least want a net payoff to be not negative, and hence the minimum value of compromised raters for which G is positive is $N - K_m$. Therefore, we can prune all $N - K$ candidates below $N - K_m$ from the adversary's strategy space regardless of the what the defender plays.



(a)

(b)

Figure 4.4. Attacker Payoff. (a) Full Sample; (b) Sub sample.

Interestingly enough, the minimum bound of $N-K_m$ compromised raters that provide a positive payoff is also the maximum payoff achievable by the adversary for any value of candidate $(N-K)$ greater than $N-K_m$, when defender uses a full sampling strategy. This can be verified from Figure 4.4(a) (orange line). Therefore, it is safe to conclude that the adversary's best strategy is to compromise $N-K_m$ (*Strategy 1*) if the defender plays a full sampling strategy.

On the other hand, let us turn our focus onto the best possible adversarial response if the defender chooses the sub-sampling approach. The Figure 4.4(b), shows the payoff of the adversary as a function of all possible $N-K$ when the defender plays a sub-sampling. We can observe that there is a particular compromised $N-K_s$ which maximizes G. Therefore, it is safe to conclude that adversary's best strategy is to compromise $N-K_s$ (*Strategy 2*), if defender chooses to play a sub-sampling approach. Note that these choices of $N-K_m$ and $N-K_s$ is dependent on p_a, because G is related to $1-P_{opt}$, and P_{opt} is directly related to p_a.

Figure 4.4(a) and 4.4(b) shows how attacker's gain G changes with number of compromised raters under both full sampling (orange line) and subsampling (blue line) strategies, for a given p_b and p_u. Thus the attacker can compute the number of raters in the whole population that it needs to compromise such that it would maximize his gain, by taking that compromised raters (on the x-axis) maximum of the blue and orange lines correspond to $N - K_s =_{N-K} (G(n))$ and $N - K_m =_{N-K} (G(N))$ for the given setting. It is evident from the diagrams that $N - K_s > N - K_m$ or $K_m > K_s$.

4.5.4. Game Theoretic Formulation. Here we present a game theoretic approach for selecting a best rational strategy in the presence of bad mouthing attack under a complete information zero sum game between the attacker and defender where both are aware of p_a and N and each other's possible set of strategies.

Here we assume a sophisticated adversary with knowledge of p_a and N; and the fact that the defender can either take a sub-sample or a full sample approach. The adversary also knows that given defender chooses the subsampling strategy it would pick a subsampling size that offers maximum number of uncompromised ratings in that sample (that yields P_{opt}). The defender MCS also knows that the adversary has two options of calculating optimal compromised number of raters that would maximize the adversary's gain, for each of its optimal subsampling or full sampling strategy. We have not included the case of no attacks as there is no loss of the attacker and maximum probability of success under the subsampling approach, even in presence of bad mouthed ratings due to the inaccuracy of honest users, the resultant subsample size from our Equation 4.12 still converges to the full sample N as shown in Figure 4.3(a).

Now an intelligent adversary would always try to maximize its net gain under different strategies by the defender. To achieve this he needs to find the optimal number of required manipulated raters under two alternative strategies:

- Strategy 1: Compromise $N - K_m$ raters which maximizes G under full sampling strategy by defender.
- Strategy 2: Compromise $N - K_s$ raters, which maximizes G, if defender incorporates the sub-sampling strategy.

The corresponding payoffs for each strategy are calculated in terms of adversary's gain/loss. Depending on the strategy incorporated by the attacker and defender, the payoffs of the players are derived as follows:

4.5.4.1. Payoff calculations for strategies.

We first calculate the two pay-offs (for Strategy 1 and Strategy 2) from the perspective of the adversary, given the defender plays a full sampling strategy.

When full sample is selected as a defense strategy, then the attacker is able to compute the minimum number of manipulation to overturn the decision of the defender. To perform this, adversary must ensure that $K.p_a \leq \frac{N}{2}$. Hence, $N - K_m = N(1 - \frac{1}{2p_a})$ and the payoff with strategy 1 (i.e. $N - K = N - K_m$) is given by: $G1(full) = (1-0)NC - (N-K_m)C = K_mC$.

Since $N - K_m < N - K_s$, if the P_{opt} drops to zero at $N - K_m$, it will remain zero under $N - K_s$ (Strategy). Therefore, the payoff under Strategy 2 is given by $G2(full) = (1-0)NC - (N-K_s)C = K_sC$.

We now calculate the two pay-offs (for Strategy 1 and Strategy 2) from the perspective of the adversary, given the defender plays an optimal subsampling strategy derived according to Equation 4.12.

Let $P_1(n_1)$ and $P_2(n_2)$ be the maximum probability of success with subsampling method (from Equation 4.12) under adversarial Strategy 1 and 2, respectively; where n_1 and n_2 are the respective optimal sampling size selections for the respective strategies. For notational simplicity, we denote $P_1(n_1), P_2(n_2)$ as just P_1, P_2 respectively.

So the attacker's payoffs with strategy 1 and strategy 2 is $G1(sub) = (1-P_1)NC - (N-K_m)C = (K_m - P_1N)C$ and $G2(sub) = (1-P_2)NC - (N-K_s)C = (K_s - P_2N)C$ respectively. As cost C is a common scaling factor in all the payoffs, the final payoff matrix of the game is given in Table 4.1 after removing C from all the individual payoffs. In Table 4.1 the defender's (D), has two strategies: SS stands for subsampling technique and FS denotes using a full sampling approach.

4.5.4.2. Game solution. Here we solve the proposed two player zero-sum game and how Nash equilibrium can be obtained. Now $p_a \leq 1$ (because $p_b, p_u \leq 1$), and both $0 \leq P_1, P_2 \leq 1$. As $K_m > K_s$, for the defender there exists a strictly dominating strategy but the attacker does not have any strictly dominating strategy. So the defender should always go for sub-sampling method to minimize his loss/attacker's gain. Thus, depending on the values of K_m, K_s, P_1, P_2, a Nash equilibrium can be obtained.

Table 4.1. Complete Information Game.

		Adversary	
		$N - K_m$	$N - K_s$
MCS	FS	$-K_m, K_m$	$-K_s, K_s$
	SS	$-(K_m - P_1N), K_m - P_1N$	$-(K_s - P_2N), K_s - P_2N$

Let us illustrate this with an example. Consider $p_b = 0.8$ with an uncertainty probability, $p_u = 0.15$. As depicted in the Figure 4.4, the $N - K_m$ and $N - K_s$ for the adversary is 40 and 55 respectively calculated from the equation of G. After calculating the payoffs as described in Table. 4.1, we get the payoff matrix as shown in Table. 4.2.

Table 4.2. Payoff Matrix for $p_b = 0.8$.

		A	
		Strategy 1	Strategy 2
MCS	FS	-55, 55	-45, 45
	SS	-19, 19	-29, 29

Clearly, the attacker does not have a strictly dominating attack strategy but the defender has sub-sampling method as a dominating strategy. So finally attacker will settle for strategy 2 as his payoff from strategy 2 is higher than that of strategy 1 and defender will choose the sub-sampling method as the defense mechanism to minimize attackers gain and that will be our Nash equilibrium.

4.5.4.3. QoI scores under optimal sample size. For beta distribution the QoI is calculated based on the following equation by dividing the uncertain ratings into positive and negative rating based on the ratio of positive and negative ratings in the population.

$$QoI^{beta}_{proposed} = \frac{\eta_\alpha(n^{beta}_{opt}) + \left(\frac{\eta_\alpha(n_{opt})}{\eta_\alpha(n_{opt})+\eta_\beta(n_{opt})}\right).\eta_\mu(n_{opt})}{\eta_{opt}+2} \tag{4.13}$$

4.6. SIMULATION RESULTS

In this section, we discuss the simulation, followed by numerical and simulation results.

4.6.1. Simulation Settings. We consider a vehicular crowd-sensing system as a proof of concept by using SUMO (Simulation for Urban Mobility [73] as a simulation environment. We extracted the Open Street Map (OSM) for a part of Manhattan and created individual vehicle trips with a minimum trip length. We introduced accidents/traffic congestion by forcing vehicles to stop at a certain location and time. Every event has an impact area and all the vehicle information in the impact area is collected in that epoch. We consider these users as potential raters who are liable to rate. As per the attack strategy, these users are divided into honest and compromised groups and bad mouthing is performed accordingly.

We considered different active rating population sizes but kept $N = 100$ since our approach is relevant for sparse samples. For certain results for each N, we have considered the compromised rater percentages to be between 30% to 70% for understanding the effect of varying attack scale.

The probability of rating accurately by an honest rater, p_b is varied between 0.7 to 0.95, while the probability of getting an uncertain rating submitted by an honest user, p_u, is varied from 0.25 to 0.

4.6.2. Implementation and Metrics of Performance. The performance is evaluated over 1000 iterations using the data collected from the SUMO environment. In each 1000 iteration, different samples are picked to provide an average case performance that reduces bias. We show results not only for the game (which contains a strategic N-K) but also all possible N-K values to give a sense of how our method will perform under adversaries that may be non-rational or non-strategic.

Finally, for an optimal subsample size for the optimal attack strategy, raters of that size are randomly selected from the population over multiple events. In each event iteration, we calculate the Beta trust score using Equation 4.13. We log the number of iterations where the QoI score under our proposed method was above 0.5 (the typical decision boundary in QoI scoring), *which counts as an event that our method successfully evaded an bad mouthing attack.* This is repeated under all possible numbers of compromised raters (N-K).

We use the following two metrics for the performance evaluation of our proposed randomized subsampling method.

- Probability of Evasion: is the probability that the QoI score obtained by using the beta trust QoI under our proposed method (MTD aware approach) gives a score of 0.5 or more (calculated over 1000 iterations in our simulation), to give the probability of evading a bad mouthing attack under our proposed approach.
- Boost in QoI: This is the raw boost in QoI score when our method is used under bad mouthing attacks, under all possible values of $N - k$, p_b, p_u.

4.6.3. Optimal Sample Sizes. Figure 4.5(a) shows a comparison between theoretical (from Equation 4.11) and experimental values of probability of success ($P(n)$) when total raters $N = 100$ and honest raters $K = 60$, with $p_b = 0.75$ and $p_u = 0.05$ and simulation result. We can conclude from Figure 4.5(a) that simulation result closely follows the numerical result, the optimal sample sizes n_{opt} from theoretical and experimental result are similar.

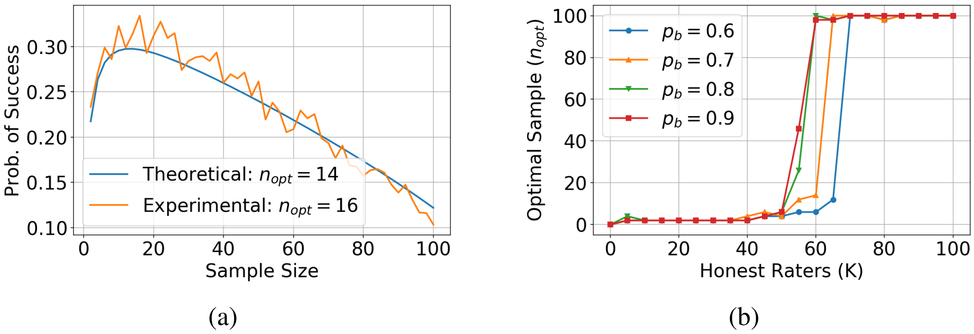


Figure 4.5. Theoretical and Simulation Results. (a) optimal Sample for $p_a = 0.7$; (b) Effect of p_b over optimal sample size.

The change in the resulting optimal sample sizes for various K value (which indirectly depends on N-K by the adversary) over different p_b values is shown in Figure 4.5(b) for N = 100 and p_e = 0.1. It is evident that optimal sample size changes with different K and p_b. As $K.p_b$ increasingly dominates the population, the optimal sample sizes tend to increase and eventually reaches N.

4.6.4. Performance Evaluation. We divide performance evaluation into two parts: (i) Illustrative Performance (ii) Average Case Performance. The illustrative result is for a specific parameter setting while average case performance evaluation is result averaged over all possible combinations of parameters involved.

4.6.4.1. Illustrative performance. We show theoretical versus experimental results of performance as well as comparison of our method versus traditional QoI with Beta distribution for setting N = 100, p_b = 0.7 and p_u = 0.25 for all N-K and also the performance at equilibrium strategies.

- Theoretical Versus Experimental Performance: The comparison of the probability of evading a bad mouthing attack successfully between theoretical and experimental results is shown in Figure 4.6(a), which verifies the accuracy of the model.

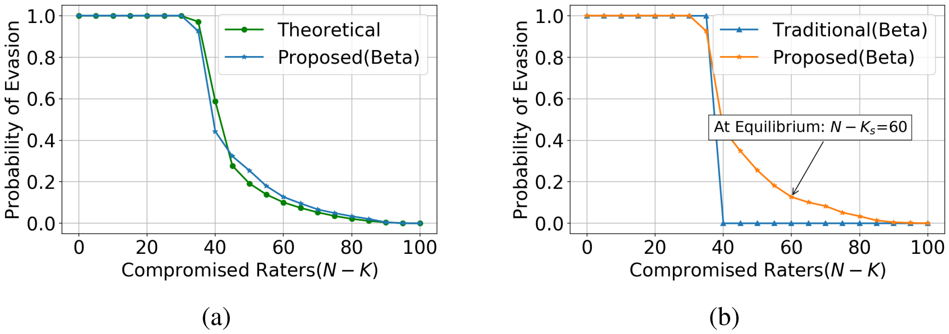
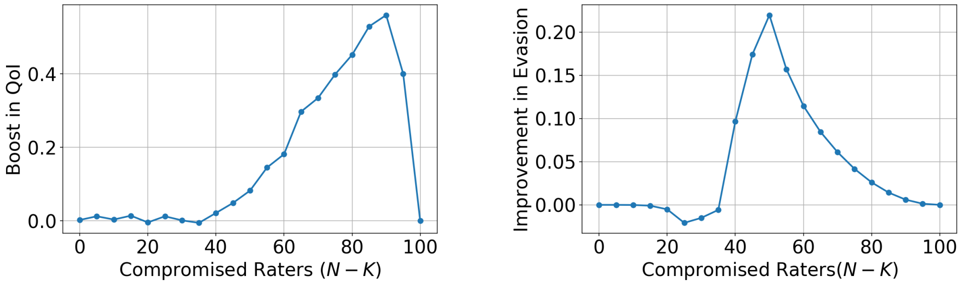


Figure 4.6. Probability of Evasion (p_b = 0.7 and p_u = 0.25). (a) Theoretical vs Experimental; (b) Comparison between Proposed and Traditional QoI scoring.

- Improvement from Traditional Beta Trust QoI: An illustrative result is shown in Figure 4.6(b) the benefit of our method as opposed to traditional QoI method in terms of probability of evasion. The percentage chances of getting QoI above 0.5 under the traditional beta trust QoI which does not implement our MTD is compared to the same under proposed beta trust with the MTD approach. We observe that our proposed method has either equal or a better chance of evasion of bad mouthing attack regardless of the $N-K$ inflicted.

Note that where the performance between traditional and proposed is equal, those are the $N-K$ which are not part of the rational strategy space of the adversary. From the game solution, we have the Nash equilibrium where adversary compromises $N-K_s = 60$ as marked in the plot and in that equilibrium, probability of evasion by using the proposed model is better than the traditional model.

4.6.4.2. Average case performance. We provide an average case performance improvement in terms of probability of evasion and boost in QoI, by averaging them over various p_b, p_u values for each $N-K$. Similarly, the average boost in QoI for using the proposed beta model is shown in Figure 4.7(a). The overall improvement in probability of evasion of the proposed model over traditional model in depicted in Figure 4.7(b). Please



(a) (b)

Figure 4.7. Improvement over Traditional QoI. (a) Boost in QoI; (b) Improvement in Probability of Evasion.

note that low boost or low improvement in the probability of evasion is seen for low $N - K$ because, the MCS gets an advantage regardless of whether MTD is used or not, and is not an indication of limiting performance of our method.

4.7. INFERENCES

Here, we have analyzed the security of a CPS system where sensing devices are mobile and the presence of malicious users are inevitable. We have used MCS as an example of such system and presented a randomized sub-sampling method to improve resilience against bad mouthing attacks even when a large fraction of the rater population (especially during cold start phase) is compromised by a strategic adversary. We showed that there exists an optimal sample size that produces an increase in QoI for each potential value of total compromised raters. Finally, we modeled the problem as a two player zero sum game to conclude that there exists a pure strategy Nash equilibrium. We also showed improvement in terms of evading the effect of bad mouthing attack and the boost in QoI of truthful events under such attacks.

5. INFLUENCE SPREAD CONTROL IN COMPLEX NETWORKS VIA REMOVAL OF FEED FORWARD LOOPS

In the previous section we investigated feedback weaponizing attacks against MCS systems which is a centralized, energy-efficient and robust data collection framework. The data from multiple users with smart devices (smartphones, tablets, wearables etc) are aggregated by the MCS platforms to build a body of knowledge to support decision making. However, due to centralized nature of the MCS system, executing MCS tasks incur cost from the users in terms of energy spent from device batteries and/or data subscription plan (if cellular connectivity is used for information transmission). Eventually users might stop participating in the data collection process and thus making the data acquisition and the decision made by the MCS server unreliable for smart city applications. Moreover, using a single-point data aggregator is always risky as it entails an inherent centrality and is heavily dependent on the functioning of central data aggregator. Thus, to prevent such single-point of network failure and to handle the large user base and data traffic volume, different distributed data collection mechanism are proposed that leverages peer devices to directly communicate with each other in the the MCS platform. One such mechanism, *bioMCS2.0* proposed in [58] attempts to utilize abundance of subgraphs (called motifs) in the MCS network to perform efficient energy-awareness participatory sensing and forwarding in a dynamic scenario where devices are both energy constrained and mobile. The proposed mechanism leverages the topological properties of 3-node motif, called Feed Forward Loop (FFL), to create robust pathways for information transmission in the network. These type of FFLs are found in abundances in social network, biological network called transcriptional regulatory network. [56, 74] also explored the possibility and advantages of FFLs in designing robust and efficient topology that enables energy efficient and sustainable data collection.

Here we investigate how the FFLs are contributing to the information spread in a large scale complex network and we propose to quantify their influence/importance (*motif score*) in the data transferring process. This analysis also gives us opportunity to identify the most influential motifs that if removed can heavily disrupt the whole data collection process. On the other hand, these influential motifs if identified can be removed by the MCS server to curb the malicious information spread at an very early stage.

Selective removal of FLLs (or 3-node motifs) based on the spread function value is one of the most powerful approaches to curb the overall influence spread in any complex network. In this paper, we first prove that any general spread function preserves both monotonicity and submodularity properties even under motif removal operations. Next, we propose a scoring mechanism as a novel spread function that quantifies the relative importance of a given motif within the overall influence spread dynamics on the complex network. We design a novel algorithm that eliminates motifs with high spread scores to curb influence spread. We evaluate the performance of our proposed spread control algorithm using simulation experiments in the context of 3-node motifs (or FFLs) in both real and synthetic network topologies. We demonstrate that high-scoring motifs intercept a high number of short paths from the pre-assigned source and sinks, because of which their elimination results in a significant effect on curbing the influence spread. Furthermore, we empirically evaluate the run-time and cost versus performance trade-off of the proposed algorithm.

5.1. INFLUENCE SPREAD CONTROL

Influence spread control in complex networks has gained attention in recent years in a broad range of applications, from healthcare and medicine (e.g., drug design [17] or curbing epidemics [18]) to social networks [19] (e.g. moderating fake information that can lead to polarization). However, the design of an effective control mechanism to curb influence spread poses important challenges: (i) high computational complexity, especially in large complex networks, (ii) ethical issues, such as harmful bias and lack of accountability in social networks, and (iii) long-term control impact, e.g., potential chronic side-effects due to new drugs, or increasing socio-economic gaps in social networks, which can only be observed in hindsight.

Here, we focus only on the first challenge, namely, designing an efficient information spread control scheme to curb the adversarial influence on the network. In particular, we propose a novel scoring mechanism that leverages the presence of *motifs* (or subgraphs) that are recurrent patterns occurring in complex networks in higher numbers than in randomized networks [22]. The presence of motifs have been reported in many applications including social networks, biological networks, ecological networks, and communication networks [20, 21]. The high degree of evolutionary conservation of motifs suggests that they play a key role in information dissemination, making their detection and analysis the first step towards investigating their contribution to spread dynamics within complex networks.

5.2. CONTRIBUTIONS

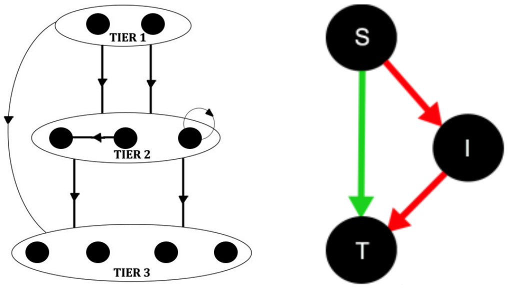
Given a complex network with a pre-determined set of seed (source) nodes spreading information, we propose a novel scoring scheme to help identify motifs that when removed, can minimize the influence spread. We investigate the properties of influence spread function and show that they preserve both monotonicity and submodularity with respect to motif elimination. We also present a novel motif scoring scheme leveraging the principle of network core, a concept from network sciences literature, to quantify the influence of motif removal towards curbing the influence spread. The scoring is based on the number as well as the path lengths between the influential source and sink nodes intercepted by a motif. Our proposed scoring scheme is highly generalizable to evaluate the relative importance of larger motif substructures.

We carry out extensive simulation experiments on a well-studied 3-node motif, called *feed forward loop* (FFL), to demonstrate the efficacy of the proposed motif scoring approach. Specifically, we consider two graph topologies, such as Erdos-Renyi (E-R) random graphs [75] and *E. coli* transcriptional regulatory networks [76], to acquire the motif scores. The performance of our scoring scheme is compared against a near-optimal influence diffusion approach [77]. Results demonstrate that high-scoring FFL motifs intercept a large number of paths connecting the pre-assigned source and sink nodes, and their removal adversely affects network connectivity. The motif scores align with the motif order of the influence diffusion approach. Moreover, the high-scoring motifs have interesting functional properties in *E. coli* transcriptional networks, such as stress response.

5.3. PRELIMINARIES

5.3.1. Graph Representation of Networks. A complex network is represented as a graph $G(V, E)$, where V is the set of vertices (nodes) and E is the set of edges between vertex-pairs. A directed graph has directed edges $(u, v) \in E$ allowing unidirectional information flow from vertex u to v.

Degree and path in directed graph: In a directed graph G, the number of edges leaving a node u is termed its *out-degree*, denoted by $deg^+(u)$, and the number of edges entering a node is its *in-degree*, denoted by $deg^-(u)$. A *directed path* $p = \{u_j, u_{j+1}, \cdots, u_n\}$ is a sequence of vertices such that there is a directed edge from a vertex to its successor in the sequence, i.e., $(u_i, u_{i+1}) \in E$ for $j \leq i < n$. A directed path is considered *simple* if it has no repeated node in the path. In an unweighted graph, the length of a simple path is defined as the number of edges on it. If G is a weighted graph, the path length is equal to the sum of edge weights on the path.



(a) (b)

Figure 5.1. Three-Tier Topology and Feed Forward Loop. (a) Three-tier topology: directed edges indicate potential edges across and within tiers; (b) Feed forward loop (FFL): S is the master regulator, I the intermediate regulator and T the regulated node. The direct edge between S and T is marked in green and the indirect path (via I) in red.

Network efficiency: It is a measure of the average shortest path length between all pairs of source and sink nodes. For a directed graph $G(V, E)$, the network efficiency is defined as

$$\mathcal{E} = \frac{1}{|V| \times (|V| - 1)} \sum_{u,v \in V, u \neq v} \frac{1}{d(u, v)}, \tag{5.1}$$

where $d(u, v)$ is the shortest path length from node u to v.

Graph density: For a directed graph $G(V, E)$, the density is defined as

$$D = \frac{|E|}{|V| \times (|V| - 1)}. \tag{5.2}$$

5.3.2. Transcriptional Regulatory Networks. Transcriptional Regulatory Networks (TRNs) are biological networks that exhibit interactions between proteins, called *transcription factors* (TFs), and genes. They are represented as directed graphs in which nodes represent TFs (or genes) and edges correspond to the regulations of target genes by TFs [76]. Validated and nearly complete TRNs of *E. coli* (*Escherichia coli*) and *Yeast* (*Saccharomyces cerevisiae*) are extracted from GeneNetWeaver [78]. *E. coli* TRN with 1,565 nodes and 3,758 edges and Yeast TRN with 4,441 nodes and 12,873 edges have respectively low graph density of $D = 0.0015, 0.00065$. Let us discuss two key aspects of TRNs.

5.3.2.1. Three-tier topology. Nodes in TRNs are classified into three tiers based on their in-degree and out-degree distributions [55, 56], as discussed below:

- *Tier-1* consists of the set of nodes with zero in-degree, i.e., $\{u \in V : deg^{-}(u) = 0\}$.
- *Tier-2* consists of the set of nodes with non-zero in-degree and out-degree, i.e., $\{u \in V : deg^{+}(u) > 0 \ \ and \ \ deg^{-}(u) > 0\}$.
- *Tier-3* comprises the set of nodes with zero out-degree i.e., $\{u \in V : deg^{+}(u) = 0\}$.

Figure 5.1(a) depicts a three-tier topology with unidirectional data flow from Tier-1 to Tier-3 of the TRN. Tier-1 and Tier-2 comprise TFs and Tier-3 are the target genes.

5.3.2.2. Feed forward loop. Complex networks such as biological networks (e.g., TRNs) and social networks are characterized by the recurrence of subgraphs, called *motifs* [79]. One such 3-node motif is the *feed forward loop* (FFL) that has a direct link between the source node S and the target node T, and an indirect path via I [80] as shown in Figure 5.1(b). S is called the *general TF*, I the *specific TF* and T the *effector operon*. Abundance of FFL results in a few graph properties.

- *Robustness due to independent paths:* Any two paths between a node-pair are called *independent* if they contain no common nodes, except the source and destination. According to Menger's theorem on vertex connectivity [81], *the minimum number of vertices whose removal disconnects two vertices is equal to the maximum number of pairwise vertex-independent paths between them.* Our prior work [56] has shown that FFL offers two independent paths between S and T, and cascades of FFLs allow them to create multiple alternate pathways that are resilient to node or link (edge) failures.
- *Increase in shortest path length:* Knocking off nodes or links in a graph may increase path lengths from u to v or disconnect them. Figure 5.1(b) shows that the indirect path between S and T via I is only a single hop longer than the direct link (S, T). Abundance of FFLs can help reduce the communication delay due to link failures.

- *FFLs as building blocks of larger motifs:* FFL motifs are building blocks of larger motifs of $4, 5, 6$ nodes [82]. This implies that understanding the dynamics of information spread due to FFLs can be a key step towards characterizing spread dynamics in large complex networks.

5.3.3. Influence Diffusion. The main goal of influence diffusion is to identify k seed nodes, which when activated, causes maximum collective information spread within a network. In [77] is presented a greedy strategy that chooses k nodes one at a time, each yielding the highest marginal spread. In a graph G at time t, the calculation of marginal spread unfolds in discrete steps, where each node u is given a single chance to activate a neighbor v with probability given by its edge weight $\zeta((u, v))$. If u is successful, v is activated at time $t + 1$ and acts as a spreader in the subsequent time instants. This process continues until no more activation is possible.

5.3.4. Network Core. Network core of any undirected graph is the residual subgraph that manifests when all nodes with degree less than l are removed. Consequently, all nodes in the l-core have degree at least equal to l; and the $(l+1)$-th core is always a subgraph of the l-th core [83]. Although the notion of network cores was originally conceived for undirected graphs, we propose to employ the same notion to a directed graph (see Sec. 5.5.1) in the proposed motif scoring approach.

5.3.5. Pearson Correlation Coefficient. The Pearson correlation coefficient measures the strength of linear association between two vectors, where the values 1 and -1 are perfect positive and negative correlations, respectively. We employ this coefficient to gauge how similar the proposed motif scoring is to the influence diffusion approach.

5.4. INFLUENCE SPREAD MODEL

Consider a directed, complex network $G(V, E)$, where V denotes the set of individuals; and the weight $\zeta((u, v))$ on directed edge (u, v) represents the influence of individual u over another individual v about a specific issue. Assume that a set of seed nodes (potentially adversarial) denoted by S feeds fake information (or malware) into the network to influence the remaining nodes in a desired manner. To curb such influence spread across the complex network, we aim to identify K FFL motifs responsible for actively spreading the information.

5.4.1. Spread Function. Given an initial seed (source) node set S, we define the spread function $\sigma_S(G)$ on the directed network G as a measure that quantifies average information spread in terms of the number of nodes activated by means of influence diffusion (see Section 5.3.3). Thus, we aim to identify a set of K FFL motifs $\phi = \{(u_1, v_1, w_1), (u_2, v_2, w_2), \ldots, (u_K, v_K, w_K)\}$ such that

$$_{\phi:|\phi|=K} \quad \sigma_S(G(V(G) - V(H), E(G) - E(H))) \tag{5.3}$$

where H is a directed graph representing the FFL motif such that $V(H) = \{u : u \in V(G), u \in \phi_i \ \forall\ i \leq K\}$ and $E(H) = \{(u, v), (v, w), (u, w) : (u, v, w) \in \phi_i, \forall i \leq K\}$. The rest of the paper makes two assumptions: (1) the set of triplet nodes (u, v, w) constituting an FFL motif is denoted as m; and (2) all non-source nodes are sink nodes.

5.4.2. Properties of Spread Functions. In this subsection, we discuss two fundamental properties of spread functions, namely *monotonicity* and *submodularity*, that are preserved under motif elimination in the context of influence spread control in complex networks.

Definition 1 (Monotonicity [84, 85]). *A function f is called monotonic if an element e removed from a set U cannot increase f, i.e.,*

$$f(U - \{e\}) \ \leq \ f(U)$$

for all elements e and set U.

In the following lemma, we show that any monotonic spread function maintains its monotonicity property even with respect to motif elimination.

Lemma 1. *Given a monotonic spread function σ_S, a directed graph H and a seed set S, for any two motif sets A and B such that $B \subseteq A$, we have*

$$\sigma_S(H - A) \leq \sigma_S(H - B). \tag{5.4}$$

Proof. Given that $B \subseteq A$, consider a motif set C such that $A = B \cup C$ and $B \cap C = \emptyset$. Since B and C are partitions of A, we have $H - A = H - B - C$. Then, $\sigma_S(H - A) = \sigma_S(H - B - C)$.

We use induction principles in the remainder of the proof. In the base case, consider the set $H - B$, and remove any one edge $e_1 \in C$ from $H - B$. Since σ_S is monotonic, we have

$$\sigma_S(H - B - \{e_1\}) \leq \sigma_S(H - B).$$

After k intermediate iterations, let $\tilde{C}_k$ be the subset of edges in C that are removed from $H - B$. Assuming that the following inequalities hold true for the k^{th} iteration:

$$\sigma_S(H - B - \tilde{C}_k) \leq \sigma_S(H - B - \tilde{C}_{k-1}) \leq \cdots \leq \sigma_S(H - B),$$

we now focus on the $(k+1)^{th}$ intermediate iteration where we remove $e_{k+1} \in C - \tilde{C}_k$ from $H - B - \tilde{C}_k$. Then, for any monotonic spread function σ_S, we have

$$\begin{aligned} \sigma_S(H - B - \tilde{C}_{k+1}) &= \sigma_S(H - B - \tilde{C}_k - \{e_{k+1}\}) \\ &\leq \sigma_S(H - B - \tilde{C}_k) \\ &\leq \cdots \leq \sigma_S(H - B). \end{aligned}$$

Therefore, by the principle of induction, when the entire set C is removed from $H - B$ (i.e., $C - \tilde{C}_K = \emptyset$ for some $K \in \mathbb{Z}$), the following inequality holds true:

$$\sigma_S(H - A) = \sigma_S(H - B - C) \leq \sigma_S(H - B).$$

□

Definition 2 (Submodularity [84, 85])**.** *A function f is called submodular if it satisfies a natural "diminishing returns" property: the marginal gain from adding an element to a set U is at least as high as the marginal gain from adding the same element to a superset of U. Formally, a submodular function satisfies:*

$$f(U \cup \{v\}) - f(U) \geq f(W \cup \{v\}) - f(W) \tag{5.5}$$

for all elements v and all pairs of sets $U \subseteq W$.

Spread functions satisfy submodularity given an initial set of active nodes [86], i.e., for all H' and H'' where $H' \subseteq H''$

$$\sigma_S(H') - \sigma_S(H' - \{e_i\}) \geq \sigma_S(H'') - \sigma_S(H'' - \{e_i\}) \tag{5.6}$$

The next lemma proves that the spread function satisfies submodularity during motif elimination.

Lemma 2. *Given a submodular spread function σ_S, and any set of edges $\{e_1, \cdots, e_k\}$ that are present in both H' and H'' where $H' \subseteq H''$, we have*

$$\begin{split}\sigma_S(H'' - \{e_1, e_2, .., e_k\}) - \sigma_S(H'') \geq \sigma_S(H' - \{e_1,\\ e_2, .., e_k\}) - \sigma_S(H')\end{split} \tag{5.7}$$

Proof. Let $\{e_1, e_2, .., e_k\} \in E$ be the set of edges to be deleted that are the common to both H' and H'' where $H' \subseteq H''$. Following Inequality (5.6), we have

$$\sigma_S(H') - \sigma_S(H' - \{e_1\}) \geq \sigma_S(H'') - \sigma_S(H'' - \{e_1\}) \tag{5.8}$$

$$or, \sigma_S(H'' - \{e_1\}) - \sigma_S(H'') \geq \sigma_S(H' - \{e_1\}) - \sigma_S(H') \tag{5.9}$$

Without loss of generality, we can extrapolate the submodularity property to the graph $H'' - \{e_1\}$ and obtain

$$\begin{aligned}\sigma_S(H'' - \{e_1, e_2\}) - \sigma_S(H'' - \{e_1\}) \geq \\ \sigma_S(H' - \{e_1, e_2\}) - \sigma_S(H' - \{e_1\})\end{aligned} \tag{5.10}$$

Similarly,

$$\begin{aligned}\sigma_S(H'' - \{e_1, e_2, e_3\}) - \sigma_S(H'' - \{e_1, e_2\}) \\ \geq \sigma_S(H' - \{e_1, e_2, e_3\}) - \sigma_S(H' - \{e_1, e_2\})\end{aligned} \tag{5.11}$$

This can be done for all the edges to be removed. Thus,

$$\begin{aligned}\sigma_S(H'' - \{e_1, e_2, .., e_k\}) - \sigma_S(H'' - \{e_1, e_2, .., e_{k-1}\}) \geq \\ \sigma_S(H' - \{e_1, e_2, .., e_k\}) - \sigma_S(H' - \{e_1, e_2, .., e_{k-1}\})\end{aligned} \tag{5.12}$$

Applying Expressions (5.9), (5.10), (5.11), and (5.12), we obtain the desired result

$$\begin{aligned}\sigma_S(H'' - \{e_1, e_2, .., e_k\}) - \sigma_S(H'') \geq \sigma_S(H' - \\ \{e_1, e_2, .., e_k\}) - \sigma_S(H')\end{aligned} \tag{5.13}$$

□

5.4.3. Greedy Algorithm Performance. Now, if a function is monotonic, submodular and non-negative, then the greedy algorithm, which starts with an empty set and at every step picks an element that maximizes the marginal benefit, provides a set achieving a good approximation, $(1 - \frac{1}{e})$ (where e represents the exponential constant), of the optimal solution [77, 87], where e is the base of the natural logarithm. Therefore, the greedy motif deletion based influence diffusion algorithm proposed in the next section removes the near-optimal motifs curbing the influence spread.

Proof. Let $M^* = \{m_1^*, m_2^*, .., m_k^*\}$ be the optimal motif set and $M = \{m_1, m_2, .., m_k\}$ the motif set generated by the greedy approach; and let Δ be the marginal gain by removing motifs from graph H. Since σ_S is monotonic and submodular, we have the following:

$$\begin{aligned}
\sigma_S(H - M^*) &\geq \sigma_S(H - \{M^* \cup M\}) \\
&= \sigma_S(H - M) \\
&\quad - \sum_{j=1}^{k} \Delta(m_j^* | M \cup \{m_1^*, m_2^*, .., m_{j-1}^*\}) \\
&\leq \sigma_S(H - M) - \sum_{m^* \in M^*} \Delta(m^* | M) \\
&\leq \sigma_S(H - M) - \sum_{i=1}^{k} \Delta(m_{i+1} | M) \\
&\leq \sigma_S(H - M) - k\Delta(m_{i+1} | M),
\end{aligned} \tag{5.14}$$

or,

$$\Delta(m_{i+1} | M) \geq \frac{1}{k}(\sigma_S(H - M^*) - \sigma_S(H - M)). \tag{5.15}$$

Let $\delta_i = \sigma_S(H - M^*) - \sigma_S(H - M)$, then

$$\begin{aligned}
\delta_i - \delta_{i+1} &= \sigma_S(H - M \cup \{m_{i+1}\}) - \sigma_S(H - M) \\
&= \Delta(m_{i+1} | M) \geq \frac{1}{k}\delta_i \quad \text{from Inequality (15),}
\end{aligned} \tag{5.16}$$

or equivalently,

$$\delta_{i+1} \leq (1 - 1/k)\delta_i, \quad \text{and} \quad \delta_k \leq (1 - 1/k)^k \delta_0. \tag{5.17}$$

From well-known bound $1 - x \leq e^{-x}$ for $x \in \mathbb{R}$, we have

$$\sigma_S(H - M) = \left(1 - \frac{1}{e}\right) \cdot \sigma_S(H - M^*) \tag{5.18}$$

□

Thus, the greedy motif deletion is $(1 - \frac{1}{e}) \approx 63\%$ approximation of spread under optimal motif elimination.

5.5. PROPOSED MOTIF SCORING FRAMEWORK

We discuss the proposed motif scoring mechanism and the baseline greedy algorithm for influence spread.

5.5.1. Motif Scoring. Given a directed graph G, adjacency matrix $\mathbf{A}$ and $k \in \mathbb{Z}$, we define $\mathbf{A}^k_{u,v}$ as the $[u, v]^{th}$ entry of the k^{th} power of adjacency matrix A that represents the number of walks from node u to v in k hops. The score reflects how many paths between the seed and other nodes are intercepted by a given FFL m. Figure 5.2 shows that the motif scoring works in two stages: connectivity from (1) a set of preassigned seeds S to m and (2) m to the preassigned sink nodes (or all nodes not in the source set S).

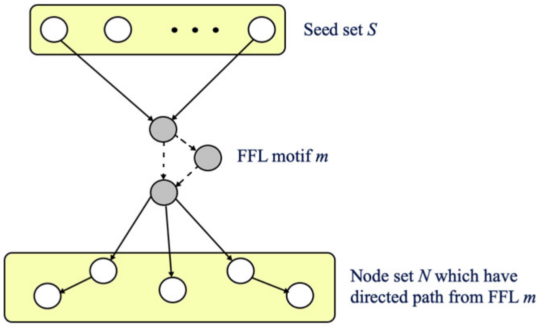


Figure 5.2. Two Stage Motif Scoring Mechanism.

In Equation 5.19, we calculate the FFL information potential, $\mathcal{P}(m)$, for motif m as the weighted sum of the score on (1) incoming links from seed nodes to m and (2) outgoing links from m to all the nodes in G. As mentioned earlier, $\mathbf{A}$ is the adjacency matrix corresponding to $G(V, E)$, α is the weighing factor, k is the maximum walk length considered, ψ_u is the individual spread of seed node u (gauged by influence diffusion discussed in Sec. 5.3.3) and p_u is the *K-directed score* (measuring spreading potential as discussed hereafter) of a node u. The motif acquires a higher score if there exist short pathways (1) into m from seeds with high spreading potential and (2) from m to successors with high spread, as shown below.

$$\begin{aligned}\mathcal{P}(m) = \alpha \times \sum_{u \in S} \left(\sum_{v \in m} \left(\sum_{k} \frac{\mathbf{A}^k_{u,v}}{k!} \right) \right) \psi_u + \\ (1-\alpha) \times \sum_{u \in V} \left(\left(\sum_{v \in m} \sum_{k} \frac{\mathbf{A}^k_{v,u}}{k!} \right) \right) \kappa_u \end{aligned} \tag{5.19}$$

It is noteworthy that this scoring technique depends on the number and length of paths between influencial nodes intercepted by a motif. Thus, it is highly generalizable to study the role of larger motifs in network spread.

- K-Directed score: We propose the K-directed score inspired from the notion of network core (mentioned in Sec. 5.3.4), which we adapt for directed networks. Given directed graph $G(V, E)$, we initialize a hash table κ. We iteratively eliminate nodes $u \in V$ with the least out-degree K, from G and assign $\kappa_u = K$ (see Algorithm 5 for details). The underlying intuition is that the nodes with high score lie at the interior of the directed network and have high spreading potential.

5.5.2. Greedy Influence Diffusion under Motif Elimination. Taking a cue from the effect of motif removal on spread function, we apply a greedy motif deletion influence diffusion as a benchmark [77]. For instance, Figure 5.2 shows that if motif m is knocked

Algorithm 5 K-directed Score

```
Input. Directed graph, $G(V, E)$
Output. $\kappa$
$K = \min(deg^{+}(G)), \kappa = \{\}$
while $|V| > 0$ do
    for $u \in V$ do
        if $deg^{+}(u) = K$ then
            $n\_list := u$
    for $u \in n\_list, \ldots$ do
        $\kappa[u] = K$
    $K = K + 1$
    Remove $n\_list$ from $G$
=0
```

off the ties between the seed and non-seed is eliminated, thereby affecting network spread. Thus, given seed nodes S, we employ the greedy baseline strategy to calculate the effect of knocking off an FFL motif on the overall spread.

5.5.2.1. A baseline algorithm. We discuss the greedy stochastic influence spread presented in [77], as discussed in Sec. 5.3.3. In Algorithm 6, given directed graph G and seed set S, we implement the influence diffusion as function $IC(G, S)$ and estimate the greedy score for each FFL motif $F = (u, v, w)$ based on the effect of spread when F is knocked off G. The final score is the mean spread across v iterations after removing nodes u, v, w from G.

Algorithm 6 Greedy Influence Diffusion

```
Input. Directed graph $G$, No. of iterations $v$, motif $F = (u, v, w)$, Seed set $S$
Output. $\frac{1}{v}\sum_{i=1}^{v}\psi_i$
$i = 0, \psi = \emptyset$
Remove $F$ from $G$, i.e.,
    $E(G) = E(G)\backslash\{(u, v), (v, w), (u, w)\}$
for $i < v$ do
    $\psi_i = IC(G, S)$
    Increment $i$
```

5.5.2.2. Validation of motif scoring approach. We first calculate the scores using the greedy baseline, where lower greedy score represents higher influence of a motif in spread. Thus, we compare the motifs ranked in the increasing order of greedy score against the list of motifs ranked in the decreasing order of proposed motif score. The effectiveness of the proposed scheme depends on the Pearson's correlation coefficient (see Sec. 5.3.5) between the two list.

Table 5.1. Description and Value of Simulation Parameters.

Parameter	Symbol	Value
Weighing factor	α	0.5
Shortest path length	p	5
Simple path length	P	10
E-R Graph Density	D	0.05, 0.1

5.5.3. Graph Topologies. We validate the motif scoring strategy with the two topologies as described below.

5.5.3.1. Random topologies. We generate Erdos-Renyi random graphs by connecting nodes randomly with directed links with a preassigned probability [75] (see Sec. 5.6 for details).

5.5.3.2. Sampled subgraphs of TRN. We sample subgraphs from the directed TRN. In Algorithm 7, we input TRN G and the order (i.e., required number of nodes in sampled subgraph) r, the of the required subgraph. We initialize an empty graph H and add a well connected node $u \in V(G)$ as the first node. Subsequently, we iteratively add nodes to H by randomly selecting a new node, say v, that belongs to the neighborhood of $u \in V(H)$ in G and include the directed links between the newly added and existing nodes u and v from G. This is required to ensure single connected component in the generated subgraph. This process terminates when the required subgraph order r is realized.

5.6. RESULTS

We carry out simulation experiments on Python. We employ the Networkx [88] library for the graph operations and validate the motif scoring method against directed Erdos-Renyi random graphs and subgraphs of order ranging from $50 - 300$ nodes sampled from *E. coli* TRN. Unless otherwise stated, we consider $\alpha = 0.5$ in the motif score (Equation 5.19). The default simulation parameters are summarized in Table 5.1.

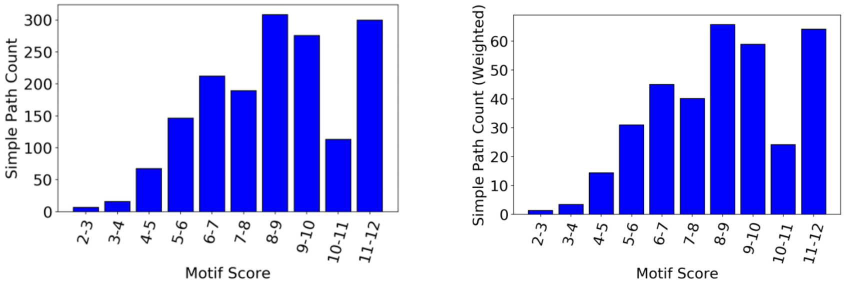
Algorithm 7 Sample - TRN

```
Input. Directed graph $G(V, E)$, $r$
Output. Directed graph $H$
Convert $G$ to undirected graph $G'$
$H = \emptyset$
Select random seed $u \in V$ with probability $\frac{deg(u)}{\sum_{v \in V} deg(v)}$
$V(H) = V(H) \cup \{u\}$
while $|H| < r$ do
    Select random neighbor of $u \in G'$ $(u \notin V(H))$, say $v$, from $G'$
    Add $V(H) = V(H) \cup \{v\}$
    for $u \in V(H)$ do
        if $(u, v) \in E(G)$ then
            Add a new edge $E(H) = E(H) \cup (u, v)$
        if $(v, u) \in E(G)$ then
            Add a new edge $E(H) = E(H) \cup (v, u)$
```

5.6.1. Motif Removal Strategies.

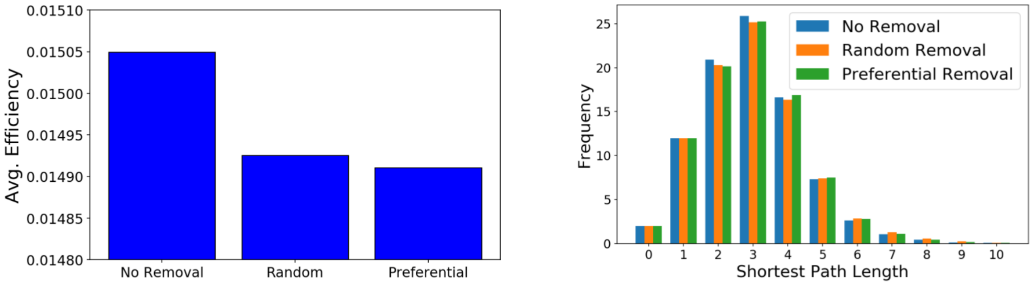
In order to remove a motif $m = (u, v, w)$ of score $\mathcal{P}(m)$, belonging to motif set M, we knock off its directed links $(u, v), (v, w), (u, w)$ from the directed graph. We consider the following two motif removal scenarios.

- *Random removal:* The probability of motif removal is independent of its motif score, i.e., $\dfrac{1}{|M|}$.
- *Preferential removal:* The probability of motif removal is proportional to its motif score, i.e., $\frac{\mathcal{P}(m)}{\sum_{m' \in M} \mathcal{P}(m')}$.



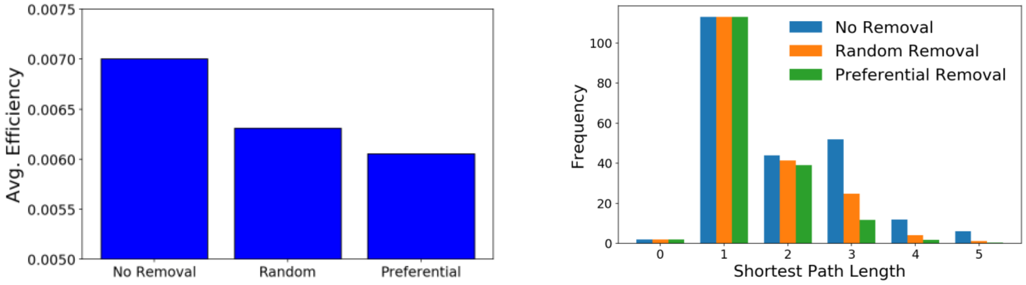
(a) (b)

Figure 5.3. Relationship Between Motif Scores and Source-Sink Path. (a) motif score and simple source-sink path count; (b) motif score and weighted source-sink path count.



(a) (b)

Figure 5.4. Failure scenarios on 150-node E-R random graphs. (a) average network efficiency; (b) frequency distribution of shortest path length.



(a) (b)

Figure 5.5. Failure on 150-node *E. coli* TRN. (a) average network efficiency; (b) frequency distribution of shortest path length.

5.6.2. Effect of Motif Score. We discuss in Sec. 5.5 that the high scoring motifs intercept bulk of the directed paths from the preassigned source to the sink nodes. Specifically, the motif score will be higher if the (1) number of source-sink paths increases and (2) source-sink paths are shorter in length. To demonstrate this, we estimate the relationship between the score $\mathcal{P}$ of a motif (refer Equation 5.19) and count and length of source-sink paths it intercepts. For these experiments, we show the mean measurements of 50 iterations on directed, weighted E-R random graphs of 300 nodes and graph density $D = 0.1$. We use the Networkx path function to enumerate simple paths of length, $P \leq 10$.

In the first experiment, we consider the two nodes with the highest out-degree nodes as source, while the remaining nodes act as sinks. Figure 5.3(a) shows that the mean count of simple paths intercepted by a motif tends to range from 10 for $\mathcal{P} \in [2, 3)$ to 300 for $\mathcal{P} \in [11, 12)$, respectively. This suggests that the number of source-sink simple paths intercepted by a motif m is indeed commensurate with $\mathcal{P}(m)$. In another experiment, we turn our attention to the length aspect of the paths intercepted by the motifs. Instead of assigning equal weight to each simple path, we gauge the importance of a path p by its "shortness", the shortest path length, i.e., $\frac{1}{|p|}$. Figure 5.3(b) once again shows that the path interception of a motif m increases with $\mathcal{P}(m)$, even when we attach more weight to shorter paths.

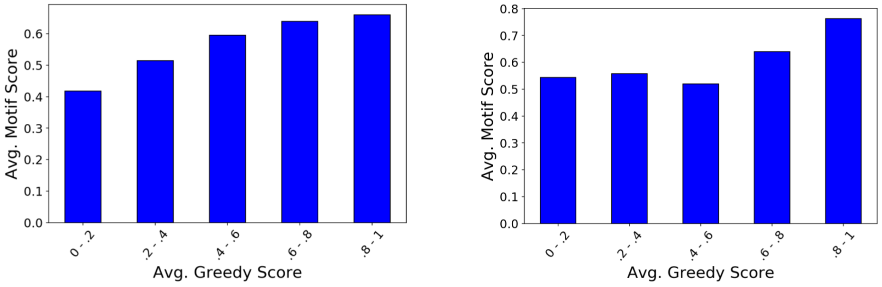
5.6.3. Failure Scenario. Let us now consider the effect of removing high-scoring FFL motifs (by virtue of failure or attack) on the overall connectivity of a directed network, measured in terms of the change in network efficiency (refer Equation 5.1 in Sec. 5.3.1 for details). Recall that we execute motif removal by knocking off its constituent links and that we consider random as well as preferential motif removal strategies. For each strategy, we carry out 10% motif removal on 150-node E-R random graphs of graph density $D = 0.1$. Once again we consider only 2 high out-degree nodes to be source and all other nodes to be sinks.

Figure 5.4(a) depicts that the network efficiency corresponding to the preferential failure is lower than that of random failure. This goes to show that the overall connectivity of the network suffers when motifs with high $\mathcal{P}$ are knocked off. Furthermore, we compare a frequency distribution of the shortest path lengths for the two failure scenarios. Figure 5.4(b) shows that while the frequency of shortest paths of length 2 for preferential failures is less than that of random failures, it exhibits higher frequency for paths of length 3. This suggests that the preferential motif failures creates longer paths, causing poorer connectivity in the directed E-R graphs.

We recreate the same scenario for 150-node subgraph sampled (using Algorithm 7) from *E. coli* TRN. Figure 5.5(a) shows a significant difference in network efficiency between preferential and random node failure scenarios. Unlike the E-R random networks where nodes exhibit average degree, TRN is a sparser topology (refer to graph density of TRN $D < 0.002$ discussed in Sec. 5.3.2) and possesses fewer source-sink pathways, making it more susceptible to failure of motifs with high $\mathcal{P}$. Similarly, we plot the frequency distribution of the path lengths to show that the preferential motif failures cause path lengths to increase or nodes to become unreachable, resulting in fewer paths of length ≤ 5 (see Figure 5.5(b)).

5.6.4. Comparison with Greedy Influence Diffusion. We compare the performance of motif scoring against standard influence diffusion approach (refer Sec. 5.3.3 for details). In each run of this experiment, we apply the greedy influence diffusion baseline approach to acquire the motif scores and classify the motifs into 5 bins of size 0.2 on the basis of scores normalized by the maximum score of that directed graph.

We apply the proposed motif scoring method to record the corresponding normalized score of the motifs in each bin based on the proposed scoring method, over 1000 iterations. Figure 5.6(a) and 5.6(b) shows that for directed 150-node E-R random graph with random edge weights and TRN subgraphs, the FFL scores increase with increasing greedy scores



(a) (b)

Figure 5.6. Pearson Correlation of Motif Scoring and Influence Diffusion. (a) on E-R random graphs; (b) on *E. coli* TRN subgraphs.

(with Pearson correlation coefficients equal to 0.84 and 0.97, respectively). This shows the efficacy of the proposed scoring in capturing the effect of FFL motifs in overall influence spread.

5.6.5. Biological Validation. The well-connected nodes in the TRNs represent transcription factor proteins or genes with well-studied and highly specialized functional properties. We validate the important FFL motifs identified by the scoring Equation 5.19 by recording the functional properties of the genes participating in the motifs in the *E. coli* TRN. To this end, we calculate the top 100 scoring motifs and rank the genes with the highest participation in them. Figure 5.7 depicts the top 20 most appearing genes in *E. coli* TRN by the (a) number of times they appear in the top 20 motifs and (b) *weighted* sum of motif scores the genes appear in. in *E. coli* TRN.

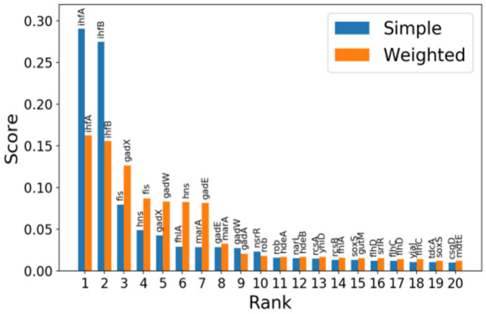


Figure 5.7. Names of Genes Participating in the Top 20 High Scoring FFL Motifs.

It is noteworthy that *ihfA*, *ihfB*, *gadX*, *gadW*, *fis* emerge as the nodes of most significant motifs. The genes ihfA, ihfB play a role in anaerobic fermentative metabolism in *E. coli* [89], while gadX, gadW are major transcriptional regulators that affects acid resistance [90]. Finally, *fis* too is a regulator influencing translation (rRNA and tRNA genes), virulence, biofilm formation, energy metabolism, stress response, central intermediary metabolism, amino acid biosynthesis, transport, cell structure, carbon compound metabolism, amino acid metabolism, etc. [91]. Also, all the above genes appear in tier 2 of the three tier hierarchy of *E. coli* TRN (see Sec. 5.3.2 for details), making them ideal candidates intercepting information flow between tiers 1 and 3. This aligns with our earlier studies [56, 82] that show the tier 2 nodes to contain key TRN motifs. The tiers of the motifs with the top 10 scores (and their tier information) are summarized in Table 5.2, where the general TFs (defined in Sec. 5.3.2.2) belong to tier 2.

Table 5.2. Top 10 Motif Scores.

Gene Motif	Weighted Score	Tier
('gadX', 'gadW', 'gadA')	10.0085	2 2 3
('gadW', 'gadX', 'gadA')	10.0085	2 2 3
('gadE', 'gadX', 'gadA')	9.93092	2 2 3
('gadX', 'gadE', 'gadA')	9.93092	2 2 3
('gadW', 'gadX', 'hdeB')	9.41128	2 2 3
('gadW', 'gadX', 'hdeA')	9.41128	2 2 3
('gadX', 'gadW', 'yhiD')	9.41128	2 2 3
('gadX', 'gadW', 'hdeA')	9.41128	2 2 3
('gadX', 'gadW', 'hdeB')	9.41128	2 2 3
('gadW', 'gadX', 'yhiD')	9.41128	2 2 3

We explore the notable research abstracts, where the genes participating in the top FFL motifs feature together. We find the 1000 research abstracts from PubMed [92] and perform text analysis to identify motif triplets in them. Figure 5.8 depicts that *PMID: 22790954* and *PMID: 17259322* deal with acid resistance [93, 94], whereas *PMID: 9097440* deals with the effect of genes on resistance to antibiotics [95]. Finally, *PMID: 9783171* explores how the over-expression of stress response genes affect tolerance to organic solvents [96]. These papers report essential functions in the transcriptional networks, explaining why the motif genes we report have high $\mathcal{P}(m)$.

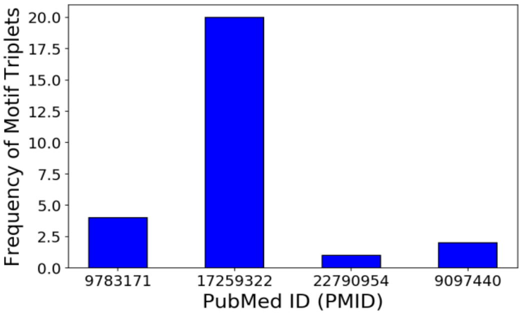
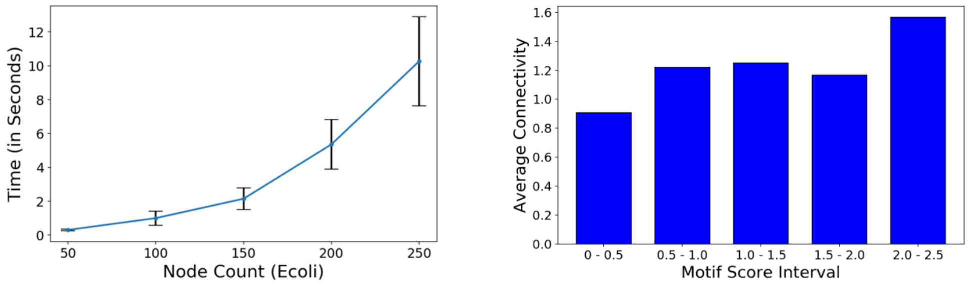


Figure 5.8. Motif Triplets Frequency in Top Four Research Abstracts.

5.7. DISCUSSIONS

In this work, we proposed a generalizable motif scoring mechanism that identifies the role of motifs in the spread of influence from preassigned spreader nodes in a complex networks. We proved the influence spread in networks under motif elimination to be monotonic, submodular and compared the accuracy of the proposed scoring technique against a near-optimal greedy influence diffusion approach. Our simulation experiments using feed forward loop (FFL) motifs in random graph and transcriptional network topologies corroborated the ability of the scoring approach to influence information spread by intercepting short paths between seed and sink nodes. Furthermore, the high-scoring FFL motifs also emerge as significant spreader in the influence diffusion approach and contribute towards important functional properties like stress response in *E. coli* transcriptional networks.

This work presents several interesting research directions. First, we explored the average running time of the motif scoring approach for *E. coli* TRN on a Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz 16 GB RAM computer. Figure 5.9(a) shows the nonlinear increase in running time (measured in seconds) for increasing graph order, suggesting that



(a) (b)

Figure 5.9. Scalability and Attack Cost vs. Damage Trade-off.

the proposed motif scoring approach may be computationally prohibitive for large complex network scenarios. This makes it imperative to design scalable yet efficient motif scoring algorithms. We are working towards a distributed motif scoring mechanism in which a node can identify motifs in its immediate neighborhood in the absence of the knowledge of the global network topology.

Second, in real world complex networks topologies, such as wireless networks and smart grid, an attacker may disrupt communication by knocking off the high scoring motifs. However, high scoring may be well-connected, thereby demanding a greater attack budget. Figure 5.9(b) shows the increasing mean degree of nodes participating in FFL motif of 50 150-node subgraphs sampled from *E. coli* TRN. Evidently, if one equates the cost of knocking off a FFL motif to its connectivity (i.e., node degree in the network), it becomes imperative for the attacker and network administrator to analyze the trade off between the damage inflicted and attack cost incurred. Third, we will validate the efficacy of the motif scoring approach on other real datasets of large-scale social, communication or biological networks. Finally, we have discussed in Sec. 5.5 that the proposed motif scoring approach is highly generalizable and can be employed to gauge the influence of larger motifs.

6. ONGOING RESEARCH

6.1. A SCALABLE MOTIF SCORING MECHANISM TO CONTROL INFORMATION SPREAD IN SOCIAL NETWORKS

6.1.1. Motivation for a Scalable Solution. Influence spread control in complex networks has gained attention in recent years in a broad range of applications, from healthcare and medicine (e.g., drug design [17] or curbing epidemics [18]) to social networks [19] (e.g. moderating fake information that can lead to polarization). However, the design of an effective control mechanism to curb influence spread poses important challenges: (i) high computational complexity, especially in large complex networks, (ii) ethical issues, such as harmful bias and lack of accountability in social networks, and (iii) long-term control impact, e.g., potential chronic side-effects due to new drugs, or increasing socio-economic gaps in social networks, which can only be observed in hindsight [97, 98, 99, 100].

Feed Forward Loop (FFL) is a $3-$node subgraph that is recurrent in social and biological networks. It is characterized by three nodes S, I and T, such that there is a direct link between S and T as well as an indirect path via I.

6.1.2. Problem Statement. Consider a social network $G = (V, E)$, where V is the set of individuals in the network and E is the set of edges that represent the influence of one individual over another regarding a specific issue. We assume that a set of adversarial seed nodes (denoted as S) feeds negative information into the social network in order to influence the rest of the individuals in a desired manner. In an attempt to alleviate this false information spread across the social network, our goal is to identify K FFL motifs which are responsible for actively spreading misinformation across the network, and eventually remove them from the network.

Consider a social graph (directed, weighted and signed) $G(V, E)$ and seed set S. The average spread within graph G for seed set S is given by $\sigma_S(G)$.

Determine a set of K FFL motifs $\phi = \{(u_1, v_1, w_1), (u_2, v_2, w_2), \ldots, (u_K, v_K, w_K)\}$, s.t.

$$\underset{\phi:|\phi|=K}{\operatorname{argmin}} \sigma_S(G(V(G) - V(H), E - E(H))). \tag{6.1}$$

Here H is a directed signed graph s.t. $V(H) = \{u : u \in V(G), u \in \phi_i \forall i \leq K\}$ and $E(H) = \{e(u,v), e(v,w), e(u,w) : (u,v,w) \in \phi_i, \forall i \leq K\}$.

6.1.3. FFL Scoring Mechanism. In this section, we present a simple scoring mechanism to gauge the significance of any given FFL motif m in the information flow within a social graph $G(V, E)$.

$$\begin{aligned} \mathcal{P}(m) = \alpha \times \sum_{u \in S} \left(\sum_{v \in m} \left(\sum_{k} \frac{\mathbf{A}_{u,v}^{k}}{k!} \right) \right) \psi_u + \\ (1 - \alpha) \times \sum_{u \in V} \left(\left(\sum_{v \in m} \sum_{k} \frac{\mathbf{A}_{v,u}^{k}}{k!} \right) \right) \kappa_u. \end{aligned} \tag{6.2}$$

$$\sigma_S(u_1, \cdots, u_M) = \sigma_S^{resid.} + \sum_{m \in \mathcal{M}} u_m \cdot \mathcal{P}(m), \tag{6.3}$$

where u_m is the decision made regarding the inclusion/exclusion of FFL m, and $\mathcal{P}(m)$ is the scoring function defined in Equation (6.2).

In Equation 6.2, we calculate the FFL information potential $\mathcal{P}$ as the weighted sum of the score on (i) incoming links from seed nodes to FFL motif m and (ii) outgoing links from m to all the nodes in G. Here A is the adjacency matrix corresponding to $G(V, E)$, w is the weighing factor, k is the maximum walk length considered, ψ_u is the individual spread of seed node u and p_u is the pagerank centrality of a node u.

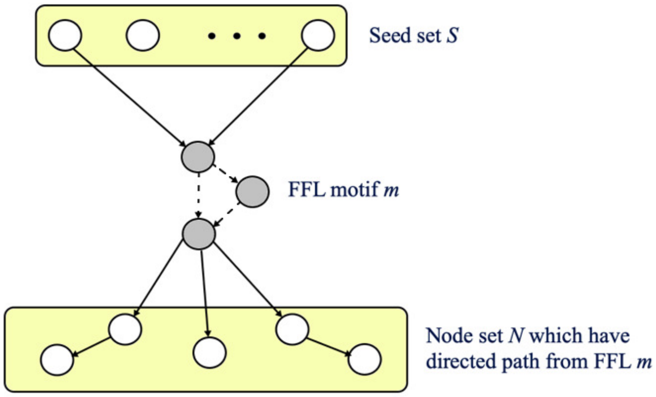


Figure 6.1. Two stage motif scoring.

Figure 6.1 represents a schematic of the two-part scoring mechanism of information potential of FFL motif m. Our metric is based on the fact that $A^k_{u,v}$ is the $[u,v]^{th}$ entry of the k^{th} power of adjacency matrix A that represents the number of walks from node u to v in k hops. The score reflects how many paths between the seed and other nodes is intercepted by a given FFL m.

6.1.4. Motif Misinformation Propagation.

$$w^{\mu}(u,v) = \frac{w(u,v) \times \mathbb{D}_u}{\max_{(u',v') \in E} w(u',v') \times \mathbb{D}_{u'}}. \tag{6.4}$$

$$\mathcal{P}^{\mu}(m) = \mathcal{P}(m) \times \prod_{v \in m} P(v|par(v,m)). \tag{6.5}$$

Here $P(v|u) = w^{\mu}(u,v)$ if $(u,v) \in E$ and 0 otherwise; $par(v,m) = \{u : u \in m, v \in m, (u,v) \in E\}$, and 0 otherwise.

6.1.5. Preliminary Experimental Results. We carry out the simulation experiments on synthetically generated networks as well as social networks. We utilize the Python Networkx [88] library to process the networks. The default simulation parameters are summarized in Table 6.1. We consider the following datasets for our validation purpose.

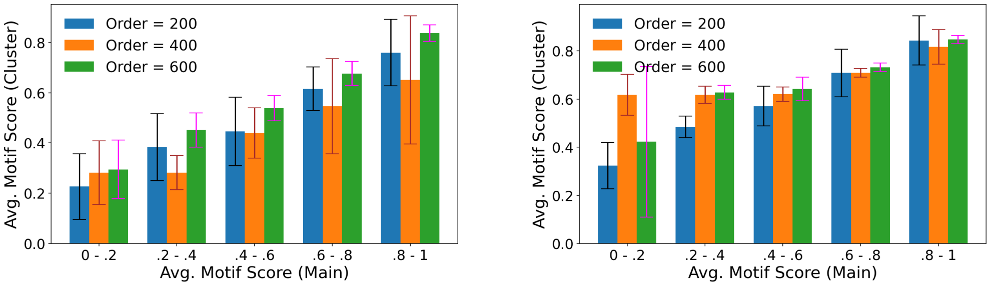
- *Wikipedia administrator requests network.* This is a directed, signed network of $10,835$ nodes and $159,388$ links, where the nodes represent the Wikipedia members and the directed links are ordered pairs of the voter and his choice of Wikipedia administrator candidate [101].
- *Clustered networks.* We use the Python Scikit-Learn library [102] *make_blobs* function to generate 200, 400, and 600 datapoints divided in 5 clusters on a 2D plane (refer Table 6.1).

Table 6.1. Scalability Parameter Description and Value.

Parameter	Symbol	Value
Weighing factor	α	0.5
Shortest path length	p	5
Simple path length	P	10
E-R Graph Density	D	0.05, 0.1
Number of nodes in wiki network	-	10835
Number of edges in wiki network	-	159388
E-R Graph Density	D	0.05, 0.1

Figure 6.2 depicts the average motif scoring using our clustered approach compared to the average motif scores obtained from the whole graph. It is evident that high scoring motifs are also getting high scores when clustered approach is used. For the next experiment we varied the cluster count and monitored the effect on execution time and average motif scores obtained from the cluster and the original graph. The result of this experiment is depicted in Figure 6.3.

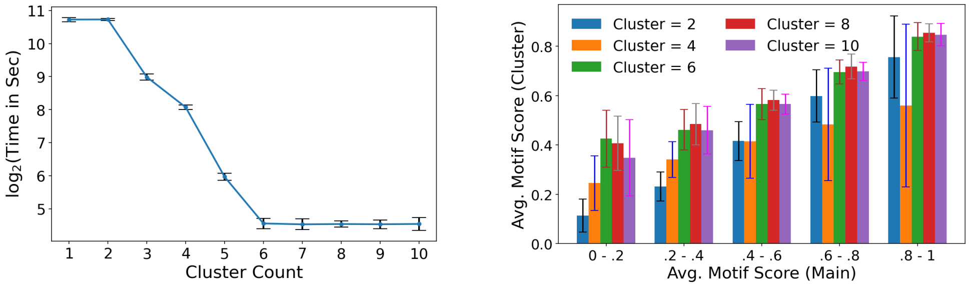
Figure 6.4 show that the running time of the scalable and original motif scoring approach in seconds on the log scale. We find that the gap in running time widens with the number of nodes. The speed-up achieved by the scalable approach on 600-node wiki network (Figure 6.4(a)) and clustered networks (Figure 6.4(b)) are approximately 16 and 64 times, respectively.



(a)

(b)

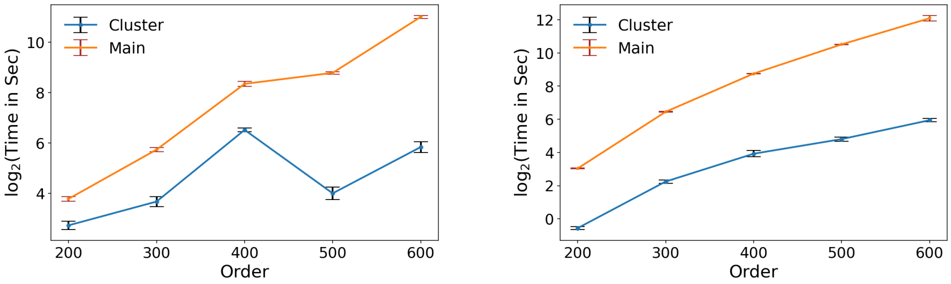
Figure 6.2. Scalability Analysis of Influence Diffusion and Motif Scoring. (a) wiki subgraph; (b) clustered networks.



(a)

(b)

Figure 6.3. Scalability Analysis Based on Cluster Count. (a) over Execution time; (b) over Motif score.



(a)

(b)

Figure 6.4. Run-time analysis of influence diffusion on (a) wiki subgraph; (b) clustered networks.

7. CONCLUSION AND FUTURE DIRECTIONS

In this dissertation, we studied the robustness of smart CPS systems under various transmission setups. We first consider a CPS system where sensing devices are static and directly transmitting to the decision maker in a single hop. We have used a distributed PMU network as a proof-of-concept application to propose a real-time anomaly based attack detection framework that uses *harmonic mean to arithmetic mean ratio* as a stable metric to identify current magnitude falsification in PMU data streams. We showed that even if the attacker has knowledge about the underlying time series we are still able to identify anomalies with a low false alarm rate in real time. Also, unlike many existing bad data detection methodologies, it does not require the topology of the grid network.

Then we studied a generic CPS system where sensing devices (or humans) are mobile and investigated the security and robustness of the system against feedback weaponizing attacks. We showed that under extreme adversarial conditions both linear and nonlinear QoI models can fail and then we proposed a randomized sub-sampling method to improve resilience against bad mouthing attacks even when a large fraction of the rater population (especially during cold start phase) is compromised by a strategic adversary. We showed that there exists an optimal sample size that produces an increase in QoI for each potential value of total compromised raters. Finally, we modeled the problem as a two player zero sum game to conclude that there exists a pure strategy Nash equilibrium. We also showed improvement in terms of evading the effect of bad mouthing attacks and the boost in QoI of truthful events under such attacks.

Finally, we consider a more generic MCS application where distributed data collection mechanisms are used that leverages peer devices to directly communicate with each other in the MCS platform and the data reaches the destination in multiple hops from the source. We investigated the role of feed forward loops (or motifs) in such a distributed framework and proposed a method for the identification of influential motifs that contribute significantly to data collection or information diffusion. We evaluate the performance of our proposed spread control algorithm using simulation experiments in the context of 3-node motifs in both real and synthetic network topologies. We demonstrate that high-scoring motifs intercept a high number of short paths from the preassigned source and sinks, because of which their elimination results in a significant effect on curbing the influence spread.

In the future, we plan to build a generic anomaly-based attack detection model that can serve as a common framework across various applications of the smart grid network. We are also working on building an all-inclusive model to tackle all possible types of feedback weaponizing attacks such as ballot stuffing, bad mouthing, reputation stuffing, etc. to strengthen the CPS system more robust and resilient to adversarial manipulations.

APPENDIX

PUBLICATIONS

1. PEER-REVIEWED CONFERENCE PAPERS

- P. Roy, S. Bhattacharjee, and S.K. Das. "Real Time Stream Mining based Attack Detection in Distribution Level PMUs for Smart Grids" IEEE Global Communications Conference (GLOBECOM) 2020.

- P. Roy, S. Bhattacharjee, and S.K. Das. "Resilience Against Bad Mouthing Attacks in Mobile Crowdsensing Systems via Cyber Deception" IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM) 2021.

- S. Roy, P. Roy, VSS Nadendla, and S.K. Das. "Influence Spread Control in Complex Networks via Removal of Feed Forward Loops" Conference on Computer Communications and Networks (ICCCN) 2021.

2. PAPERS UNDER REVIEW

- VPK Madhavarapu, P. Roy, S. Bhattacharjee, and S.K. Das. "Active Learning Augmented Folded Gaussian Model for Anomaly Detection in Smart Transportation"

REFERENCES

[1] Ragunathan Rajkumar, Insup Lee, Lui Sha, and John Stankovic. Cyber-physical systems: the next computing revolution. In *Design automation conference*, pages 731–736. IEEE, 2010.

[2] Yang Li, Jiachen Sun, Wenguang Huang, and Xiaohua Tian. Detecting anomaly in large-scale network using mobile crowdsourcing. In *IEEE INFOCOM 2019-IEEE Conference on Computer Communications*, pages 2179–2187. IEEE, 2019.

[3] Federico Montori, Prem Prakash Jayaraman, Ali Yavari, Alireza Hassani, and Dimitrios Georgakopoulos. The curse of sensing: Survey of techniques and challenges to cope with sparse and dense data in mobile crowd sensing for internet of things. *Pervasive and Mobile Computing*, 49:111–125, 2018.

[4] Bin Guo, Zhu Wang, Zhiwen Yu, Yu Wang, Neil Y Yen, Runhe Huang, and Xingshe Zhou. Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM computing surveys (CSUR)*, 48(1):1–31, 2015.

[5] Satyaki Roy, Nirnay Ghosh, Preetam Ghosh, and Sajal K Das. biomcs: A bio-inspired collaborative data transfer framework over fog computing platforms in mobile crowdsensing. In *Proceedings of the 21st International Conference on Distributed Computing and Networking*, pages 1–10, 2020.

[6] A Von Meier, UC Berkeley, K Brady, UC Berkeley, M Brown, UC Berkeley, GR Cotter, and I Llc. Synchrophasor monitoring for distribution systems: Technical foundations and applications a white paper by the naspi distribution task team. *NASPI: Berkeley, CA, USA*, 2018.

[7] BREAKOUT SESSION DISCUSSIONS. Metrics for measuring progress toward implementation of the smart grid. 2008.

[8] Eric Lightner and S Director. Evolution and progress of smart grid development at the department of energy. In *DOE presentation at FERC-NARUC Smart Grid Collaborative Workshop, Washington, DC*, 2008.

[9] Alexandra Von Meier, David Culler, Alex McEachern, and Reza Arghandeh. *Micro-synchrophasors for distribution systems*. IEEE, 2014.

[10] Lingjun Pu, Xu Chen, Jingdong Xu, and Xiaoming Fu. Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing. In *IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9. IEEE, 2016.

[11] Ellie D'Hondt, Matthias Stevens, and An Jacobs. Participatory noise mapping works! an evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring. *Pervasive and Mobile Computing*, 9(5):681–694, 2013.

[12] G Enrico Santagati and Tommaso Melodia. U-wear: Software-defined ultrasonic networking for wearable devices. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 241–256, 2015.

[13] Alexandros Zenonos, Sebastian Stein, and Nicholas R Jennings. Coordinating measurements for air pollution monitoring in participatory sensing settings. 2015.

[14] Baoqi Huang, Guodong Qi, Xiaokun Yang, Long Zhao, and Han Zou. Exploiting cyclic features of walking for pedestrian dead reckoning with unconstrained smartphones. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 374–385, 2016.

[15] Sarfraz Nawaz and Cecilia Mascolo. Mining users' significant driving routes with low-power sensors. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, pages 236–250, 2014.

[16] Francesco Restuccia, Nirnay Ghosh, Shameek Bhattacharjee, Sajal K Das, and Tommaso Melodia. Quality of information in mobile crowdsensing: Survey and research challenges. *ACM Transactions on Sensor Networks (TOSN)*, 13(4):1–43, 2017.

[17] Zengrui Wu, Weihua Li, Guixia Liu, and Yun Tang. Network-based methods for prediction of drug-target interactions. *Frontiers in pharmacology*, 9:1134, 2018.

[18] Weiming Wang, Yongli Cai, Mingjiang Wu, Kaifa Wang, and Zhenqing Li. Complex dynamics of a reaction–diffusion epidemic model. *Nonlinear Analysis: Real World Applications*, 13(5):2240–2258, 2012.

[19] Le Wu, Peijie Sun, Yanjie Fu, Richang Hong, Xiting Wang, and Meng Wang. A neural influence diffusion model for social recommendation. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval*, pages 235–244, 2019.

[20] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. *Science*, 298(5594):824–827, 2002.

[21] Chrysanthi Kosyfaki, Nikos Mamoulis, Evaggelia Pitoura, and Panayiotis Tsaparas. Flow motifs in interaction networks. *arXiv preprint arXiv:1810.08408*, 2018.

[22] Uri Alon. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, 8(6):450–461, 2007.

[23] Mohammad Farajollahi, Alireza Shahsavari, Emma M Stewart, and Hamed Mohsenian-Rad. Locating the source of events in power distribution systems using micro-pmu data. *IEEE Transactions on Power Systems*, 33(6):6343–6354, 2018.

[24] Daniel B Arnold, Ciaran Roberts, Omid Ardakanian, and Emma M Stewart. Synchrophasor data analytics in distribution grids. In *2017 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2017.

[25] Mahdi Jamei, Anna Scaglione, Ciaran Roberts, Emma Stewart, Sean Peisert, Chuck McParland, and Alex McEachern. Anomaly detection using optimally placed *mu*-pmu sensors in distribution grids. *IEEE Transactions on Power Systems*, 33(4):3611–3623, 2017.

[26] Armin Aligholian, Mohammad Farajollahi, and Hamed Mohsenian-Rad. Unsupervised learning for online abnormality detection in smart meter data. In *2019 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2019.

[27] Seemita Pal, Biplab Sikdar, and Joe H Chow. Classification and detection of pmu data manipulation attacks using transmission line parameters. *IEEE Transactions on Smart Grid*, 9(5):5057–5066, 2017.

[28] Jun Jiang, Xinghui Zhao, Scott Wallace, Eduardo Cotilla-Sanchez, and Robert Bass. Mining pmu data streams to improve electric power system resilience. In *Proceedings of the Fourth IEEE/ACM International Conference on Big Data Computing, Applications and Technologies*, pages 95–102, 2017.

[29] Vivek Kumar Singh and Manimaran Govindarasu. Decision tree based anomaly detection for remedial action scheme in smart grid using pmu data. In *2018 IEEE Power & Energy Society General Meeting (PESGM)*, pages 1–5. IEEE, 2018.

[30] Zhenghao Zhang, Shuping Gong, Aleksandar D Dimitrovski, and Husheng Li. Time synchronization attack in smart grid: Impact and analysis. *IEEE Transactions on Smart Grid*, 4(1):87–98, 2013.

[31] Meng Wu and Le Xie. Online detection of false data injection attacks to synchrophasor measurements: A data-driven approach. In *Proceedings of the 50th Hawaii international conference on system sciences*, 2017.

[32] Xinlei Oscar Wang, Wei Cheng, Prasant Mohapatra, and Tarek Abdelzaher. Artsense: Anonymous reputation and trust in participatory sensing. In *2013 Proceedings IEEE INFOCOM*, pages 2517–2525. IEEE, 2013.

[33] A Jøsang. An algebra for assessing trust in certificate chains. In *The Internet Society Symposium on Network and Distributed System Security February*, 1999.

[34] Bin Yu and Munindar P Singh. An evidential model of distributed reputation management. In *Proceedings of the first international joint conference on Autonomous Agents and Multiagent Systems: Part 1*, pages 294–301, 2002.

[35] Shameek Bhattacharjee, Nirnay Ghosh, Vijay K Shah, and Sajal K Das. *qnq* q n q: Quality and quantity based unified approach for secure and trustworthy mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 19(1):200–216, 2018.

[36] Kuan Lun Huang, Salil S Kanhere, and Wen Hu. Are you contributing trustworthy data? the case for a reputation system in participatory sensing. In *Proceedings of the 13th ACM international conference on Modeling, analysis, and simulation of wireless and mobile systems*, pages 14–22, 2010.

[37] Haleh Amintoosi and Salil S Kanhere. A reputation framework for social participatory sensing systems. *Mobile Networks and Applications*, 19(1):88–100, 2014.

[38] Liang Xiao, Yanda Li, Guoan Han, Huaiyu Dai, and H Vincent Poor. A secure mobile crowdsensing game with deep reinforcement learning. *IEEE Transactions on Information Forensics and Security*, 13(1):35–47, 2017.

[39] Xinlei Wang, Wei Cheng, Prasant Mohapatra, and Tarek Abdelzaher. Enabling reputation and trust in privacy-preserving mobile sensing. *IEEE Transactions on Mobile Computing*, 13(12):2777–2790, 2013.

[40] Haiming Jin, Lu Su, Danyang Chen, Hongpeng Guo, Klara Nahrstedt, and Jinhui Xu. Thanos: Incentive mechanism with quality awareness for mobile crowd sensing. *IEEE Transactions on Mobile Computing*, 18(8):1951–1964, 2018.

[41] Tie Luo, Hwee-Pink Tan, and Lirong Xia. Profit-maximizing incentive for participatory sensing. In *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pages 127–135. IEEE, 2014.

[42] Tie Luo, Salil S Kanhere, Hwee-Pink Tan, Fan Wu, and Hongyi Wu. Crowdsourcing with tullock contests: A new perspective. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2515–2523. IEEE, 2015.

[43] Tie Luo, Salil S Kanhere, and Hwee-Pink Tan. Sew-ing a simple endorsement web to incentivize trustworthy participatory sensing. In *2014 Eleventh Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, pages 636–644. IEEE, 2014.

[44] Yufeng Zhan, Yuanqing Xia, Yang Liu, Fan Li, and Yu Wang. Incentive-aware time-sensitive data collection in mobile opportunistic crowdsensing. *IEEE Transactions on Vehicular Technology*, 66(9):7849–7861, 2017.

[45] Lingjie Duan, Takeshi Kubo, Kohei Sugiyama, Jianwei Huang, Teruyuki Hasegawa, and Jean Walrand. Motivating smartphone collaboration in data acquisition and distributed computing. *IEEE Transactions on Mobile Computing*, 13(10):2320–2333, 2014.

[46] Chen-Khong Tham and Tie Luo. Quality of contributed service and market equilibrium for participatory sensing. *IEEE Transactions on Mobile Computing*, 14(4):829–842, 2014.

[47] Tie Luo, Jianwei Huang, Salil S Kanhere, Jie Zhang, and Sajal K Das. Improving iot data quality in mobile crowd sensing: A cross validation approach. *IEEE Internet of Things Journal*, 6(3):5651–5664, 2019.

[48] Ahmed F Abdelzaher, Ahmad F Al-Musawi, Preetam Ghosh, Michael L Mayo, and Edward J Perkins. Transcriptional network growing models using motif-based preferential attachment. *Frontiers in bioengineering and biotechnology*, 3:157, 2015.

[49] Shmoolik Mangan and Uri Alon. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences*, 100(21):11980–11985, 2003.

[50] Marcus Märtens, Jil Meier, Arjan Hillebrand, Prejaas Tewarie, and Piet Van Mieghem. Brain network clustering with information flow motifs. *Applied Network Science*, 2(1):1–18, 2017.

[51] Nadav Kashtan, Shalev Itzkovitz, Ron Milo, and Uri Alon. Topological generalizations of network motifs. *Physical Review E*, 70(3):031909, 2004.

[52] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. *Science*, 353(6295):163–166, 2016.

[53] Thomas E Gorochowski, Claire S Grierson, and Mario Di Bernardo. Organization of feed-forward loop motifs reveals architectural principles in natural and engineered networks. *Science advances*, 4(3):eaap9751, 2018.

[54] Azade Nazi, Mayank Raj, Mario Di Francesco, Preetam Ghosh, and Sajal K Das. Efficient communications in wireless sensor networks based on biological robustness. In *2016 International Conference on Distributed Computing in Sensor Systems (DCOSS)*, pages 161–168. IEEE, 2016.

[55] Satyaki Roy, Vijay Shah, and Sajal Das. Characterization of e. coli gene regulatory network and its topological enhancement by edge rewiring. In *Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS)*, pages 391–398, 2016.

[56] Satyaki Roy, Vijay K Shah, and Sajal K Das. Design of robust and efficient topology using enhanced gene regulatory networks. *IEEE Transactions on Molecular, Biological and Multi-Scale Communications*, 4(2):73–87, 2018.

[57] Vijay K Shah, Satyaki Roy, Simone Silvestri, and Sajal K Das. Bio-drn: Robust and energy-efficient bio-inspired disaster response networks. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pages 326–334. IEEE, 2019.

[58] Satyaki Roy, Nirnay Ghosh, Preetam Ghosh, and Sajal K Das. biomcs 2.0: A distributed, energy-aware fog-based framework for data forwarding in mobile crowdsensing. *Pervasive and Mobile Computing*, 73:101381, 2021.

[59] Lewis M Branscomb, Richard D Klausner, et al. Making the nation safer: the role of science and technology in countering terrorism. *Committee on Science and Technology for Countering Terrorism, National Research Council*, 2002.

[60] Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 388, 2016.

[61] Thomas M Chen. Stuxnet, the real start of cyber warfare?[editor's note]. *IEEE Network*, 24(6):2–3, 2010.

[62] Thomas Morris, Shengyi Pan, Jeremy Lewis, Jonathan Moorhead, Nicholas Younan, Roger King, Mark Freund, and Vahid Madani. Cybersecurity risk testing of substation phasor measurement units and phasor data concentrators. In *Proceedings of the seventh annual workshop on cyber security and information intelligence research*, pages 1–1, 2011.

[63] Daniel P Shepard, Todd E Humphreys, and Aaron A Fansler. Evaluation of the vulnerability of phasor measurement units to gps spoofing attacks. *International Journal of Critical Infrastructure Protection*, 5(3-4):146–153, 2012.

[64] Emma Stewart, Anna Liao, and Ciaran Roberts. Open μpmu: A real world reference distribution micro-phasor measurement unit data set for research and application development. 2016.

[65] Marco Pignati, Miroslav Popovic, Sergio Barreto, Rachid Cherkaoui, German Dario Flores, Jean-Yves Le Boudec, Maaz Mohiuddin, Mario Paolone, Paolo Romano, Styliani Sarri, et al. Real-time state estimation of the epfl-campus medium-voltage grid by using pmus. In *2015 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pages 1–5. IEEE, 2015.

[66] David I Urbina, Jairo A Giraldo, Alvaro A Cardenas, Nils Ole Tippenhauer, Junia Valente, Mustafa Faisal, Justin Ruths, Richard Candell, and Henrik Sandberg. Limiting the impact of stealthy attacks on industrial control systems. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, pages 1092–1105, 2016.

[67] Shameek Bhattacharjee and Sajal K Das. Detection and forensics against stealthy data falsification in smart metering infrastructure. *IEEE Transactions on Dependable and Secure Computing*, 2018.

[68] Audun Josang and Roslan Ismail. The beta reputation system. In *Proceedings of the 15th bled electronic commerce conference*, volume 5, pages 2502–2511, 2002.

[69] Shameek Bhattacharjee, Nirnay Ghosh, Vijay K Shah, and Sajal K Das. Qnq: Quality and quantity based unified approach for secure and trustworthy mobile crowdsensing. *IEEE transactions on mobile computing*, 19(1):200–216, 2018.

[70] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[71] Shameek Bhattacharjee, Saptarshi Debroy, and Mainak Chatterjee. Quantifying trust for robust fusion while spectrum sharing in distributed dsa networks. *IEEE Transactions on Cognitive Communications and Networking*, 3(2):138–154, 2017.

[72] Sushil Jajodia, Anup K Ghosh, Vipin Swarup, Cliff Wang, and X Sean Wang. *Moving target defense: creating asymmetric uncertainty for cyber threats*, volume 54. Springer Science & Business Media, 2011.

[73] Pablo Alvarez Lopez, Michael Behrisch, Laura Bieker-Walz, Jakob Erdmann, Yun-Pang Flötteröd, Robert Hilbrich, Leonhard Lücken, Johannes Rummel, Peter Wagner, and Evamarie Wießner. Microscopic traffic simulation using sumo. In *The 21st IEEE International Conference on Intelligent Transportation Systems*. IEEE, 2018.

[74] Satyaki Roy, Mayank Raj, Preetam Ghosh, and Sajal K Das. Role of motifs in topological robustness of gene regulatory networks. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.

[75] Paul Erdos, Alfréd Rényi, et al. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci*, 5(1):17–60, 1960.

[76] Trevor R Sorrells and Alexander D Johnson. Making sense of transcription networks. *Cell*, 161(4):714–723, 2015.

[77] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[78] Thomas Schaffter, Daniel Marbach, and Dario Floreano. Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. *Bioinformatics*, 27(16):2263–2270, 2011.

[79] Shai S Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of escherichia coli. *Nature genetics*, 31(1):64–68, 2002.

[80] Michael E Wall, Mary J Dunlop, and William S Hlavacek. Multiple functions of a feed-forward-loop gene circuit. *Journal of molecular biology*, 349(3):501–514, 2005.

[81] Mark Newman. *Networks*. Oxford university press, 2018.

[82] Satyaki Roy, Preetam Ghosh, Dipak Barua, and Sajal K Das. Motifs enable communication efficiency and fault-tolerance in transcriptional networks. *Scientific reports*, 10(1):1–15, 2020.

[83] Pavol Hell and Jaroslav Nešetřil. The core of a graph. *Discrete Mathematics*, 109(1-3):117–126, 1992.

[84] Andreas Krause and Daniel Golovin. Submodular function maximization. *Tractability*, 3:71–104, 2014.

[85] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134, 2007.

[86] Elchanan Mossel and Sébastien Roch. Submodularity of influence in social networks: From local to global. *SIAM Journal on Computing*, 39(6):2176–2188, 2010.

[87] George L Nemhauser and Laurence A Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of operations research*, 3(3):177–188, 1978.

[88] Aric Hagberg, Pieter Swart, and Daniel S Chult. Exploring network structure, dynamics, and function using networkx. Technical report, Los Alamos National Lab.(LANL), Los Alamos, NM (United States), 2008.

[89] Manika Kargeti and KV Venkatesh. The effect of global transcriptional regulators on the anaerobic fermentative metabolism of escherichia coli. *Molecular BioSystems*, 13(7):1388–1398, 2017.

[90] Angela Tramonti, Michele De Canio, and Daniela De Biase. Gadx/gadw-dependent regulation of the escherichia coli acid fitness island: transcriptional control at the gady–gadw divergent promoters and identification of four novel 42 bp gadx/gadw-specific binding sites. *Molecular microbiology*, 70(4):965–982, 2008.

[91] Meranda D Bradley, Michael B Beach, AP Jason de Koning, Timothy S Pratt, and Robert Osuna. Effects of fis on escherichia coli gene expression during different growth stages. *Microbiology*, 153(9):2922–2940, 2007.

[92] Kathi Canese and Sarah Weis. Pubmed: the bibliographic database. *The NCBI Handbook*, 2:1, 2013.

[93] Yuki Yamanaka, Akira Ishihama, and Kaneyoshi Yamamoto. Induction of ydeo, a regulator for acid resistance genes, by ultraviolet irradiation in escherichia coli. *Bioscience, biotechnology, and biochemistry*, 76(6):1236–1238, 2012.

[94] Aaron K Mates, Atef K Sayed, and John W Foster. Products of the escherichia coli acid fitness island attenuate metabolite stress at extremely low ph and mediate a cell density-dependent acid resistance. *Journal of bacteriology*, 189(7):2759–2768, 2007.

[95] Hiroyuki Asako, Harushi Nakajima, Kei Kobayashi, Masato Kobayashi, and Rikizo Aono. Organic solvent tolerance and antibiotic resistance increased by overexpression of mara in escherichia coli. *Applied and Environmental Microbiology*, 63(4):1428–1433, 1997.

[96] Rikizo Aonoa. Improvement of organic solvent tolerance level of escherichia coli by overexpression of stress-responsive genes. *Extremophiles*, 2(3):239–248, 1998.

[97] Pei Wang, Jinhu Lü, and Xinghuo Yu. Identification of important nodes in directed biological networks: A network motif approach. *PloS one*, 9(8):e106132, 2014.

[98] Dong Li, Zhi-Ming Xu, Nilanjan Chakraborty, Anika Gupta, Katia Sycara, and Sheng Li. Polarity related influence maximization in signed social networks. *PloS one*, 9(7):e102199, 2014.

[99] Jiliang Tang, Yi Chang, Charu Aggarwal, and Huan Liu. A survey of signed network mining in social media. *ACM Computing Surveys (CSUR)*, 49(3):1–37, 2016.

[100] Maria-Evgenia G Rossi, Bowen Shi, Nikolaos Tziortziotis, Fragkiskos D Malliaros, Christos Giatsidis, and Michalis Vazirgiannis. Mati: An efficient algorithm for influence maximization in social networks. *PloS one*, 13(11):e0206318, 2018.

[101] Robert West, Hristo S Paskov, Jure Leskovec, and Christopher Potts. Exploiting social network structure for person-to-person sentiment analysis. *Transactions of the Association for Computational Linguistics*, 2:297–310, 2014.

[102] Thomas P Trappenberg. Machine learning with sklearn. In *Fundamentals of Machine Learning*, pages 38–65. Oxford University Press, 2019.

VITA

Prithwiraj Roy was born in Midnapore, West Bengal, India. He graduated with a Bachelor of Engineering (B.E.) in Electronics and Telecommunication Engineering in 2010 from Indian Institute of Engineering Science and Technology, Shibpur, India. Subsequently, he worked for Tata Consultancy Services (TCS), and IBM as a software engineer during 2010-2017. He joined Missouri University of Science and Technology, Rolla, USA as Ph.D. scholar in Computer Science in Fall 2017 under Dr. Sajal K. Das. During this time, he worked as a graduate research assistant at the Center for Research in Wireless Mobility and Networking (CReWMaN) lab and he also served as graduate teaching assistant in two graduate courses. In December 2021, he received his Ph.D. in Computer Science from Missouri University of Science and Technology.