Resilient Control and Intrusion Detection for SCADA Systems

Bonnie Xia Zhu

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2014-34
http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-34.html

May 1, 2014

Acknowledgement

Many heartfelt thanks are due -- To my advisor, Prof. Shankar Sastry, for inspiring me with his passion for exploring new technologies and new research areas; To my mentor, Prof. Vern Paxson, for guiding me through my early attempts on network security; To my committee members: Profs. Doug Tygar, Anthony Joseph, Andrew Packard; To numerous Berkeley professors for having remarkably positive impact on me, including but not limited to: Profs. K.P., Seth Sanders, Richard Karp, Scott Shenker, Randy Katz, Jean Walrand, Pravin Varaiya; To Ruth Gjerde for her resilience and her wise advices at needed times; To MaMa, for her unapologetic self-efficacy and unconditional sacrifice; To BaBa, for his genuine love for work and innovation, perpetual perseverance and optimism.

Resilient Control and Intrusion Detection for SCADA Systems

by

Xia Bonnie Zhu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Science

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor S. Shankar Sastry, Chair
Professor J. Doug Tygar
Professor Anthony D. Joseph
Professor Andrew K. Packard

Fall 2011

Resilient Control and Intrusion Detection for SCADA Systems

Abstract

Resilient Control and Intrusion Detection for SCADA Systems

by

Xia Bonnie Zhu

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Science

University of California, Berkeley

Professor S. Shankar Sastry, Chair

Supervisory Control and Data Acquisition (SCADA) systems are deeply ingrained in the fabric of critical infrastructure sectors. These computerized real-time process control systems, over geographically dispersed continuous distribution operations, are increasingly subject to serious damage and disruption by cyber means due to their standardization and connectivity to other networks. However, SCADA systems generally have little protection from the escalating cyber threats. To achieve defense-in-depth for SCADA systems by means of intrusion detection and resilient control, this dissertation strives for a robust stochastic signal and system approach without being overly-pessimistic. Its main elements are (1) two SCADA-specific comprehensive taxonomies with one on cyber attacks and the other on intrusion detection system to layout the lay of the land and shed light to the workspace, (2) one overall framework/architecture for intrusion detection and resilient control – *Xware* (3) its measurement fusion assurance component – *Trust counter*, (4) one signal-based early-detection and resilient estimation scheme with proved theoretical performance bounds, for SCADA systems in general. Especially the said *Robust General Likelihood Ratio Test* (RGLRT) is generic enough and has been applied to linear dynamical systems in general and beyond. (5) The application of RGLRT in network traffic anomaly detection. (6) The application of RGLRT to anomaly detection for SCADA systems in smart grids through model construction and identification for both clean renewable energy supply and variable consumer demand.

First, in order to understand the potential danger and to protect SCADA systems, we highlight their difference from standard *Information Technology* (IT) systems and present a set of security property goals. Furthermore, we systematically identify and classify likely *cyber attacks* including cyber-induced *cyber-physical attacks* on SCADA systems are according the SCADA's hierarchy. Determined by the impact on control performance of SCADA systems, we use the attack categorization criteria to stress the commonalities and important features of such attacks that define unique challenges posed to securing SCADA systems versus traditional IT systems.

Second, in order to address the big challenge of how to modify conventional IT intrusion detection techniques to suit the needs of SCADA, we explain the nuance associated with the task of SCADA-specific intrusion detection and frame it in the domain interest of control's researchers to illuminate problem space. We present a taxonomy and a set of metrics for SCADA-specific intrusion detection techniques through heightening their possible use in SCADA systems. In particular, we enumerate a list of *Intrusion Detection Systems* (IDS) that have been proposed to

undertake this endeavor. Drawing upon the discussion, we identify the deficits and voids in current research. Based upon this taxonomy and analysis on which SCADA-specific IDS strategies are most likely to succeed, we offer recommendations and future research venues in part through presenting a prototype of such efforts towards this goal.

Third, we present the overall architecture for instruction detection and resilient control *Xware*. It is comprised of two strong footings – *Normalcy Checking*, a control theoretic, domain knowledge specific, specification-based payload inspection system and a high-speed, real-time, behavioral-based *Network Intrusion Detection System* (NIDS). *Xware* integrates a *Trust Counter* to verify the truthfulness of sensor measurements. It also provides exfiltration of confidential information from within the intranet. Moreover, *Xware* hardens SCADA system with compensation schemes when intrusion evades NIDS or unexpected fault occurs to guarantee its performance. It puts things in perceptive and highlights the overall systematic and holistic approach.

Fourth, we propose the *Trust Counter* to deal the cases when the possible manifestation of those potential disruption from cyber attacks can affect the Kalman filter, the primary recursive estimation method used in the control engineering field. Whereas, to improve such estimation, data fusion may take place at a central location to fuse and process multiple sensor measurements delivered over the network. In an uncertain networked control system where the nodes and links are subject to attacks, false or compromised or missing individual readings can produce skewed results. To assure the validity of data fusion, a centralized trust rating system is proposed. It evaluates the trustworthiness of each sensor reading on top of the fusion mechanism. The ratings are represented by Beta distribution, the conjugate prior of the binomial distribution and its posterior. Then an illustrative example demonstrates its efficiency.

Fifth, RGLRT is an earlier anomaly detection and resilient estimation scheme for the cyber-physical systems, networked control systems to be specific, in an uncertain network environment. It *robustly* identifies and detects outliers among real-time multidimensional measurements of dynamical systems by using an online window-limited sequential *Robust Generalized Likelihood Ratio* (RGLR) test without any prior knowledge of the occurrence time and distribution of the outliers. The robust sequential testing and quick detection scheme achieves the optimal stopping time with low rates in both false alarm and misdetection. We propose a set of qualitative and quantitative *metric* to measure its optimality in the context of cyber-physical systems. Further, this resilient and flexible estimation scheme *robustly* rectifies and cleans data upon both isolated and patchy outliers while maintain the optimality of the Kalman Filter under the nominal condition. Its approximated optimality of the robustification performance is shown through *stochastic approximation*.

Sixth, we give a network anomaly detection scheme as one of the applications of RGLRT. The time series model of Autoregressive Integrated Moving Average (ARIMA) progress, finds its wide usage including network security applications. Model building and anomaly detection based on such models are often a first and important step towards monitoring unexpected problems and assuring the soundness and security of those systems being studied. The time variability by the coefficients in those dynamic regression models is particularly relevant and possibly indicative. To address this issue, a corresponding framework and a novel anomaly detection approach based on the Kalman filter for identifying those dynamic models including their parameters and a General Likelihood Ratio (GLR) test for detecting suspicious changes in the parameters and therefore the

models is proposed. The idea is shown through experiments and show its promising potential in terms of accuracy and robustness.

Seventh, we apply RGLRT to anomaly detection for SCADA systems in smart grids. While the utilization of clean energy resources including wind and solar power sets to grow from filling the gap of peak hours to taking a larger share in the upcoming smart grid and efficient infrastructure, the price-incentivized electricity consumption shall alleviate peak hours and reduce power outages. Both benign faults and malicious attacks threat the reliability and availability of the new grid. We address these duo problems are from the angle of one fundamental technique used. The ARIMA time series models play roles at both ends in this new ecosystem: namely, predicting the variable clean energy resource on the supply side and forecasting the flexible load demand on the consume side. Model construction and anomaly detection based on such models are often a first and important step towards monitoring unexpected problems and assuring the soundness and security of those systems being studied. The time variability of the coefficients in those dynamic regression models is particularly relevant and possibly indicative. Thus a corresponding framework and a novel anomaly detection approach is introduced. It's based on a robustified Kalman Filter for identifying those dynamic models including their parameters and a RGLRT for detecting suspicious changes in the parameters and therefore the models. Currently, the effectiveness and robustness of this method is shown through simulation.

Citius, Altius, Fortius

To My Parents.

致爸爸和妈妈

Contents

List of Figures

Acknowledgments

Many heartfelt thanks are due,

To my advisor, Prof. Shankar Sastry, for inspiring me with his passion for exploring new technologies and new research areas, for opening the window to nonlinear system theory for me and teaching me that nothing is impossible yet equations are forever, and for his trust in my capabilities;

To my mentor, Prof. Vern Paxson, for guiding me through my early attempts on network security beyond his thought-provoking class, more importantly for influencing me to be grounded with the appreciation that *the devil is in the details*, and for his kindness;

To my committee members:

Prof. Doug Tygar for sharing his passion in teaching, for encouraging my security research efforts, for coaching me on how research paper should be rightly written and a good presentation well delivered;

Prof. Anthony Joseph for giving his consistent support to my research on SCADA security from day one and for consistently dedicating his time and expertise to every meeting when I hosted and organized the Securing SCADA Berkeley Study/Work Group;

Prof. Andrew Packard for his time, patience and helpful feedbacks;

To numerous Berkeley Professors for having remarkably positive impact on me, including but not limited to: Prof. K.P., Prof. Seth Sanders, Prof. Richard Karp, Prof. Scott Shenker, Prof. Randy Katz;

To many domain experts with whom I have been lucky enough to have interactions:

Prof. Roy Maxin at CMU for sharing his take on anomaly detection with me and pointing out doing cross-discipline research means to master both disciplines;

Dr. Ulf Lindqvist at SRI for his professional guidance and especially his feedback on the *Xware* architecture during IEEE S&P Oakland WIP session in 2008;

Dr. John James for his intellectual generosity beyond help on data access attempts;

Dr. Karsten Nohl at Security Research Labs for his suggestion on smart meter related security issues;

Dr. Matthew Stillerman at ATC-NY for sharing his work on SCADA power grids security;

Dr. Walt Heimerdinger at Honeywell ACS Laboratories for sharing EPRI report on intrusion detection for the electric power grid.

Mr. Tom Phinney for sharing his field knowledge on SCADA systems and his attentive and detailed correspondence;

Dr. Rob Cunningham MIT Lincoln Lab for help on data access attempt;

Mr. Dale Peterson at Digital Bond for data access attempt;

To every Security Reading Group folk for many stimulating discussions and Prof. David Wagner for many insightful comments;

To the whole NetEcom group meeting group and Prof. Jean Walrand and Prof. Pravin Varaiya in particular for their advices, comments and feedbacks on the progress of my research work;

To the entire *Tea*, the machine learning research meeting group for their consistent supply on latest machine learning techniques and especially Prof. Michael Jordan for his input on my application of graphic modeling tools;

To the Robust Statistics reading group and its organizer Blaine Nelson particularly for useful discussions on robust statistics;

To the financial support for this research work in part from the National Science Foundation Award CCF-0424422 for the Team for Research in Ubiquitous Secure Technology (TRUST) and the following organizations: AFOSR (#FA9550-06-1-0244), BT, Cisco, DoCoMo USA Labs, EADS, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, TCS, Telecom Italia, and United Technologies ;

To the full scholarship from NSF and DoE for attending Process Control Systems Summer School 2008;

To the 2009 Energy and Sustainability Fellowship for carrying out experiments on SCADA wind power system;

To the travel grants for attending Grace Hopper, USENIX conferences;

To WiCSE for giving me the opportunity to serve as a co-president;

To Berkeley Fencing Club for inspiring me with good sportsmanship and new research ideas on attacks and games;

To Berkeley for nurturing me with intellectual nutrients, stimulating me with cerebral excitements and lavishing me with phytogenetic sceneries, a beautiful California sky and uphill trails as marathon training grounds.

To my friends afar and close, office buddies and campus peers: Ardain Mettler, Ari, Arel, David Molnar, Fengming, Tamara, Rebecca, Lynn, Elaine, Dr.Qiu, Alison, Beth, Sarah, Alex, Annarita, Edga, Phebous, Ted, Humberto, Fernando, Ram, Maryam, Sam, Andrew, Lily, Anil, Galina, Saurahb, Sally Alcala, Larry, Phil, Garry Givens, Jessica Gamble, Mary Stewart, Christen Gates, Eddie, Amiee, Dana and folks at CUSG help desk;

To Ruth Gjerde for her resilience and her wise advices at needed times;

To Dr. David Levine, Dr. Jung Hwan Ahn, and Dr. Karen Langer at NYU for their tremendous help and for fostering my interest in neuroscience and neuropsychology;

To my sister's family especially my two lovely nephews Nolie and JuJu for their non-ending inquisitive questions;

To WaiPo, for her love;

To MaMa, for her unapologetic self-efficacy and unconditional sacrifice;

To BaBa, for his genuine love for work and innovation, perpetual perseverance and optimism.

Chapter 1

Introduction

Due to their standardization and connectivity to other networks, *Supervisory Control and Data Acquisition* (SCADA) systems are increasingly subject to damage and disruption by cyber means. However, the issues facing securing SCADA system are: (1) regulation-wise: Lack of policies or standards, (2) technology-wise: the need for *availability*, *integrity*, *confidentiality* is only met with limited specialized solutions, (3) economics- and finance-wise: lack of economic justification, (4) markets-wise: they are legacy systems, where lack of demands from operators: organizational priorities conflict.

In particular, SCADA present challenges for security engineering due to their requirements for continuous availability, real-time operation, potential impact on the populace and the physical world, and legacy deployments. They further play crucial roles in the fabric of critical infrastructure such as electric power grids, water distribution systems, petroleum and natural gas pipelines, and manufacturing operations.

The cyber-physical security of real-time, continuous systems necessitates a comprehensive view and holistic understanding of network security, control theory and the physical system. Ultimately, any viable technical solutions and research directions in securing SCADA systems must lie in the conjunction of computer security, communication network and control engineering. However, the very large installed base of such systems means that in many instances we must for a long time to come rely on retrofitted security mechanisms, rather than having the option to design them in from scratch. This leads to a pressing need for robust SCADA-specific intrusion detection systems (IDS) and resilient control.

The goals of this effort are to develop IDS and resilient control technology that can (1) efficiently detect and block cyber intrusions into SCADA systems in entrenched operational environments, in real-time, (2) without interrupting the control performance of the protected system, (3) without creating extra operational burden or operational reservations due to false alarms, (4) in the presence of both malicious and messily benign network traffic, (5) and lastly rectify and compensate the system performance in case some intrusions succeed. The system must operate in a real-time, robust fashion, with performance adequate to meet the demands of the dynamic cyber-physical interactions inherent to SCADA systems.

To this end, we formulate a number of objectives,

- Conceptualize control performance - oriented metrics for mentioned security measures,

- Develop usage- and goal-oriented taxonomies of cyber attacks on SCADA system and SCADA-specific IDS to shed insight onto the problem domain.
- Establish prudent and plausible threat models,
- Characterize the system architecture, protocol use, network topology, and network activity of SCADA systems used in power grid, particularly.
- Create models of both normal operation and the allowed range of operation (ala' specification-based intrusion detection) to enable detection of new attacks while maintaining low false alarm rates during legitimate changes of a SCADA system's dynamics and permitted variations in its traffic, including valid safety system responses at extreme cases. Unique to this problem domain, such models can draw upon insight into expected and allowed behavior that we can "analytically" derive form the underlying control system principles and properties.
- Find asymptotic performance bounds on these models.
- Integrate a network IDS with these models to enable a resilient, defense-in-depth, SCADA-domain network monitoring, and online data clearing & control compensation in case certain intrusions succeed.
- Construct a test environment to verify the IDS performance in terms of its resistance to evasion and ability to detect and block attacks against a given SCADA system with acceptable low false alarm rate.
- Conduct experiments to confirm the system's resilience level in case certain attacks succeed.

Chapter 2

A Taxonomy of Cyber Attacks on SCADA Systems

> Example is the school of mankind, and they will learn at no other.
>
> *Letters on a Regicide Peace*
> EDMUND BURKE

Supervisory Control and Data Acquisition (SCADA) systems are deeply ingrained in the fabric of critical infrastructure sectors. These computerized real-time process control systems, over geographically dispersed continuous distribution operations, are increasingly subject to serious damage and disruption by cyber means due to their standardization and connectivity to other networks. However, SCADA systems generally have little protection from the escalating cyber threats. In order to understand the potential danger and to protect SCADA systems, in this paper, we highlight their difference from standard IT systems and present a set of security property goals. Furthermore, we focus on systematically identifying and classifying likely *cyber attacks* including cyber-induced *cyber-physical attacks* on SCADA systems. Determined by the impact on control performance of SCADA systems, the attack categorization criteria highlights commonalities and important features of such attacks that define unique challenges posed to securing SCADA systems versus traditional Information Technology (IT) systems.

The utilization of *Supervisory Control and Data Acquisition* (SCADA) systems facilities the management with remote access to real-time data and the channel to issue automated or operator-driven supervisory commands to remote station control devices, or *field devices*. They are the underlying control system of most critical national infrastructures including power, energy, water, transportation, telecommunication and are widely involved in the constitutions of vital enterprises such as pipelines, manufacturing plants and building climate control.

Remote locations and proprietary industrial networks used to give SCADA systems a considerable degree of protection through isolation [153, 78]. Most industrial plants now employ networked process historian servers for storing process data and other possible business and process interfaces. The adoption of Ethernet and transmission control protocol/Internet protocol TCP/IP for process control networks and wireless technologies such as IEEE 802.x and Bluetooth has

further reduced the isolation of SCADA networks. The connectivity and de-isolation of SCADA system is manifested in Figure 2.1.

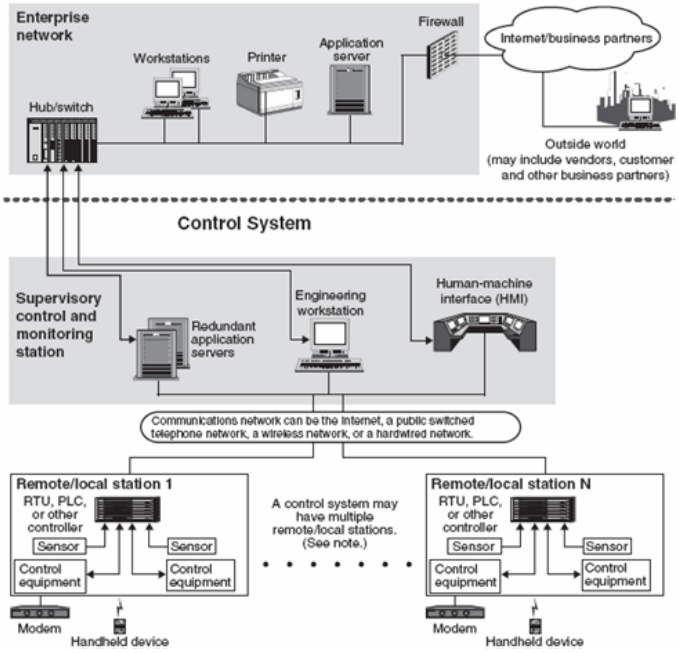


Figure 2.1: Typical SCADA Components Source: United States Government Accountability Office Report. GAO-04-354 [78]

Furthermore, the recent trend in standardization of software and hardware used in SCADA systems makes it even easier to mount SCADA specific attacks. Thus the security for SCADA systems can no longer rely on obscurity or on being a function of locking down a system.

These attacks can disrupt and damage critical infrastructural operations, cause major economic losses, contaminate ecological environment and even more dangerously, claim human lives.

The British Columbia Institute of Technologys Internet Engineering Lab (BCIT/IEL) maintains an industrial cyber security incident database [28] with more than 120 incidents logged since the initiation. Baker et al at McAfee in their 2011 sequel report [19] surveyed 200 IT security executives in 14 counties from critical electricity infrastructure enterprises, where SCADA systems are widely used, and found out most facilities have been under cyber attacks.

Being one of most sophisticated SCADA malware known to date[1], Stuxnet according to Falliere et. al at Symantec [70], takes advantage of multiple Windows zero-day vulnerabilities and targets the command-and-control software installed in industrial control systems world-wide. It sabotages facilities by reprogramming *Programmable Logic Controllers* (PLCs) to operate as the attackers intend them, most likely out of their specified boundaries while its "misreporting" feature hides the incident from the network operations center. As of April 21st 2011, more than 50 new Stuxnet-like attacks beckon SCADA threats have been discovered [194].

[1]In McAfee's report [19], nearly half of those being surveyed in the electric industry said that they had found Stuxnet on their systems.

Most related works have focused on the classification and categorization of attacks on standard IT systems such as [104, 115, 144], communication standards and/or protocols [167], communication devices [171]. There are work done to enumerate possible attacks on small embedded systems [82, 225]. More recently, SCADA-specific security solutions are proposed [75] and SCADA-specific *Intrusion Detection Systems* (IDS) are evaluated [302].

The remainder of this chapter is organized as the follows. Section 2 compares SCADA systems with standard IT properties that attribute to their security concerns. Section 3 defines desired security properties, trust model and threat model. Section 4 states vulnerabilities that embedded in SCADA systems. Section 5,6,7 numerate cyber attacks on hardware, software, communication stacks respectively. Section 8 concludes.

2.1 Difference from IT

In SCADA systems, or control systems in general, the fact that any logic execution within the system has a direct impact in the physical world dictates safety to be paramount. Being on the first frontier to directly face human lives and ecological environment, the field devices in SCADA systems are deemed with no less importance than central hosts [2] [42]. Also certain operating systems and applications running on SCADA systems, which are unconventional to typical IT personnel, may not operate correctly with commercial off-the-shelf IT cyber security solutions.

Furthermore, factors like the continuous availability demand, time-criticality, constrained computation resources on edge devices, large physical base, wide interface between digital and analog signals, social acceptance including cost effectiveness and user reluctance to change, legacy issues and so on make SCADA system a peculiar security engineering task.

SCADA systems are *hard real-time systems* [251] because the completion of an operation after its deadline is considered useless and potentially can cause cascading effect in the physical world. The operational deadlines from event to system response imposes stringent constraints: missing deadline constitutes a complete failure of the system. Latency is very destructive to SCADA system's performance: the system does not react in a certain time frame would cause great loss in safety, such as damaging the surroundings or threatening human lives.

It's not the length of time frame but whether meeting the deadline or not distinguishes hard real-time system from soft real-time system. In contrast, *soft real-time systems*, such as live audio-video systems, may tolerate certain latency and respond with decreased service quality, eg. dropping frames while displaying a video. Non-major violation of time constraints in soft real-time systems leads to degraded quality rather than system failure.

Furthermore due to the physical nature, tasks performed by SCADA system and the processes within each task are often needed to be interrupted and restarted. The timing aspect and task interrupts can preclude the use of conventional encryption block algorithms.

As *Real-time operating system* (RTOS), SCADA's vulnerability also rises from the fact that memory allocation is even more critical in an RTOS than in other operating systems. Many field

[2] Although arguably, a compromised central serverl/controller may cause server harm if the field devices don't have their own individual and local protection.

level devices in SCADA system are embedded systems that run years without rebooting but accumulating fragmentation.

Thus, buffer overflow is more problematic in SCADA than in traditional IT.

2.2 Problem Statement

Before we state the security properties that are desirable for SCADA systems to achieve, we must point out that there are many trade-offs between security and control performance goals. And we will group attacks according to the hierarchy of the SCADA system.

2.2.1 Security Property Goal

Control systems have many characteristics that are different from traditional IT systems in terms of risks and operational priorities thus render unique performance and reliability requirements besides the use of operating systems and applications being unconventional to typical IT personnel.

Even where security is well defined, the primary goal in the Internet is to protect the central server and not the edge client. In process control, an edge device, such as PLC or smart drive controller, is not necessarily merited less importance than a central host such as data historian server [42], as they are on the first frontier facing human lives and ecological environment.

These differences between SCADA systems and IT systems demand an adjusted set of security property goals and thus security and operational strategies.

In the traditional IT community, the set of common desirable security properties are *confidentiality*, *integrity* and *availability*, or *CIA* in short. The paramount, in IT's world is confidentiality and integrity while in control systems is system availability and data integrity as result of human and plant safety being its primary responsibility.

Particularly, most of computer security research focus on confidentiality. To be SCADA system specific, we prioritize security properties of SCADA systems in the order of its importance and desirability in industry, especially in control engineering sector. The modification we make addresses the special needs incurred from the unique characteristics of SCADA systems, namely the time criticality, dispersed distributed-ness and continuous availability.

There are different versions of definition and use of security properties [12] with slight variations. However, in light to differentiate the uniqueness of control systems from standard IT systems, it's necessary for us to stress and explain some more relevant subtleties. Nevertheless, it's not to say that these properties we want to highlight are mutual exclusive, absent of over-lapping.

Timeliness

explicitly expresses the time-criticality of control systems, a given resulted from being real-time system, and the concurrencies in SCADA systems due to being widely dispersed distributed systems.

It includes both the *responsiveness* aspect of the system, e.g. a command from controller to actuator should be executed in real-time by the latter, and the timeliness of any related data being

delivered in its designated time period, by which, we also mean the *freshness* of data, i.e., the data is only valid in its designated time period. Or in a more general sense, this property describes that any queried, reported, issued and disseminated information shall not be stale but corresponding to the real-time and the system is able and sensitive enough to process request, which may be of normal or of legitimate human intervention in a timely fashion, such as within a sampling period. In reality, if arrives late or repeatedly to the specified node, a message is no longer any good, be it a correct command to an actuator or a perfect measurement from a sensor with intact content. As a matter of fact, any replay of data easily breaches this security goal.

Moreover, this property also implicitly implies the order of updates among peered sensors, especially if they are observing the same process or correlated processes. The order of data arrival at *central monitor room* may play an important factor in the representation of process dynamics and affect the correct decision making of either the controlling algorithms or the supervising human operators.

In a nutshell, all right data should be processed in *right* time, which unfolds an underpinning security goal – *secure time* provision.

Availability

means when any component of a SCADA system, may it be a sensory or servomechanical device, communication or networking equipment, or radio channel; computation resource and information such as sensor readings and controller commands etc. that transmits or resides within the system should be ready for use when is needed. Most of SCADA controlled processes are continuous in nature. Unexpected outages of systems that control industrial processes are not acceptable. This desired property for both SCADA systems control performance and security goal requires that the security mechanism employed onto SCADA systems, including but not limited to the overall cryptographic system, shall not degrade the maintainability, operability , and its accessibility at emergency, of the original SCADA system without those security oriented add-ons.

Integrity

requires data generated, transmitted, displayed, stored within a SCADA system being genuine and intact without unauthorized intervention, including both its content, which may also include the header for its source, destination and time information besides the payload itself. A very related terminology is *authenticity*, in the context of SCADA system, it implies that the identity of sender and receiver of any information shall be genuine. Using our definition of integrity, then authenticity falls within the same category. One can image how disastrous the consequence can be, if a control command is redirected to an actuator other than its intended receiver or fake or wrong source information of a sensor measurement being reported to the central controller. The *intra-message integrity* means specifically the content of message to be genuine and *inter-message integrity* refers to assure data integrity, the protocol must prevent an adversary from constructing unauthentic messages, modifying messages that are in transit, reordering messages, replaying old messages, or destroying messages without detection.

Confidentiality

refers to that unauthorized person should not have any access to information related to the specific SCADA system. At current stage, this need is dwarfed by the desirability of availability in a control performance-centric setting. SCADA systems measure and control physical processes that generally are of a continuous nature with commands and responses are simple and repetitive. Thus the messages in SCADA systems are relatively easy to predict. Hence confidentiality is secondary in importance to data integrity.

However, the confidentiality of critical information such as passwords, encryption keys, detailed system layout map and etc. shall rank high when it comes to security concerns in industry. Applicable reinforcement should be imposed in this aspect. Also, the information regarding physical content flowed within the control algorithm may be subject to leaking critical message to side channel attacks.

The drastic difference in the ordering of desired security properties is mostly due to that SCADA systems are demanded to be real-time operating and continuously functioning.

Graceful Degradation

requires the system being capable of keeping the attack impact local and withholding tinted data flow within *tainted* region without further escalating into a full scale, full system cascading event.

Again, all these desired security properties are not mutual exclusive but closely related. For example, by breaching integrity, an adversary can change control signals to cause a device malfunction which might ultimately affect the availability of the network. Overall, a tightly enforced **access control** may render confidentiality, integrity, availability , timeliness and graceful degradation as well.

2.2.2 Trust Model

Given that we focus on the cyber attacks on SCADA system, we restrain our attention to attacks mounted through cyber means [3] and assume the basic physical security is provided. Particularly, the *SCADA server* or *Master Terminal Unit* is physically secure, i.e., we assume there are no direct physical tampering on the server where the main control and estimation algorithms reside. Brute force physical sabotage such as cutting wires and cables from communication and power supply or hammering devices or radio jamming are out the scope of this paper.

Furthermore, we assume that the control and estimation algorithms are programmed securely.

[3] As stated in previous sections, these cyber attacks are most likely resulted in physical destruction in SCADA systems.

2.2.3 Threat Model

Typical threats to sensor networks and to conventional IT systems are also threats to SCADA systems if the adversarial have means to exploit the vulnerabilities of SCADA systems[4]. The adversary sources include but not limited to hostile governments, terrorist groups, foreign intelligence services, industrial spies, criminal groups, disgruntled employees, bot-network operators, phishers, spywaremalware authors, spammers, and attackers [80]. We assume attacks come from one side of SCADA center only and there's no collusion.

2.3 Vulnerability

The current common practice of SCADA system leaves window open to various vulnerabilities. To name a few, the entrenched factors are not limited to public information like a company's network infrastructure, insecure network architecture, operating system vulnerabilities enabled trap doors to unauthorized users and the use of wireless devices. In particular, the lack of real-time monitoring and proper encryption is very detrimental.

Cyber attacks on SCADA system can take routes through Internet connections, business or enterprise network connections and or connections to other networks, to the layer of control networks then down the level of field devices. More specifically, the common attack vectors are

- Backdoors and holes in network perimeter
- Vulnerabilities in common protocols
- Attacks on field devices through cyber means
- Database attacks
- Communications hijacking and *Man-in-the-middle* attacks
- *Cinderella* attack on time provision and synchronization

From the point view of a control engineer, possible attacks can be grouped into following categories

- bogus input data to the controller introduced by compromised sensors and/or exploited network link between the controller and the sensors
- manipulated and misleading output data to the actuators/reactors from the controller due to tempered actors/ reactors or compromised network link between the controller and the actuators
- controller historian

[4] Note we are making a rather conservative assumption in light of exploring the potentials of cyber security issues in the SCADA system domain. Any further suitable and refined threat model depends on the cost effectiveness of the security measures.

- Denial of Service – missing the deadlines of needed task actions.

There is still little reported information about actual SCADA attacks nor scenarios designed by red-teams, despite the growing awareness of security issues in industrial networks. However, by leveraging the existing solution and understanding of the conventional IT system, we use the SCADA hierarchy as a reference plane. Then the classification of cyber attacks can fall into the following categories.

2.4 Cyber Attacks on Hardware

Attacker might gain unauthenticated remote access to devices and change their data set points. This can cause devices to fail at a very low threshold value or an alarm not to go off when it should. Another possibility is that the attacker, after gaining unauthenticated access, could change the operator display values so that when an alarm actually goes off, the human operator is unaware of it. This could delay the human response to an emergency which might adversely affect the safety of people in the vicinity of the plant. Some of the detailed procedure of achieve such attacks are given out in later section when we describe specific SCADA protocols.

The main issue in preventing cyber attacks on hardware is access control. With that in mind, we should mention one of the representative attacks in this category, namely the doorknob-rattling attack. The adversary performs a very few common username and password combinations on serval computers that results in very few failed login attempts. This attack can go undetected unless the data related to login failures from all the hosts are collected and aggregated to check for doorknob-rattling from any remote destination.

2.5 Attacks on Software

As listed in earlier sections, SCADA system employs a variety of software to meet its functionality demands. Also there are large databases reside in data historians besides many relational database applications used in cooperate and plant sessions.

Hosting centralized database , data historians contain vital and potentially confidential process information. These data are not only indispensable for technical reasons, such as that many control algorithms rely on past process data to make correct decisions, but also for business purposes, such as electricity pricing.

Although we've assumed the algorithms of these softwares are trustworthy, there are still vulnerabilities associated with their implementations. The most common implementation flaw is buffer overflow among others such as format string, integer overflow and etc. The fact that most control applications are written in C requires us to take extra precaution with this vulnerability.

2.5.1 No Privilege Separation in Embedded Operating System

VxWorks was the most popular embedded operating system in 2005 and claimed 300 million devices in 2006 [212], which is a platform developed by Wind River Systems and has since been

acquired by Intel [190]. VxWorks has been used to power everything from the Apple Airport Extreme access points to the Mars rovers and the C-130 Hercules aircraft [182]. VxWorks itself is essentially a monolithic kernel with applications implemented as kernel tasks, This means that all tasks generally run with the highest privileges and there is little memory protection between these tasks.

2.5.2 Buffer Overflow

Many attacks boil down to cause buffer overflow as their eventual means to corrupt the intended behavior of the program and cause it to run amok. Some general methods are stack smashing and manipulating function pointer.

The effect of such attacks can take forms such as resetting passwords, modifying content, running malicious code and so on.

The buffer overflow problem in SCADA system takes two fronts. One front is on the workstations and servers which are similar to standard IT systems.

For example, WellinTech KingView 6.53 HistorySvr, an industrial automation software for historian sever widely used in China, has a heap buffer overflow vulnerability that could potentially become the risk of a Stuxnet type mishap if not patched [32] .

The other front manifests itself in field devices and other components that rely on RTOS thereof inherent the susceptible memory challenge. Exploits can take advantage of the fixed memory allocation time requirement in RTOS system to have more successful launchings. Let alone that many field devices run for years without rebooting. Therefore, these SCADA components, especially in legacy networks, are subject to accumulated memory fragmentation, which leads to program stall.

The Hardware/Software Address Protection (HSAP) technique offered by [246] including hardware boundary check method and function pointer XOR method to deal with stack smashing attack and function pointer attack in embedded systems , respectively.

2.5.3 SQL Injection

Most small and industrial- strength database applications can be accessed using Structured Query Language (SQL) statements for structural modification and content manipulation. In light of data historians and web accessibility in current SCADA systems, SQL injection, one of the top Web attacks, has a very strong implication on the security of SCADA system.

The typical unit of execution of SQL which comes in many dialects loosely based around SQL-92 ANSI standard is *query*, which is a collection of statements that typically return a single *result set*. SQL injection occurs when an adversary is able to manipulate data input into an Web application, which fails properly sanitize user-supplied input, and to insert a series of unexpected SQL statements into a query. Thus it is possible to manipulate a database in several unanticipated ways. Moreover, if a “command shell” store procedure is enabled, an attacker can move further to prompt level. The process will run with the same permissions as the component that executed the command. The impact of this attack can allow attackers to gain total control of the database or even execute commands on the system.

In the case studied in [206], where the store procedure in SQL server (shown in Fig.2.2) is enabled by default. Thus an attacker still can get into SCADA system even though two LAN cards are installed.

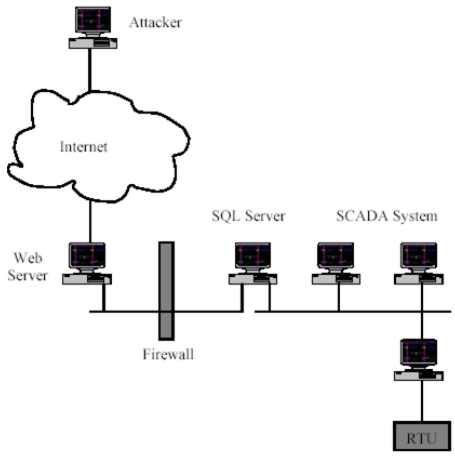


Figure 2.2: SQL Attack

Intentionally malicious changes to databases can cause catastrophic damage.

2.6 Attacks on the Communication Stack

We break down the attacks on the communication stack by using the TCP/IP or the Internet reference model and highlight some of those may have more potentials in harming SCADA systems, in particular on *network layer, transport layer, application layer* and the *implementation of protocols*.

The UDP back door on port $0x4321$ on thousands of devices is known in the public since at least spring 2002.

There are many well-known TCP/IP attacks in literature, readers please refer to [115, 104] for more details.

2.6.1 Network Layer

Diagnostic Server Attacks through UDP port

Adversaries have access to the same debugging tools that any RTOS developers do. They can read symbol tables, step through the assembly, etc., considering also that many attackers don't even need code-level knowledge. For example Wind River Systems VxWorks weak default hashing algorithm in standard authentication API for VxWorks is susceptible to collisions, an attacker can brute force a password by guessing a string that produces the same hash as a legitimate password [5]. Or through VxWorks debug service runs UDP on port 17185, which is enabled by

[5] US-Cert VU #840249.

default, an attacker can execute the following attacks without any authentication required while maintaining a certain level of stealthiness such as remote memory dump, remote memory patch, remote calls to functions, remote task management [6].

The VxWorks Wind DeBug (WDB) is an RPC-based protocol which uses UDP can explored over the Internet by downloading hacking software and adding targets to a host list before running the script.

Idle Scan

is to blind port scan by bouncing off a dumb "zombie" host, often a preparation for attack. Both MODBUS and DNP3 have scan functionalities prone to such attacks when they are encapsulated for running over TCP/IP.

Smurf

is a type of address spoofing, in general, by sending a continuous stream of modified *Internet Control message Protocol*(ICMP) packets to the target network with the sending address is identical to one of the target computer addresses. In the context of SCADA systems, if an PLC acts on the modified message, it may either crash or dangerously send out wrong commands to actuators.

Address Resolution Protocol (ARP) Spoofing/Poisoning

The ARP is primarily used to translate IP addresses to Ethernet Medium Access Control (MAC) addresses and to discover other connected interfaced device on the LAN. The ARP spoofing attack is to modify the cached address pair information.

By sending fake ARP messages which contain false MAC addresses in SCADA systems, an adversary can confuse network devices, such as network switches. When these frames are falsefully sent to another node, packets can be sniffed; or to an unreachable host, DoS is launched; or intentionally to an host connected to different **actuators**, then *physical disasters* of different scales are initiated.

Static MAC address is one of the counter measures. However, certain network switches do not allow static setting for a pair of MAC and IP address. Segmentation of the network may also be a method to alleviate the problem in that such attacks can only take place within same subnet.

Chain/Loop Attack

In a chain attack, there is a chain of connection through many nodes as the adversary moves across multiple nodes to hide his origin and identity. In case of a loop attack, the chain of connections is in a loop make it even harder to track down his origin in a wide SCADA system.

[6] US-Cert VU #362332

2.6.2 Transport Layer

SYN flood is to saturate resources by sending TCP connection requests faster than a machine can process.

SCADA protocols, particularly those running over top of transport protocols such as TCP/IP have vulnerabilities that could be exploited by attacker through methodologies as simple as injecting malformed packets to cause the receiving device to respond or communicate in inappropriate ways and result in the operator losing complete view or control of the control device.

2.6.3 Application Layer

Currently, there is no strong security control in protocols used in SCADA systems, such as DNP3 without secure authentication, Modbus,*Object Linking and Embedding (OLE) for Process Control* (OPC), *Inter-Control Center Communications Protocol* (ICCP). Practically there is no authentication on source and data such that for those who have access to a device through a SCADA protocol, they can often read and write as well. The write access and diagnostic functions of these protocols are particular vulnerable to cyber and cyber induced physical attacks.

One of possible attacks in both SCADA and conventional IT systems is *DNS forgery*. Such attack is to send a fake DNS reply with a matching source IP, destination port, request ID, but with an attacker manipulated information inside, so that this fake reply may be processed by the client before the real reply is received from the real DNS server. For more details on those attacks studied in conventional IT systems, please refer to [104].

Next, we list potential attacks associated with more SCADA specific protocols.

MODBUS

Modbus [187] is a *de facto* standard of application layer protocol used in industrial networks. It comes with different flavors from plain Modbus to Modbus+ to Modbus/TCP. A Modbus client (or master) can send a request to a Modbus server (or slave)[7] with a *function code* that specifies the action to be taken and a *data field* that provides the additional information. The general Modbus frame is shown in Figure (2.3).

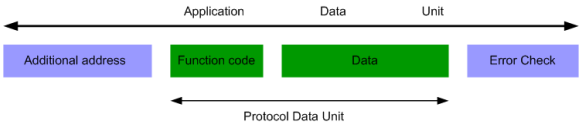


Figure 2.3: A typical Modbus frame

Among currently little published accounts on attacks against Modbus, Digital Bond [210] has conducted intrusion detection work on studying its potential weakness. Their detection rules

[7]Initially, Modbus was a master-slave protocol for serial buses. When implementing Modbus over TCP, a Modbus master is a TCP client, and a Modbus slave is a TCP server.

include denial of service (e.g., rebooting Modbus servers, configuring them to provide no servicecalled listen-only mode, and crashing servers with a large size request), reconnaissance (e.g., unauthorized reading of data, and gathering device information), and unauthorized write requests.

Byres and his company have used Achilles Vulnerability Test Platform to perform security tests on Modbus to discover vulnerabilities [42, 43] .

Given that Modbus does not have encryption or any other security measures, there are many ways to directly explore such weakness on the function code level. The function codes 0x05 and 0x0F are used to write a single or multiple outputs (coils) to either ON or OFF in a remote device, respectively. This means that an adversary can turn off and suppress output(s) remotely thus to create a false sense of situation at the HMI end. Unauthorized writes can be accomplished through using function codes 0x06 and 0x10. Accordingly, the forged data may be written to either a single or multiple registers in a remote device. If Modbus is implemented on serial line, function code 0x11 can be used to gather information from a remote device, such as a controller's description. Function code 0x08 is used for diagnostics on serial line. However, combined with subfunction code 0x01, it can initialize and restart the slave (server) port and clear out the communication event counter, which is a ideal attack vector. When combined with subfunction code 0x04, the diagnostics function code can force a remote device into its Listen Only Mode. Similarly, Modbus+ has a function code (08) for log cleaning that can enable an attacker to clear stats of data manipulation and denial of service events.

DNP3

DNP3 is used between master control stations and remote computers or controllers called *outstations* for the electric utility industry and water companies. DNP3 is implemented by several manufacturers due to its small memory consumption. Its function code 0x0D can reset and reconfigure DNP3 outstations by forcing them to perform complete power cycle. During the re-initialization to default values, many devices clear all queues as well. An attacker can take advantage of this property to cause delay in outstations before they accept requests again. Furthermore, function code 0x13 enable loading new outstation configurations. With unauthorized access, an attacker can manipulate the remote devices with manipulated setting values, suppress output and or create false alarms.

2.6.4 Attacks on Implementation of Protocols

Protocol vulnerabilities can reveal themselves as segmentation faults, stack, heap or buffer overflows, etc., all of which can cause the protocol implementation to fail resulting in a potential exploit.

Meanwhile, certain protocol implementations, such as ICCP servers, only allow users to read values, and there are a number of protocols that are in the process of adding security controls to address this deficiency.

Nevertheless, [210] argues that SCADA implementation vulnerabilities are more important than lack of security controls in SCADA protocols.

TCP/IP

First of all, in light of the migration to Windows from UNIX in operating system used by many sectors in SCADA systems, there are several attacks specifically exploit the implementation of TCP/IP protocols in Windows. Although there are patches available, restrained to be on-line continuously, it's very likely that these machines do not have up-to-dated patches. Here, we only name a few well known ones.

- WinNuke takes advantage of the absence of status flag URG in handling the TCP protocol.
- TearDrop/NearTear and Ssping utilize implementation error of fragmentation handling in TCP/IP protocol.

A nightmare scenario can be that one company's network is compromised and a polymorphic worm takes down most servers and any unpatched SCADA servers running Windows.

Secondly, these protocol stacks can and do suffer from various vulnerabilities commonly found due to poor software design and coding practices.

OPC

OPC servers use Microsoft's OLE technology[8] to provide real-time information exchange between software applications and process hardware.

At the OPC interface level, the item write function takes two parameters: an item handle and a value to write to it. If the server maps handles to memory addresses and fails to validate a client-provided handle, the IO interfaces write function allows an attacker to write any value to any memory address, a primitive which can be easily exploited to run arbitrary code on the server (e.g. through stack return addresses). It is an even larger issue that an OPC server can be remotely compromised and used to launch attacks on other systems. Because OPC servers are often exposed in the Demilitarized Zone (DMZ), this could be a communication chain that could allow control system exploitation from the enterprise network or Internet.

[27] gives three possible OPC attack scenarios, of which are all associated with extra open ports:

- Collateral Damage by OPC-Unaware Malware;
- Opportunistic OPC Denial of Service Attack;
- Intelligent, aggressive attack against OPC hosts through a man-in-the-middle (MITM) technique

ICCP

The most serious and exposed SCADA protocol stacks are those that are used to exchange information with business partners, such as ICCP, or those used to exchange information between the corporate network and control center network.

[8] Also known as the Component Object Model, or COM

According to the LiveData ICCP Server white paper [268], LiveData ICCP server contains a heap-based buffer overflow. The LiveData implementation of ISO Transport Service over TCP (RFC 1006) is vulnerable to a heap-based buffer overflow. By sending a specially crafted packet to a vulnerable LiveData RFC 1006 implementation, a remote attacker may be able to trigger the overflow to execute arbitrary code or crash a LiveData ICCP Server to cause a denial of service.

UCA

UCA was expected to be more robust standard than DNP3 when the Electric Power Research Institute (EPRI) decided to use it to serve the SCADA needs of the electric utilities. It's based on the Manufacturing Message Specification from ISO standard 9506.

MMS

Tamarack MMS*d* is an implementation of *Manufacturing Message Specification* (MMS) protocol, an international standard (ISO 9506), dealing with messaging system for transferring real time process data and supervisory control information between networked field devices and/or computer applications.

Tamarack MMS*d*[9] components do not properly handle malformed RFC 1006 packets either. This vulnerability may allow a remote, unauthenticated attacker to cause a denial of service condition.

2.7 Discussion

The cyber-physical security of real-time, continuous systems necessitates a comprehensive view and holistic understanding of network security, control theory and the physical system. Ultimately, any viable technical solutions and research directions in securing SCADA systems must lie in the conjunction of computer security, communication network and control engineering. The idea of looking into the problem in the context of control performance holds its solid bearings. However, the very large installed base of such systems means that in many instances we must for a long time to come rely on retrofitted security mechanisms, rather than having the option to design them in from scratch. This leads to a pressing need for robust SCADA-specific intrusion detection systems (IDS) and resilient control.

Our next step is to categorize the attacks in terms of their manifestation and realization in order to shed more light into intrusion prevention and detection.

[9] Vulnerability Note VU#372878

Chapter 3

SCADA-specific Intrusion Detection/Prevention Systems: A Survey and Taxonomy

Due to standardization and connectivity to the Internet, Supervisory Control and Data Acquisition (SCADA) systems now face the threat of cyber attacks. SCADA systems were designed without cyber security in mind and hence the problem of how to modify conventional Information Technology (IT) intrusion detection techniques to suit the needs of SCADA is a big challenge. We explain the nuance associated with the task of SCADA-specific intrusion detection and frame it in the domain interest of control's researchers to illuminate problem space. We present a taxonomy and a set of metrics for SCADA-specific intrusion detection techniques by heightening their possible use in SCADA systems. In particular, we enumerate Intrusion Detection Systems (IDS) that have been proposed to undertake this endeavor. We draw upon the discussion to identify the deficits and voids in current research. Finally, we offer recommendations and future research venues based upon our taxonomy and analysis on which SCADA-specific IDS strategies are most likely to succeed, in part through presenting a prototype of our efforts towards this goal.

Defined by IEEE Standard (C37.1-1994) [45] , a Supervisory Control and Data Acquisition (SCADA) system includes all control, indication, and associated telemetering equipment at the master station, and all of the complementary devices at the (Remote Terminal Unit) RTU(s)[1]. A typical SCADA system includes hardware, software and communication protocols that connect together the different layers in the hierarchy. For more detailed exposition of SCADA system compositions, readers please refer to resources such as [256, 153]

Being one of the primary categories of control systems, SCADA systems are generally used for large, geographically dispersed distribution operations, such as electrical power grids, petroleum and gas pipelines, water and wastewater (sewage) systems and other critical infrastructures [256]. They not only provide management with remote access to real-time data from Distributed Control Systems (DCSs) and Programmable Logic Controllers (PLCs) but also enable operational con-

[1]RTUs are special purpose data acquisition and control units designed to support SCADA remote stations. These field devices are often equipped with wireless radio interfaces to support remote situations where wire based communications are unavailable.

trol center to issue automated or operator-driven supervisory commands to remote station control devices.

One of the enabling elements in SCADA systems is the set of various communication protocols employed within the hierarchical system [12, 64, 153]. Their functionalities range from processing raw data transmission to handling high-level exchange between different networks and domains. These protocols have strong implications on the security of SCADA system. We name a few most popular ones: Modbus, Profibus, Distributed Network Protocol (DNP3) and Utility Communications Architecture (UCA), Foundation Fieldbus, Common Industrial Protocol (CIP), Controller Area Network(CAN), Object Linking and Embedding (OLE) for Process Control (OPC) and Inter-Control Center Communications Protocol [153].

Most industrial plants now employ networked process historian servers storing process data and other possible business and process interfaces, such as using remote Windows sessions to DCSs or direct file transfer from PLCs to spreadsheets. This integration of SCADA networks with other networks has made SCADA vulnerable to various cyber threats. The adoption of Ethernet and TCP/IP for process control networks and wireless technologies such as IEEE 802.x, Zigbee, Bluetooth, WiFi, plus WirelessHART and ISA SP100 [64, 153] has further reduced the isolation of SCADA networks. The connectivity and de-isolation of the SCADA system is manifested in Fig.2.1.

Furthermore, the recent trend in standardization of software and hardware used in SCADA systems [153] potentially makes it even easier to mount SCADA-specific attacks[2]. These attacks can disrupt and damage critical infrastructural operations, contaminate the ecological environment, cause major economic losses and, even more dangerously, claim human lives [90, 5, 81]. These likely "penalty costs" due to lack of protection and *aversion to loss* [138, 267, 242] push us to consider seeking protection measures with reasonable cost-effectiveness [196].

3.0.1 Why SCADA-specific Intrusion Detection Systems?

Had we not started with the legacy systems but been freed from difficulties such as interoperability [161, 204] instead, we may apply and implement many known security measures directly. Among them, a sound implementation and viable deployment of one Intrusion Detection System (IDS) can manifest itself as an add-on intelligence component to the existing SCADA systems with minimum hardware cost or operational changes, leveraging many entrenched SCADA component infrastructures and technologies.

To this end, the industrial and academic control security community has started to build Intrusion Detection Systems (IDS) specifically for SCADA systems ([49, 191, 195, 204, 230, 233, 262, 263, 287]).

Nevertheless, it is important to realize that when we borrow tools from other fields, there are situations and conditions that our original set of assumptions might not hold. A SCADA system is different from the conventional IT system in the following ways [256]: it is a **hard real-time** system; its **timeliness** and **availability** at all times is very critical and its terminal devices have limited computing capabilities and memory resources [59]. Additionally, in the existing SCADA systems, there are weak authentication mechanisms to differentiate human users or privilege separation or

[2]In the paper, we interchange the use of intrusion and attack equivalently.

user account management to control access and so on [204]. Such fundamental weakness in access control leaves door open to attacks. These differences challenge design and implementation of SCADA-specific IDSs.

Meanwhile, among the attempts to date, some authors [49] may consider that SCADA systems usually have a relatively static topology[3], a *presumably* regular network traffic[4] and use simple protocols, hence monitoring them may not be more difficult than doing so in enterprise systems. But such assumptions are not fully validated yet as barely any mentioned work has been tested on real operational SCADA system network traffic. The related details are to be discussed in subsequent sections.

Furthermore, the cyber-physical security of real-time, continuous systems necessitates a comprehensive view and holistic understanding of network security, control theory and physical systems. The focus and terminologies by convention in each field have partial overlaps and their own field-specific interpretations for these overlapped lingoes. One of the barriers faced by researchers in IDS for SCADA is the occupational or cultural and lingo differences between IT and control personnel. Thus this paper aims to convey the idea of intrusion detection and prevention in the setting a SCADA system by leveraging the classic control engineering and theory view point.

The ultimate goal of much needed work in this area is to achieve satisfactory control performance in a continuous 24×7, real-time, realistic environment, where normalized behavior co-exists with benign noises, honest mistakes, natural components and or systems faults plus potential malicious cyber intrusions.

Towards concrete progress beyond generic discussions, it's important for us to survey and evaluate up-to-date research efforts in this area and reflect on the soundness of the overall methodologies. We may want to ask:

- Whether these techniques and approaches have addressed the specifical needs of SCADA systems ? Furthermore,
- Whether we are being simply handicapped by the special needs of current SCADA systems in terms of security engineering efforts? Or
- Whether we are leveraging the entrenched SCADA infrastructure components and technologies?

3.0.2 Contribution

In this paper, we make the following contributions:

- First systematic and thorough effort in investigating and assessing the landscape of up-to-date SCADA-specific intrusion detection techniques and systems;
- Explain the nuance of SCADA-specific IDS and provide clear definitions plus a taxonomy and a set of metrics of SCADA-specific IDS;

[3]Under the assumption that there is no wireless sensor network involved.

[4]Due to the scarce accessibility to operational SCADA traces known to the public, we are conservative at taking the leap of faith yet.

- Ease the interoperability between conventional IT security and control systems research by framing the intrusion detection problem in a setting favorable to SCADA systems' continuous operation, withstanding the possible presence of adversary and unintentional faults;
- Bring in cross-discipline insights to tailor the special needs entailed by SCADA systems by leveraging entrenched SCADA components and technologies and provide future direction;
- Show a prototype of our efforts in this arena.

3.0.3 Definitions and Difficulties from Ambiguities

To resolve the ambiguity of same terminologies that bear different meanings in control theory (including systems & control and fault detection & isolation) and IT (particularly, operating system and security engineering), we intend to unify the terms to ease the misunderstanding and highlight the end goal of providing engineers and researchers insights into the problems facing networked control systems [304].

Fault: a non-hostility-induced deviation from the system's specified behavior including honest mistakes caused by honest people and component failures or defects.

Anomaly: refers to malicious and intrusive event plus abnormal yet non-intrusive behavior including (faulty and noisy/messy) actions;

Misuse: includes both malicious and unintentional misuse;

Detection: alarm alerts issued in the presence of true anomaly or misuse.

False alarm/positive: alarm alerts issued in the absence of real **anomaly** and/or **misuse** when there is normal traffic/behavior only..

False negative or missed detection: missed detection in the presence of a real intrusion.

Note: Any large network is a very "noisy" environment even at the packet level.

3.0.4 Related Work

Since SCADA-specific IDS research is a rather new arena, we decide to resort to the classics in the standard IT field for references.

As observed by John McHugh in [176]

> The point is that the taxonomy must be constructed with two objectives in mind: describing the relevant universe and applying the description to gain insight into the problem at hand.

Both Stefan Axelsson [15] and John Mchugh [177] have thorough work on classification of intrusion detection systems. Many evaluation and assessment principles on SCADA-specific IDS in this paper are derived from their works.

The unified view is to consider intrusion detection as a signal detection problem as framed by Stefan Axelsson [16], where we consider the normal network traffic as background data. If we view background data and responses as noise and attack data and responses as signal, the IDS problem can be characterized as one of detecting a signal in the presence of noise. This school of thought is much in line with the standard control theory [46].

3.1 On Real Time Intrusion Detection Types

We adapt a taxonomy of real-time intrusion detection to facilitate the choice for control's researchers as well.

In the early days of IDS research, two major approaches known as **signature detection** and **anomaly detection** were developed.

In between these two approaches, there lie the probabilistic- and specification-based methods for intrusion detection. A **probabilistic approach** is also termed as a *statistical* or a *Bayes* method [152] with probabilistically encoded models of misuse. It has some potential to detect unknown attacks. A **specification-based approach** constructs a model of what is allowed, enforces its predefined policy and raises alerts when the observed behavior is outside this model. It has a high potential for generalization and leverages against new attacks [20]. This technique has been proposed as a promising alternative that combines the strengths of signature-based and anomaly-based detection.

Instead of finding the deviation and unknowns, specification-based method [20, 148] defines what's allowable in terms of network traffic behavior/patterns. This method sounds promising. But it might be tedious to enumerate all possibly allowable patterns.

Complementary to the above knowledge based classification, there are also **behavioral detection** approaches[5]. They capture behavior patterns associated with certain attacks which are not necessarily illegitimate in semantic sense. They may also abstract allowable normal interaction as well. Such methods are quite promising, especially in conjunction with other methods [290].

Table 3.1 gives the overall comparison.

Knowledge based or *behavioral based*	*Approach*	*Basis*	*Attacks Detected*	*Generalization*
Knowledge	Signature	Misuse	Known	No
Knowledge	Anomaly	Learned models of normal	Must appear anomalous	Yes
Knowledge	Probabilistic	Model learning	Match patterns of misuse	Some
Hybrid	Specification	Construct normal model	Must violate specs	Yes
Behavioral	Behavioral	Capture behavioral pattern	Match patters of behavior	Yes

Table 3.1: Comparison of Intrusion Detection System Approaches

[5] A thoroughly stringent and meticulous categorization is not the focus of this paper. Interested readers may refer to [15, 177] for more detailed taxonomies on IDS

3.2 Proposed SCADA-specific Intrusion Detection/Prevention Systems

3.2.1 Model-Based IDS for SCADA Using Modbus/TCP

The group at SRI [49] adapted the specification-based approach for intrusion detection to SCADA systems that rely on Modbus/TCP. This work renders a multi-algorithm IDS appliance containing pattern anomaly recognition, Bayes analysis of TCP headers, and stateful protocol monitoring complemented with customed Snort rules. Alerts are forwarded to the correlation framework.

They offer three model-based techniques to characterize the expected/acceptable system behavior according to the Modbus/TCP specification and to detect potential attacks that violate these models.

3.2.2 Anomaly-Based Intrusion Detection

We discuss two anomaly-based intrusion detection systems in this section.

AutoAssociative Kernel Regression and Statistical Probability Ratio test SPRT

Yang et al [287] use the AutoAssociative Kernel Regression (AAKR) model coupled with the Statistical Probability Ratio test (SPRT) and apply them to a simulated SCADA system.

The fundamental methodology is pattern matching. Predetermined features representing network traffic and hardware operating statistics are used by the AAKR model to predict the "correct" behavior. Then new observations are compared with past observations denoted as normal behavior. The comparison residuals are fed into SPRT to determine whether is anomalous or not.

Besides DoS attacks, ping flood, jolt2 attacks, bubonic attacks, simultaneous jolt2 and bubonic attacks, the authors also consider insider attack scenarios.

Multi-Agent IDS Using Ant Clustering Approach and Unsupervised Feature Extraction

Tsang and Kwong [262] propose an unsupervised anomaly-learning model - the Ant Colony Clustering Model (ACCM) in a multi-agent, decentralized IDS to reduce data dimensionality and increase modeling accuracy. The idea is bio-inspired from nature to construct statistical patterns of network data into near-optimal clusters for classification.

3.2.3 Configurable Middleware-Level Detection

Næss et al [195] presents a configurable Embedded Middleware-level Intrusion Detection System (EMISDS) framework. It's implemented within *MicroQoSCORBA*, a CORBA-based middleware framework, with high configurability achieved with the Interface Definition Language (IDL) compiler and code generation tools [178].

The system model is comprised of anomaly and misuse detection while leaving the flexibility to specify the interaction of middle-level information within the IDS.

3.2.4 Intrusion Detection and Event Monitoring in SCADA Networks

Oman and Phillips [204] from the University of Idaho give a very clear exposition on the implementation of a SCADA power-grid testbed for intrusion detection and event monitoring. They are producing comprehensive intrusion signatures for unauthorized access to SCADA devices besides baseline-setting files for those devices.

3.2.5 Model for Cyber-Physical Interaction

Power Plant interfacing Substations through Probabilistic validation of attack-effect bindings (PVAEB)

Rrushi and Campbell [233] look into the attacks on IEC 61850 [126], the protocol used for communication between electricity substation and power plant (a nuclear power plant is referred).

The authors present the semantic correlation between the dynamics of nuclear reactors in the power plant and those of the generated electricity provision in the substation through structural equations modeling (SEM). For each logical node of IEC 61850, they apply Bayesian Belief Networks (BBN) to enumerate probability distributions attributed by its associated data individually. Then the authors use Stochastic Activity Network (SAN) to verify such bindings and to spot intrusions.

All construction of attack-effects are based on *known* failure models.

Workflow-based non-intrusive approach for enhancing the survivability of critical infrastructures in Cyber Environment

Xiao et al [282] proposed an approach based on workflow, a technique to automate existing processes to incorporate the detections of both *known* attack patterns and *known* unsafe states.

This work leverages the *presumably* existing survivability-related knowledge and protection scheme. They consider that each essential component in the physical layer has a corresponding node in the workflow.

A simplified water treatment system is studied through simulation to illustrate the idea.

3.3 Comparison

The overall comparisons of the proposed systems are listed in Table 3.2 and Table 3.3. The rationale behind choosing the features we used for comparison is out of operational concerns besides performance issues.

3.3.1 Intrusion Detection

Particularly, we'd like to look into the intrusion detection methods used in each system, seen in Table 3.4

Name of System	Publ. year	Degree of SCADA Specific	Specific Domain	Detection Prevention Principle	Malicious Intrusions only?	Threat model	Time of Detection	Security	Fallacy Analysis	Unit of analysis
PVAEB [233]	2008	high	electrical power	proba.	fault & intrusion	no	N/A	low	no	packet
IBM NADS [191]	2008	medium	N/A	anomaly, spec, behavioral	extensible	outsider not explicit	Non-real	low	no	flow--based
SRI Modbus [49]	2007	high	N/A	spec. proba.	extensible	outsider	real	medium	no	packet
WFBNI [282]	2007	high	water treatment system	signature	unintent. faults unsafe states	not explicit	on-line prediction	low	no	N/A
SHARP [230]	2008	medium	N/A	spec. encryp.	extensible	insider or outsider	on-line	high	no	N/A
IDEM [204]	2007	high	power grid	signature	yes	unauth. access	real	low	N/A	packet
AAKR--SPRT [287]	2006	high	no	anomaly	yes	insider & outsider	real	low	no	packet
EMISDS [195]	2005	low	N/A	anomaly, spec., signature	yes	N/A	real	low	no	procedural interval
MAAC--UFE [262]	2004	medium	N/A	anomaly	yes	both	real	N/A	yes	N/A

Table 3.2: Comparison of Intrusion Detection System Approaches

Name of System	Data Proc.	Data Coll.	Scalab--ility	Granul-arity	Audit Source	Type of Response	Inter-oper.	Imple-ment.	Deploy. ment	Real traces
PVAEB [233]	centr.	centr.	medium	batch	host	passive	N/A	yes	no	testbed
IBM NADS [191]	centr.	dist.	high	cont.	network	passive	yes	yes	N/A	N/A
SRI Modbus [49]	dist.	dist	high	cont.	both	active	yes	yes	no	testbed
WFBNI [282]	centr.	dist.	high	cont.	network	passive	maybe	yes	no	simulation
SHARP [230]	centr.	centr.	low	cont.	network	active	yes	no	N/A	N/A
IDEM [204]	centr.	centr.	low	cont.	network	passive	yes	yes	no	testbed
AAKRSPRT[287]	centr.	centr.	low	cont.	host	passive	yes	yes	no	testbed
EMISDS [195]	dist.	dist.	high	batch.	both	N/A	N/A	no	no	simulation w/o intrusion
MAACUFE [262]	dist.	dist.	high	N/A	both	active	N/A	yes	no	KDD-cup

Table 3.3: Comparison of Intrusion Detection System Approaches: Contd.

Name of System	Detection Type	Intrusion only	Detection Method / Algorithm
PVAEB [233]	anomaly	fault intrusion	Structural Equation Modeling, Bayesian Belief Networks, Stochastic Activity Networks
IBM NADS [191]	anomaly, behavioral specification	N/A	net flow matching
SRI Modbus [49]	spec., prob.	extensible	descriptive statistics, simple rule based
WFBNI [282]	signature	fault intrusion	matching fault model
SHARP [230]	spec.	extensible	N/A
IDEM [204]	signature	yes	N/A
AAKRSPRT[287]	anomaly	yes	AAKR, SPRT, pattern matching
EMISDS [195]	anomaly, spec. signature	yes	simple rule based, sliding window
MAACUFE [262]	anomaly	yes	ACCM, PCA

Table 3.4: Comparison of Intrusion Detection Method in Each Proposed System

3.3.2 SCADA-Specific-ness

We compare how SCADA's special needs are being addressed in each proposed system with results shown in Table 3.5

3.4 Evaluation

3.4.1 Design Pitfalls and Evaluation Criteria

Looking at IT standard IDSs, McHugh [176] criticizes many aspects of the DARPA/LL evaluation. In terms of modeling, both signature and probabilistic IDSs model misuse, the *illegal* behavior of an intrusion. Anomaly-based IDSs empirically and statistically model normal system usage and behavior. Specification-based IDSs define what is allowable under protocol and policy specification. All these model-based approaches bear certain common drawbacks:

- Inaccurate models can lead to false alarms and/or missed detections.
- Modeling can be expensive and difficult if the system and/or user activity is complex.

Anderson states [12] "In general, if you build an intrusion detection system based on data-mining techniques, you are at serious risk of discriminating."

Paxson has a similar argument, even more from a technical point of view [208] that one of the pitfalls of machining learning based IDS techniques is the lack of illumination for the rationale behind many approaches on how they decide to take such approach; and why they succeed in doing so or why they fail in achieving.

According to Axelsson [15], McHugh [177] and Paxson [208], we shall look for

- soundness
- completeness
- timeliness
- choice of metrics, statistical models, profiles
- system design;
- social implications
- feedback: or how to decide actionable events

The SCADA-specific angles we look at are: What are their contributions, limitations or room for improvement, extensibleness in terms of

- How do they frame the work including assumptions, logics and conclusions?
- What kind of security properties do they want to achieve? Do they achieve and how?

Name of System	Security Properties			Inter. oppp	Use of SCADA Components					Interaction between Cyber – Physical
	Time--liness	Availability			Domain/ Industry	HW	SW	communication		
		Self Security	Type Response					hardware	protocol	
PVAEB [233]		low	passive	N/A	electrical power			simulated IED	IEC 61850 DNP3	yes
IBM		low	passive	yes					Modbus	
NADS [191]										
SRI Modbus [49]		medium	passive	yes	N/A				Modbus	
WFBNI [282]		low	passive	N/A	water					yes
SHARP [230]		high	active	yes	N/A					
IDEM [204]		low	passive	yes	electrical power	yes				
AAKRSPRT [287]		low	passive	yes	N/A				SNMP	
EMISDS [195]	yes	low	passive	N/A	N/A					
MAACUFE [262]		N/A	active	N/A	N/A		yes			

Table 3.5: Comparison of SCADA's Special Needs Being Addressed in Each Proposed System

- What are their trust model, threat model and attack scenarios? How plausible?
- What are the illuminations they bring into the problem space;
- What's the selling point of their approach?
- What kind of detection algorithms they've used that suit SCADA systems particularly well
    1. either through leveraging the entrenched components and/or technologies used in the specific SCADA physical systems under their study;
    2. or restrict their attention to a more focused and potentially narrowed workspace that are more relevant to specific SCADA physical system under their study when applying generic methods.
- What are the subtle points they bring out that might have been simply left out by a non-SCADA-security expert?
- What's unique in the cyber-physical interactions?
- How is the detection performance in terms effectiveness and efficiency? Effectiveness is reflected through high detection rate and low false alarm rate; efficiency overheads.

3.4.2 Evaluation Results

Strength

Intrusion detection research for SCADA systems to date has been quite limited, with the three most prominent and critical deficiencies being

- the lack of a well-considered threat model;
- the absence of addressing false alarm and false negative (mis-detection) rates; and
- the need to empirically ground the development of IDS mechanisms in the realities of how such systems operate in practice, including the diversity of traffic they manifest and the need to tailor IDS operation to different SCADA environments.

From the above evaluation of existing IDSs for SCADA systems, we can see that the current bottleneck problems faced by research and design henceforth implementation and deployment of IDS for SCADA are the scarcity in access to operational SCADA system (network traffic) traces and the lack of prudent yet novel threat models, or attack scenarios.

Barely any of these systems has a performance evaluation on the false alarms that it generates. However, given the availability demand of SCADA systems, we believe this is an issue that must be addressed well before IDS can be implemented and deployed in SCADA systems at large scale.

3.5 Future Directions

Ultimately, any viable technical solutions and research directions in securing SCADA systems must lie in the conjunction of computer security, communication network and control engineering. However, the very large installed base of such systems means that in many instances we must for a long time to come rely on retrofitted security mechanisms, rather than having the option to design them in from scratch. This leads to a pressing need for deployable, robust, SCADA-specific intrusion detection systems (IDS).

We shall aim to capture the characteristics of a specific SCADA system under study with full situational awareness, including the dynamics of the physical plant being monitored, its communication patterns, system architecture, network traffic behavior, and specific application-level protocols used.

3.5.1 Our Future Work

We propose a JIE [6], a *viable* intrusion detection and self-hardening system for SCADA system.

In terms of the functionalities of intrusion detection and prevention, our proposed JIE would be able to

- efficiently detect and block cyber intrusions into SCADA systems in real operational environments, and in real-time,
- without interrupting the control performance of the protected system,
- without creating extra operational burdens or operational reservations due to false alarms,
- in the presence of both malicious and messily benign network traffic. The system must operate in a real-time, robust fashion, with performance adequate to meet the demands of the dynamic cyber-physical interactions inherent to SCADA systems.

3.6 Discussion

As argued by Rakaczky [224], the ease of deployment requires the intrusion detection/prevention strategy to minimize the associated personnel overhead.

The model-based system for SCADA system using Modubs/TCP addresses Modbus protocol encapsulated within TCP/IP. The idea can be generalized to other control system protocols as well.

Since SCADA networks are built of resource-constrained embedded systems, the IDS using the middleware-level detection has the advantage of directly accessing message signatures and

[6] This is the 40th hexagram of *I Ching*, or, *Yi Jing*, *The Book of Changes*, comprising of 64 hexagrams plus their commentaries and transformations as strategic interpretation of chance event. It literally means *Problem Solving* or *Deliverance*. The essence of this strategy is: Don't trouble troubles until trouble troubles you; If it does, then act quick.

parameter values without decoding the raw network packets. But there is a tradeoff in the risk involved in handling embedded responses to attacks.

Both model-based intrusion detection and middleware-level intrusion detection build models to specify the normal behavior of the network traffic and compare the SCADA traffic against these models to detect potential anomalous behavior. Model-based detection is an important complement to signature-based approaches.

The specification-based IDS has an inviting advantage to SCADA systems and networked control systems in general.

Chapter 4

Xware – an Overall Architecture of a SCADA-specific Security Solution

> Security is a process, not a product.
>
> BRUCE SCHNEIER

A SCADA-specific defense-in-depth security engineering solution framework: **Xware** as shown in figure. 4.1 is presented in this chapter.

This system tailors the special needs entailed by SCADA systems through leveraging the entrenched SCADA components and technologies. It provides reliable performance in the face of malicious intrusion, unintentional faults, honest mistakes, benign noise, extreme cases besides predefined allowable behavior thus very low in both false positive and false negative rates. We give an overview of the system's design with emphasis on prudent threat model. *Xware* is comprised of two strong footings – *Normalcy Checking*, a control theoretic, domain knowledge specific, specification-based payload inspection system and a high-speed, real-time, behavioral-based NIDS (Network Intrusion Detection System). *Xware* integrates a *Trust Counter* to verify the truthfulness of sensor measurements. It also provides exfiltration of confidential information from within the intranet. Moreover, *Xware* hardens SCADA system with compensation schemes when intrusion evades NIDS or unexpected fault occurs to guarantee its performance. It puts things in perceptive and highlights the overall systematic and holistic approach.

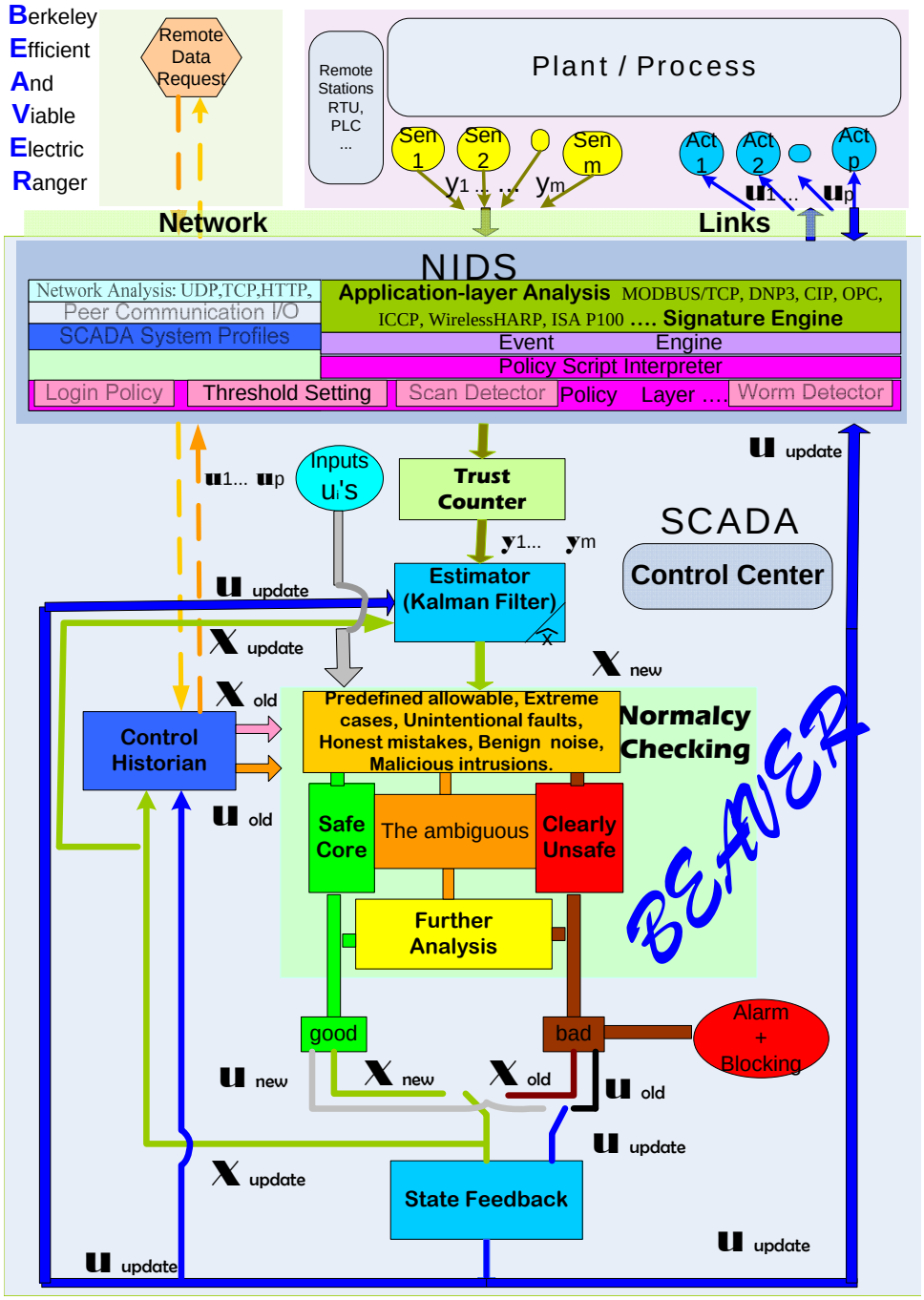


Figure 4.1: ***Xware***: the overal architecture of a SCADA-specific Security Solution

Chapter 5

Trust Counter –Data Fusion Assurance for the Kalman Filter in Uncertain Networks

Trust is cheaper than control

JON MELL

This chapter depicts *Trust Counter*, an important component of the proposed *Xware* that measures trustworthiness of each sensor reading before fusing them in an estimation-performance-centric way and feeding it to a central location.

Due to standardization and connectivity to other networks, networked control systems, a vital component of many nations' critical infrastructures, face potential disruption. Its possible manifestation can affect the Kalman filter, the primary recursive estimation method used in the control engineering field. Whereas, to improve such estimation, data fusion may take place at a central location to fuse and process multiple sensor measurements delivered over the network. In an uncertain networked control system where the nodes and links are subject to attacks, false or compromised or missing individual readings can produce skewed results. To assure the validity of data fusion, this paper proposes a centralized trust rating system that evaluates the trustworthiness of each sensor reading on top of the fusion mechanism. The ratings are represented by Beta distribution, the conjugate prior of the binomial distribution and its posterior. Then an illustrative example demonstrates its efficiency.

Control systems[1] are deeply ingrained in the fabric of critical infrastructure sectors including power grids; oil and gas pipeline systems; water treatment and distribution; railroads and mass transit; and widely involved in the constitutions of vital enterprises such as manufacturing plants and building climate control [79].

Most industrial plants now employ networked process historian servers storing process data plus other possible business and process interfaces[2]. This integration of networked control systems with other networks has made control systems vulnerable to various cyber threats. The adoption of

[1] Control Systems are computer-based systems that are used in many industries to monitor and control sensitive processes and physical functions [79].

[2] For example, using remote Windows sessions to Distributed Control Systems or direct file transfer from Program Logic Controllers to spreadsheets.

Ethernet and TCP/IP for process control networks and wireless technologies such as IEEE 802.x, Zigbee, Bluetooth, WiFi [64, 153] and so on has further reduced the isolation of control networks. The connectivity and de-isolation of a control system is manifested in Fig.**??**. Furthermore, the recent trend in standardization of software and hardware used in control systems makes it possible to mount control specific attacks. The continuous availability, hard deadline, legacy issues and low computation power of the end devices are among the things that have been keeping ready security measures from immediate implementation and deployment.

Such uncertainty may potentially affect the performance of networked control systems. Specifically, we address its likely manifested impact on the Kalman filter based estimation, a key functionality of control systems, and propose a possible countermeasure.

Typically, a central location collects measurements from multiple sensors to achieve higher accuracy in estimation as shown in Fig 5.1.

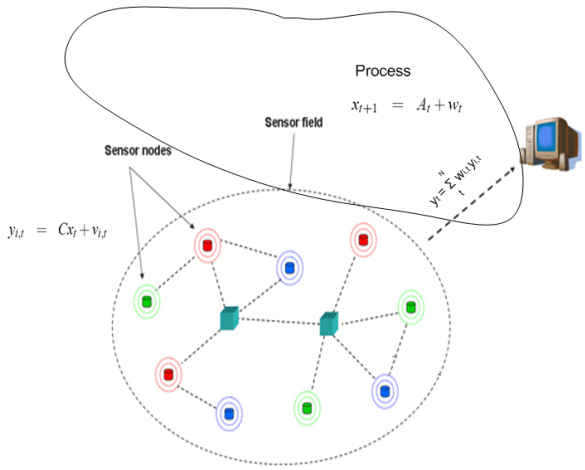


Figure 5.1: An Example of Centralized Data Fusion for Networked Control Systems

The discrete time linear dynamical system and measurement model are the following, where i is the index of sensors.

$$x_{t+1} = Ax_t + w_t \tag{5.1}$$
$$y_{i,t} = C_i x_t + v_{i,t} \tag{5.2}$$

where $x_t \in \Re^n$ is the state vector, $y_t \in \Re^m$ is the output vector, $w_t \in \Re^p$ is white Gaussian noise with zero mean and covariance $Q > 0$ and $v_{i,t}$'s $\in \Re^m$ are white Gaussian noises with covariance $R_i > 0$. w_t and $v_{i,t}$'s are independent. The initial system state x_0 is Gaussian with zero mean and covariance $\sum_0$. We assume x_0 is independent of w_t and $v_{i,t}$'s.

Then individual measurements $y_{i,t}$ undergo fusion before feeding into the Kalman filter, which will be further discussed in later sections.

Furthermore, we shall briefly recap the standard Kalman filtering algorithm and the Kalman filter based data fusion methods in a theoretically benign setting plus mention two well known examples of trust rating systems in practice, dealing with potential malicious situations.

5.0.1 Standard Kalman Filter

$$\begin{aligned} P_{t|t} &= \mathbb{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^{'}|\mathbf{y_t}] \\ \hat{x}_{t+1|t} &= \mathbb{E}[x_{t+1}|\mathbf{y_t}] \\ P_{t+1|t} &= \mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^{'}|\mathbf{y_t}] \\ \hat{y}_{t+1|t} &= \mathbb{E}[y_{t+1}|\mathbf{y_t}]. \end{aligned}$$

The prediction phase for $\hat{x}_{t+1|t}$ and $P_{t+1|t}$ of the Kalman filter is independent of the observation process with:

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} \tag{5.3}$$

$$P_{t+1|t} = AP_{t|t}A^{'} + Q \tag{5.4}$$

For the update phase of the Kalman filter, we have

$$\begin{aligned} \hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + P_{t+1|t}C^{'}(CP_{t+1|t}C^{'} + R)^{-1} \\ &\quad (y_{t+1} - C\hat{x}_{t+1|t}) \qquad (5.5) \\ P_{t+1|t+1} &= AP_tA^{'} + Q - P_{t+1|t}C^{'}(CP_{t+1|t}C^{'} + R)^{-1} \\ &\quad CP_{t+1|t} \qquad (5.6) \end{aligned}$$

The accuracy of measurement improves as more sensors collaborate. Naturally, this leads to the question of how to fuse data from multiple sensors.

5.0.2 Data Fusion

The two most commonly used methods for the Kalman filter based data fusion are state-vector fusion and measurement fusion [76]. State-vector fusion involves fusing a joint state estimate through individual estimates produced by each sensor from its individual Kalman filter, whereas the measurement fusion method directly fuses the sensor measurements to obtain a weighted measurement and feeds it into a single Kalman filter to derive a final state estimate.

The measurement fusion method provides a better overall estimation performance and demands a relative lower computation load on each sensor node. The state-vector fusion method is only effective when the Kalman filters are consistent [76], whereas modeling errors introduced by linearization in many realistic applications often violate this condition. For this reason, we focus our attention on measurement fusion to illustrate the idea.

Note so far we only discuss things in a benign setting whereas in reality there are many malicious situations. To motivate our problem formulation and proposed solution, we name two of the well-known examples in practice that handle such uncertainty.

5.0.3 Trust Rating Systems

Google uses robots to crawl the web pages and then to store their information into their database to calculate the pagerank value. Therefore, Google is characterized as a centralized reputation system [286].

Netscape 8 includes a new "Trust Rating" system that attempts to tell users which sites are "safe". Netscape shows an on-screen indication when it believes a site to be trustworthy [65].

Each system includes a component, or **trust counter**, to compute and store related trustworthiness information.

Paper Organization

After motivating the problem, section 5.1 gives the problem formulation including the fusion method, trust and threat model and the overall assurance idea; section 5.2 explains the details of how the trust rating system works with section 5.3 showing a simple illustrative example.

5.1 Problem Formulation

Among several possible methods for measurement fusion, we choose to fuse observations from different sensors with the inverse of the sensor's variance as weighting factor.

$$y_t = [\sum_{i=1}^{N} R_i^{-1}(t)]^{-1} \sum_{i=1}^{N} R_i^{-1}(t) y_{i,t} \tag{5.7}$$

This method is optimal in the sense of minimum-mean-square-error (MMSE) with a consistent observation vector dimension to have a lower computational load. Note the noise covariance of fused measurement takes the form $R_t = [\sum_{i=1}^{N} R_i^{-1}(t)]^{-1}$. We name this functionality as **fuser**.

Before moving on to the details of assurance system, it's necessary to outline the trust and threat model.

5.1.1 Trust Model

We assume the central location, where the *fuser* and *trust counter* reside, is secure [3].

5.1.2 Threat Model

We assume that the nodes and links are in an uncertain environment, which is subject to attacks from the outside world. Attacks can affect the integrity and availability of the data, such as the man-in-the-middle attack, that may change or delete the data content. Or by taking down certain links, the absence of data from certain nodes may be mistreated as readings being zero.

[3] By resorting to central processing, we restrain ourselves from potential attacks such as *bad mouthing* in distributed systems.

5.1.3 Assurance

Facing these potential threats, we add a trust rating system (Fig 5.2) with details in section 5.2.

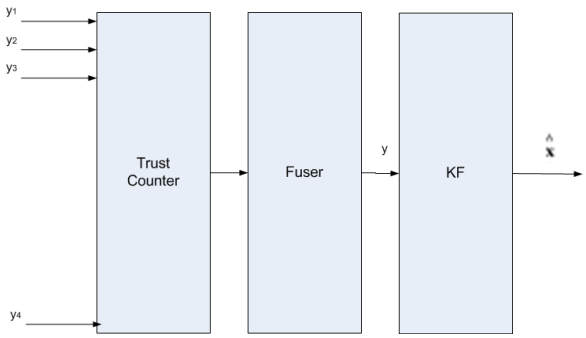


Figure 5.2: The Architecture for Fusion Assurance

The architecture adds a *trust counter* that maintains the trustworthiness and untrustworthiness values of each nodes, on top of the original fusion mechanism, seen in Fig 5.2.

5.2 Trust Rating System

α_i and β_i represent the corresponding ratings for $node_i$ and are determined by equation 5.8. These two values range with (0,1) and depend on the offset contributed by the variation of the existing overall median upon the introduction of the reading from this particular node. If the new median is off beyond a preset threshold value, namely $|\hat{m}_i - m| > Threshold$, the node has untrustworthiness of 1 and trustworthiness 0. Or if its reading doesn't introduce notable difference from the existing median, then the node has trustworthiness 1 and untrustworthiness 0. Otherwise, if the resulted change is within the threshold, $|\hat{m}_i - m| < Threshold$, then its trustworthiness is proportional to the change it introduced versus the threshold value $\frac{Threshold - |\hat{m}_i - m|}{Threshold}$. It's worth pointing out that the median of all measurements y_i is a *robust metric* to quantify the individual measurement [118].

$$(\alpha_i, \beta_i) = \begin{cases} (1,0), & if \quad |\hat{m}_i - m| = 0 \\ (\frac{T - |\hat{m}_i - m|}{T}, 0) & if \quad |\hat{m}_i - m| < T \\ (0,1), & if \quad |\hat{m}_i - m| > T \end{cases} \tag{5.8}$$

In fact, the trust ratings are represented by *Beta distribution* [135] with α and β as its parameters.

$$Beta(\alpha, \beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \tag{5.9}$$
$$\forall 0 \leq x \leq 1, \alpha \geq 0, \beta \geq 0.$$

The central counter updates the trust ratings of $node_i$ based on r_i truthful and s_i bogus observations. Given that the two sets of observations are binary, i.e., truthful or not and bogus or not,

they follow *Binomial distribution*. Indeed, the Beta distribution is the *conjugate prior* of the Binomial distribution and its posterior as well. By using a Bayesian parameter estimation of binomial distribution, it follows that

$$\frac{Bin(r_i+s_i,r_i)*Beta(\alpha_i,\beta_i)}{Normalization} = Beta(\alpha_i+r_i,\beta_i+s_i) \tag{5.10}$$

5.2.1 Update Algorithm

The sequences of truthful/bogus observations of a given measurement evolve, as the status of the uncertain network may vary. We must update the ratings in order to reflect the latest status.

$$\begin{aligned} r_i^t &= \lambda r_i^{t-1} + \alpha_i \\ s_i^t &= \lambda s_i^{t-1} + \beta_i, \end{aligned} \tag{5.11}$$

where λ is a discounting factor ranging from 0 to 1 to reflect the fact that the older the information, the less it worths.

Thus the future (projected) truthfulness of a measurement from a given node can be estimated as

$$\begin{aligned} T_i &= E[Beta(r_i+1,s_i+1)] \\ &= \frac{r_i+1}{r_i+s_i+2} \end{aligned} \tag{5.12}$$

Hence the fused measurement under assurance is

$$y_t = [\sum_{i=1}^{N} T_i R_i^{-1}(t)]^{-1} \sum_{i=1}^{N} T_i R_i^{-1}(t) y_{i,t} \tag{5.13}$$

where, T_i is the truthfulness for each corresponding node measurement determined by the central trust rating system.

5.3 Example

As an illustration, in this section, we demonstrate the idea through simple examples.

There are 30 identical sensors uniformly distributed over the surveillance region. We model the discrete dynamics and measurement of the evader as

$$\begin{aligned} x_{t+1} &= A^e x_t + w_t \\ y_{i,t} &= C_i x_t + v_{i,t} \end{aligned} \tag{5.14}$$

where w and v are white Gaussian noises with zero mean and covariance Q^e = diag $(0.15^2, 0.15^2, 0.15^2, 0.15^2)$

and $R_i = R = \text{diag}(0.15^2, 0.15^2)$, and $\delta = 0.5$ is the sampling period.

$$A^e = \begin{bmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & \delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad C_i = C \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^T \tag{5.15}$$

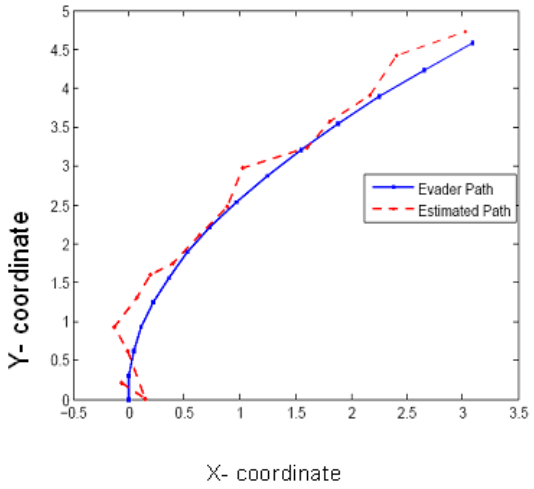


Figure 5.3: Tracking without Trust Rating

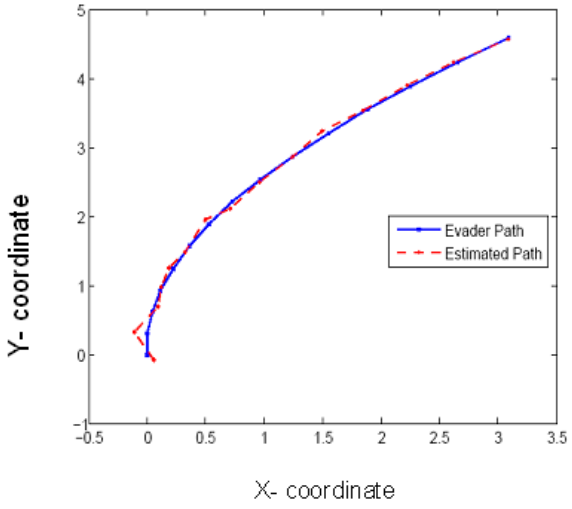


Figure 5.4: Tracking with Trust Rating

From Fig.5.3 and Fig.5.5, we can see the accuracy improves for measurements with trust rating.

The similar holds true when we use 1000 nodes and observe how the estimation error varies as more readings are compromised, shown in Fig.5.5

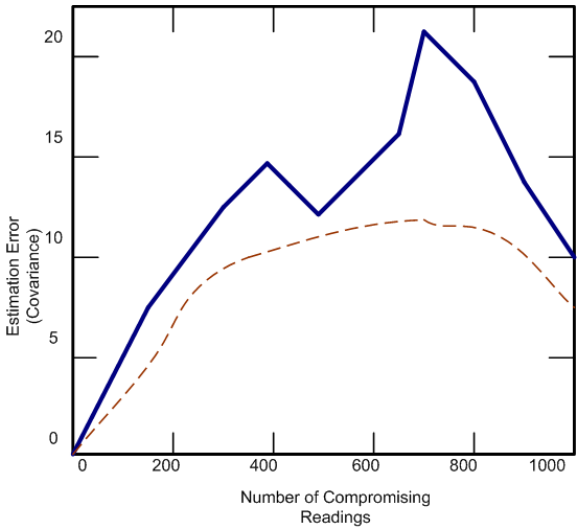


Figure 5.5: Estimation Error: ... dot line indicates with trust rating, – solid line without

5.4 Related Work

There are works making the effort to use reputation frameworks in distributed systems such as [77]. However, it's hard to work around the problems such as compromised nodes being message passing leader or bad mouthing from compromised nodes.

While in our setting, we think it's doable to apply this method in module fashion such that the trust computing base can be limited to the central location only.

5.5 Discussion

In a networked control system setting, where the nodes and links are subject to attacks, the usage of a centralized trust rating system shows the potential to assure the validity of nodes' readings. By using Beta distribution, it only requires storing two parameters thus it's simple yet intuitive. This approach provides intermediate assurance to the data fusion used by the Kalman filter before full-scale implementation of security solutions to the networked control systems. Particularly, this mechanism can facilitate the disambiguation between honest yet rare events and malicious ones. It's implemented in our follow-on work.

Chapter 6

Robust General Likelihood Ratio Test

Faster Higher Stronger

Olympic Motto

This chapter gives the gist of *Robust General Likelihood Ratio Test* (RGLRT) in the context of SCADA security in particular.

The adaptation of large-scale *Wireless Sensor Networks* (WSN) has enabled *Supervisory Control And Data Antiquation* (SCADA) systems with critical remote monitoring. Meanwhile the large networks are prone to benign components failures and malicious attacks. To address such problems, we present an earlier anomaly detection and resilient estimation scheme for the cyber-physical systems, networked control systems to be specific, in an uncertain network environment. It *robustly* identifies and detects outliers among real-time multidimensional measurements of dynamical systems by using an online window-limited sequential *Robust Generalized Likelihood Ratio* (RGLR) test without any prior knowledge of the occurrence time and distribution of the outliers. The robust sequential testing and quick detection scheme achieves the optimal stopping time with low rates in both false alarm and misdetection. We propose a set of qualitative and quantitative *metric* to measure its optimality in the context of cyber-physical systems.

Further, this resilient and flexible estimation scheme *robustly* rectifies and cleans data upon both isolated and patchy outliers while maintain the optimality of the Kalman Filter under the nominal condition. We show the approximated optimality of the robustification performance through *stochastic approximation*. We also offer a simple simulation example to illustrate our ideas.

Supervisory Control And Data Antiquation (SCADA) systems are deeply ingrained in the fabric of critical infrastructure sectors including power grids; oil and gas pipeline systems; water treatment and distribution; railroads and mass transit; and widely involved in the constitutions of vital enterprises such as manufacturing plants and building climate control [79]. The *Wireless Sensor Network* (WSN) has been an emerging application in SCADA systems. In the monitoring and control of moving or remote machinery , wireless sensor networks have compelling economic and engineering advantages over their wired counterparts [218]). They may also deliver crucial information in real-time from environments and processes where data collection is impossible or impractical with wired sensors. Individual sensors simultaneously sense an process and transmit

measured information over a lossy wireless network to a control center, which processes the data and produces an optimal estimate of the state.

However, the uncertainties in the SCADA system itself [296] and in the wireless sensor networks including both benign component faults and malicious attacks may skew the sensor measurements and thus that of the estimation and control command results.

What motivates us to address the issue of outlier-detection and -mitigation is multifaceted. First, outliers are often a clear indication of environmental noise level and potentially faults in sensors or malicious attacks in the system [306]. As for their impact on the applications, in general the performance of linear least squares estimates may degrade remarkably when plant or observation disturbances are non-Gaussian, particularly when the non-Gaussianness, i.e., outlier, is of a heavy-tailed variety giving rise to occasional very large values [265, 116, 117]. In light of the prevalent and broad usage of the Kalman filter in engineering fields and SCADA systems in particular, we are mostly interested the skewing impact of outliers [179] having on the Kalman filter among many other decision making algorithms that are subject to outliers. The state estimation error can grow without bound since the estimate is a linear function of the observation noise. Outliers skew and affect the performance of many decision making algorithms, the standard Kalman filter , and potentially leads to divergency [74] and instability [238] and destabilize the whole controller.

On the other hand, the difficultly of online detection of outliers lies in that moments-based procedures themselves are *not* robust upon outliers [30, 120]. Furthermore, the fact that the adversaries have control over inputs makes the detection task more complicated.

The CUSUM (Cumulative Summation) method and its variants are widely used for anomaly detection. As pointed out in [25],[254], its major drawback is that it requires *a priori* knowledge on information after change, i.e. the intensity of the anomaly etc. But in practice, such information are not predicable. Given that our work is closely related to CUSUM, sequential analysis and hypothesis testing in general, we deem that the related sequential testing approaches deserve a brief exposition in more details in the following Section 6.1.

To address robustness issues, [310] proposes a filtering technique that ensures an estimation error variance with a guaranteed upper bound given the norm-bounded time-varying parameter uncertainty in both the system state and output measurement matrices. Their focus doesn't include outlier detection though. [260] uses a weighted least squares-like approach by introducing weights for each data sample. A data sample with a smaller weight has a weaker contribution when estimating the current time step's state. They treat the problem as an *expectation maximization* (EM) learning problem with maximization over all available data points at every time step while using a variational factorial approximation of the true posterior distribution to get analytically tractable inference. [132] removes the drifting tracking points using Kalman filter when the flow based tracking approach is possibly prone to outliers due to its aperture problem.

Hammes [95] studies robust positioning algorithms for transmitter devices over wireless networks where the non-line-of-sight propagation effects lead to erroneous signal parameter estimates. The framework of an extended Kalman filter (EKF) is rewritten into a linear regression model at each time step while non-parametric pdf estimation is used for position estimation within a parametric signal model to solve for position and velocity of the user equipment.

Contribution of our work:

- we offer a simplified taxonomy/comparison of change detection methods;
- we present a resilient and flexible estimation scheme *robustly* rectifies and cleans data upon both isolated and patchy outliers while maintain the optimality of the Kalman Filter under the nominal condition;
- we propose an online window-limited sequential *Robust Generalized Likelihood Ratio* (RGLR) test without any prior knowledge of the occurrence time or the distribution of the outliers;
- the robust sequential testing bears optimal stopping time,i.e., asymptotically shortest detection delay time while maintaining lowest false alarm rate.

The rest of this paper is organized as the following, Section 2. gives a brief exposition of hypothesis testing and a taxonomy/comparison of related work; Section 3 states the problem formulation including performance metrics; Section 4. presents the resilient estimation; Section 5. describes the scheme for outlier detection; Section 6. shows simulation results, evaluation and discussion. Section 7. Concludes.

6.1 Hypothesis Testing

In this section, we give an overall review of hypothesis testing, sequential analysis and detection before listing a simplified taxonomy.

Let $\mathfrak{M}$ be the set of probability measures on the real line $\mathbf{R}$ and let P_0, P_1 be two distinct elements of $\mathfrak{M}$, having densities p_0, p_1 with respect to some measure ω. Denote $\{z_k\}_0^m$ sequence of *identically independently distributed* (***iid***) observations of a random variable Z with distribution D. The testing problem is hypotheses

$$\begin{cases} H_0 & : \quad D = P_0 \\ H_1 & : \quad D = P_1 \end{cases} \tag{6.1}$$

Let p_{θ_i}, dependent on a parameter θ, be the respective densities of P_i for $i = 0, 1$ with respect to some dominating measure ω.

To discriminate between two we may either use the likelihood ratio test provided by the Neyman-Pearson lemmma, or Wald's sequential probability ratio test.

Recall that *log-likelihood ratio* is defined as

$$\begin{cases} s(\theta, z, i) & = \quad log \frac{p_{\theta_1}(z)}{p_{\theta_0}(z)} \\ S_n & = \quad \sum_{i=1}^n s(\theta, z, i) = \sum_{i=1}^n log \frac{p_{\theta_1}(z_i)}{p_{\theta_0}(z_i)} \end{cases} \tag{6.2}$$

6.1.1 Fixed Sample Size Test

For the **Neyman-Pearson** test, the sample size is fixed and we reject hypothesis H_0 if S_n is too large.

6.1.2 Sequential Probability Ratio Testing

Wald's *Sequential Hypothesis Testing* (SHT), or the *Sequential Probability Ratio Testing* (SPRT) scheme [270] in 1947 not only enjoys the benefits of relatively small sampling size as that of single sampling schemes in the detection of large changes, but also retains a desirable expected sampling size before action is taken when dealing with small changes in magnitude [205].

The task of SHT becomes

$$\begin{cases} S_0 &= 0 \\ S_{k+1} &= log\frac{p_1(Z_k)}{p_0(Z_k)} + S(k), \qquad k \geq 1 \\ N &= \inf\{n \geq 1 : S_n \notin [L,U]\}, \end{cases} \tag{6.3}$$

The SHT decision rule d_N follows,

$$d_N = \begin{cases} H_1 & \text{if} \quad S_N \geq U \\ H_0 & \text{if} \quad S_N \leq L \end{cases} \tag{6.4}$$

where $L \approx ln\frac{F_N}{1-F_A}$ and $U \approx ln\frac{1-F_N}{F_A}$ with F_A being the predefined *false alarm rate* and F_N the predefined *false negative rate* or the missed detection rate upon user's choice and tuning.

Under the assumptions that hypothesis H_0 is of the distribution P_0 with a probability function p_0 and H_1 of P_1 and p_1. Pick 2 numbers a,b with $a < 0 < b$ and define the decisive sample number (the stopping rule or the detection rule)

$$N = \inf\{n \geq 1 : S_n \leq a \ \text{ or } \ S_n \geq b\} \tag{6.5}$$

with $\inf 0 \neq \infty$.

Wald [270] proved that N is almost surely finite under both P_0 and P_1. The testing procedure is to stop at stage N and reject T_0 if $S_n \geq b$ and accept H_0 if $S_n \leq a$ (hence reject H_1). We denote this test $SPRT(a,b,P_0,P_1)$. The average sample numbers are $\mathbb{E}_j[N], j = 0,1$, where $\mathbb{E}_j$ denotes expectation under P_j. The error probabilities are $\alpha = P_0(S_n \geq b)$ and $\beta = P_1(S_n \leq a)$.The SPRT is optimum in the following sense. Consider any other testing procedure with corresponding elements $\alpha', \beta', \mathbb{E}_0, \mathbb{E}_1$ then (cf. Lehmann 1959 [159]), it holds that

$$\begin{cases} \alpha' &\leq \alpha \\ \beta' &\leq \beta \end{cases} \Rightarrow \begin{cases} \mathbb{E}_0[N] &\leq \mathbb{E}_0[N]' \\ \mathbb{E}_1[N] &\leq \mathbb{E}_1[N]' \end{cases} \tag{6.6}$$

SPRT's major strength lies in two-fold that it's a recursive online scheme and optimal in sample size for both hypothesis with theoretical proof on bounds. However, it assumes θ_1, the distribution after change is known, while in reality, especially for the goal of this paper, it is not.

Sequential Detection

Closely related to sequential testing theory is the theory of sequential change-point detection. Page [205] and Shiryaev [248] modified Wald's SPRT and developed the **cu**mulative **sum** (CUSUM) [205] and the Shiryaev-Roberts charts [248] respectively to improve the sensitivity of

the Shewhart charts [247]. The goal of optimality in the *Shiryaev-Roberts-Pollak* (SRP) sense is to minimize the *worst-case average delay* subject to the upper bound of a false alarm whereas in *Lorden's* sense is to minimize the upper bound of the worst case delay subject the upper bound of a false alarm [166].

The CUSUM [26, 33, 88, 188] test is one of the most successful algorithms of sequential change detection. The CUSUM procedure developed in 1954 calculates the cumulative sum of samples from a process X_n with weights ω_n in the following fashion,

$$\begin{cases} S_0 &= 0 \\ S_{n+1} &= max(0, S_n + X_n - \omega_n) \end{cases}$$

The stopping rule or the detection rule is that: when the value of S exceeds a certain threshold value, a change in value has been found [1].

Widespread applications and theory development in quality control [168, 188, 235], fault detection [51, 276], surveillance [121, 133], anomaly detection [252, 172] are stemmed from CUSUM and/or CUSUM alike procedures.

Some of the methods proposed over the years were originally ad hoc procedures and were later proven to possess optimality properties including both Wald's SPRT or Page's CUSUM. Others remain popular though sub-optimal such as Shewhart [247] and *Exponentially-Weighted Moving Average* (EWMA) [228] control charts.

The overall comparison and a simplified taxonomy is summarized in Table. 6.1.

For a more detailed review on sequential analysis or sequential change-point detection involving multivariate and dependent observations, interested readers please refer to [154] and [25] respectively.

6.2 Problem Formulation

First we recap estimation and identification in state-space models and the statistical approach based on the Kalman filter and likelihood techniques.

6.2.1 A General State Space Model Setting

Let positive integer $k = 0, 1, \ldots$ denotes discrete time, then stochastic state-space model in discrete time has the following form

$$\text{state:} \quad x_{k+1} = F_k x_k + G_k u_k + w_k \tag{6.7}$$

$$\text{observation:} \quad y_k = H_k x_k + J_k u_k + v_k \tag{6.8}$$

where $x_k \in \mathbf{R^n}$ is the (hidden) internal state vector,
$u_k \in \mathbf{R^r}$ is the input vector,

[1] Note the above formula (6.7) only detects changes in the positive direction. When negative changes need to be found as well, the min operation should be used instead of the max operation, and this time a change has been found when the value of S is below the (negative) value of the threshold value.

Work Name	Observation Sequence	Statistical Parameter Knowledge			Problem Formulation	Optimality Criteria	Test Statistics	Key Point
		Occurrence Time	Q_0	Q_1				
Page '54 [205]	iid	unknown	known	known	minimax	Lorden	maximum likelihood	recursive CUSUM
Shiryaev '63 [248]	iid	geometric distribution	known	known	Bayesian	SRP		
Roberts '66 [229]	iid	unknown	known	known	Bayesian	SRP		
Lorden '71 [166] GLR	iid	unknown		unknown	minimax	Lorden	log-likelihood	one-sided SPRT
Wilsky-Jones '76 [277] Window-limited GLR		unknown	known	known			general likelihood	limited- -window
Pollak '85 [215]	iid	unknown	known	known	Bayesian	SRP		almost minimax
Moutiades '86 [192]	iid	unknown	known	unknown	minimax	Lorden		
Lai '98 [155] Reduced-window GLR	dependent	unknown	known	known		change-of-measure		reduced window
Robust-GLR	dependent	unknown	known	unknown		change-of-measure		robust

Table 6.1: A Simplified Taxonomy/ Comparison of Sequential Analysis / Change Point Detection Methods

$y_k \in \mathbf{R^m}$ is the output i.e. observation (measurement) vector.
$w_k \in \mathbf{R^r}$, the process (plant) (6.7) noise vector, is a white Gaussian noise sequence with zero mean and covariance matrix $Q_k > 0$.
$v_k \in \mathbf{R^m}$, the observation (measurement) (6.8) noise vector, is a white Gaussian noise sequence with zero mean and covariance matrix $R_k > 0$.
$\{F_k\}$ the state transition matrix, $\{H_k\}$ the observation matrix, $\{G_k\}$ and $\{J_k\}$ the control matrices are known sequences of matrices with appropriate dimensions.
The initial system state vector x_0 is Gaussian with zero mean and covariance matrix P_0. We assume that the initial state x_0 and the two noise sequences w_k, v_k are mutually independent. We will use observation and measurement interchangeably.

In summary, (6.7) is a recursive state model of the linear dynamical process (plant), and (6.8), a linear observation model of the system. Note such a model (6.7)-(6.8) is a Markov model, namely the pair $(X_{k+1}; Y_k)$ is a Markov process.

6.2.2 Kalman Filter

The Kalman filter provides one particular estimate of the state x_k of the system (6.7)-(6.8). It's a *minimum variance* estimate of the state, namely the conditional mean [2] of x_k given the past observations $\{:::; y_{k-2}; y_{k-1}\}$. We denote this *one-step ahead prediction* as $\hat{x}_{k+1|k}$.

As shown in Fig.6.1, the overall flow diagram of the Kalman Filter, it's an on-line recursive algorithm. To illustrate its recursion, we decompose its procedure into two phases, namely the *predication phase* and *measurement update phase*.

Fig.6.2 illustrates the recursive procedure of the Kalman filter, noting at each time step, only current and previous step are involved. That is to say no batch operation is required. This is precisely what makes the Kalman filter an *online* algorithm.

6.2.3 Outliers' Distribution Model

We shall point out that employing a outliers' distribution model only gives us a somewhat plausible and trackable model for generating outliers [174] and for illustrating the impact of outliers on estimation performance. That is not to say that our detection scheme is dependent on the outliers' distributions, otherwise it is not robust nor effective.

There are several types *heavy-tailed* or alternatively referred to as *fat-tailed* distributions [3] in wide use[175]. Alternatively, the *contaminated normal distributions* is one specific instance of the more generic mixture distribution model for outliers [93] which will suffice for purposes of our current exposition. To be more specific, the outliers are generated through the contaminated-

[2] When the Gaussian assumption concerning the noises is removed, the Kalman filter gives the linear minimum variance estimate of the state, namely the smallest unconditional error covariance among all linear estimates, but, in general, this estimate is not the conditional mean (Goodwin and Sin, 1984).

[3] A fat tail is a property of some probability distributions exhibiting extremely large kurtosis particularly relative to the ubiquitous Gaussian which itself is an example of an exceptionally thin tail distribution. Fat tail distributions have *power law* decay.

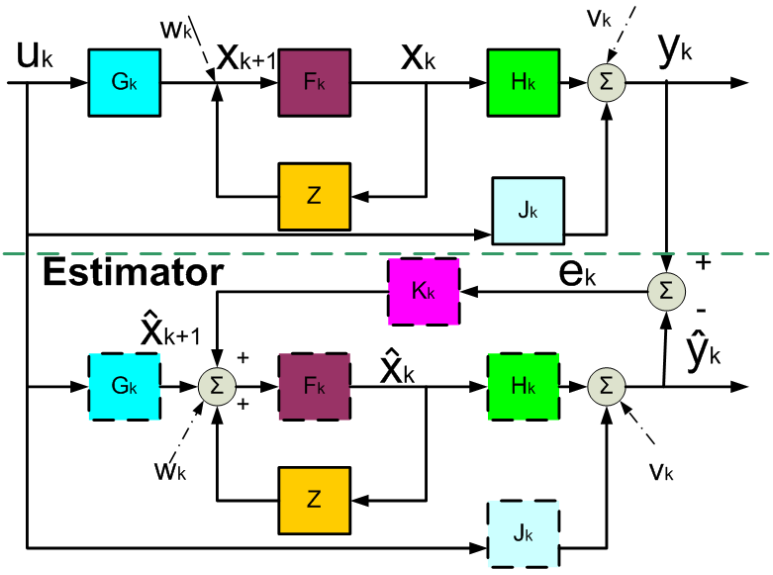


Figure 6.1: The Kalman Filter Flow Chart
innovation: $e_{k+1} = y_{k+1} - \hat{y}_{k+1}$: to "correct"
update: $\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1}e_{k+1}$;
1-step predication: $\hat{x}_{k+1|k} = F_k\hat{x}_{k|k} + G_k u_k$
x_k, u_k, y_k, w_k, v_k: the state, input, observation, process noise, observation noise vector; F_k,H_k,G_k and J_k: the state "transition", observation, control matrices.

normal distribution with *degenerate* central component [174]

$$CN(t;\gamma,\sigma^2) = (1-\gamma)N(t;0,0) + \gamma N(t;0,\sigma^2) \tag{6.9}$$

That is to say the process x_t is observed perfectly about $100(1-\gamma)$ percent of the time and is corrupted by outliers about 100γ percent of the time, where $0.01 \leq \gamma \leq 0.25$.

6.2.4 Further Property Assumptions

Furthermore, for some integer d, let $(\mathbf{R}^d, \mathbf{B}, \lambda)$ be a measure space, where $\mathbf{R}$ is the real line, $\mathbf{B}$ the Borel σ-algebra, and λ the Lebesgue measure. Let $\boldsymbol{F}$ be a zero-mean probability measure on $(\mathbf{R}^d, \mathbf{B})$ such that $\boldsymbol{F}$ is absolutely continuous with respect to λ and admits the density f in accordance with Radon-Nikodym theorem.

We have a sequence of *identically independently distributed* (**iid**) observations $\{z_k\}_0^m$ of a random variable Z with a probability density $p_\theta(Z)$ that is dependent on one *scalar* parameter only. The parameter $\theta = \theta_0$ before a *unknown change time* ν and $\theta = \theta_1$ after ν.

Note that change time ν is *unknown*. We either consider ν as a nonrandom unknown value or a random unknown value with unknown distribution. In other words, we deal with a nonparametric approach as far as this change time ν is concerned. In practice, either it is very difficult to have *a priori* information about the distribution of the change times, or this distribution is nonstationary (i.e. it doesn't have an invariant mean nor variance). This is particularly meaningful for our

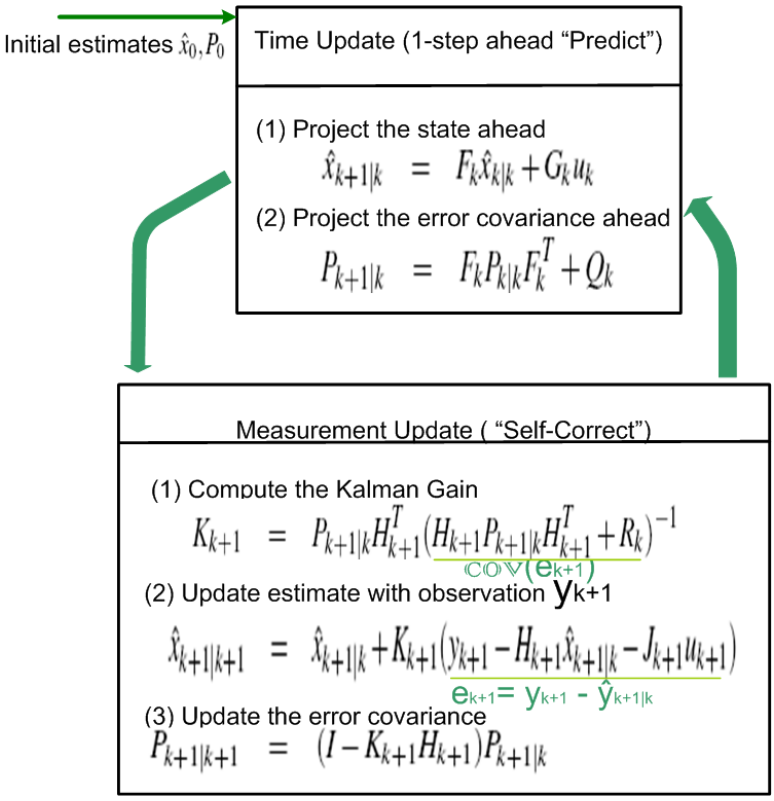


Figure 6.2: The recursive operation of the Kalman Filter: a combination of the high-level diagram in Fig.6.1 and the formulations in section 6.2.2

problem setting, giving that we have no *a priori* knowledge of when the intrusion thus outliers or anomalies would occur at all. That's the reason why certain basic tools can't directly suit our problem.

Our **security model** is that the SCADA center itself is secure and so are the core programs. We assume the attack is session based, should it arise over the network.

By "**resilient**", we stress the importance of the flexility and parsimoniousness of the overall strategy. Without incurring too large overhead, it shall maintain the systems's optimal performance under nominal conditions while strive for near optimal performance should atypical situations arise without being unduly affected by spurious observations.

6.2.5 Meaningful Metrics for Recursive Robust Estimation

It's only appropriate to bring up the issue of the robustness of estimation schemes when we address outliers. Conceptually, the definition of robustness[4] we use here stipulates that small changes from an assumed nominal model would only introduce small changes in estimate, according to both Tukey [265] and Huber [213]. Furthermore, *robust-resistant*, a purely data-oriented notion defined by Tukey [266], refers that an estimate is called resistant if changing a small fraction of the

[4]The word "robust" is loaded with many if not often inconsistent meanings.

data by large amounts results in little change to the estimate. That is to say the capability against gross error and outliers.

Formulation wise, while the minimax approach is pessimistic, it provides an optimum lower bound on performance. Let: $\mathfrak{T}$ be a class of estimates, $\mathfrak{F}$ a class of distributions, and $V(T,F)$ the *asymptotic variance* of $T \in \mathfrak{T}$ when the distribution is $F \in \mathfrak{F}$. Then the minimax robust estimate T_0 and its associated *least favorable distribution* F_0 satisfy

$$\min_{T\in\mathfrak{T}} \max_{F\in\mathfrak{F}} V(T,F) = V(T_0,F_0) = \max_{F\in\mathfrak{F}} \min_{T\in\mathfrak{T}} V(T,F) \tag{6.10}$$

Naturally, this can be viewed as a game in which we choose $T \in \mathfrak{T}$, nature chooses $F \in \mathfrak{F}$ and $V(T,F)$ is the payoff. This game has a saddle point pair (T_0,F_0) if T_0 and F_0 satisfy the above (6.10).

Furthermore, for multivariate, dependent Markovian (state space model) without process noises, analytically the *asymptotic variance* is still a good choice of

Plus, in this paper, this goal is to achieve optimally estimating and tracking the state of stochastic time-variant linear dynamic system rather than obtaining minimum asymptotic estimation error. Thus approximations of a conditional mean estimator which is known for its unbiasedness and minimum error variance [11], are targeted [241].

6.2.6 Sequential Detection Performance Measure

False Alarm Constraints

Often the methodology of optimal change-point detection pursues stopping rules that achieve the best balance of the mean detection delay and the rate of false alarms or minimize the mean delay under a fixed false alarm probability [22]. In order to establish a sound sequential detection performance measure, we must first lay out the associated false alarm probability constraints that the asymptotic lower bound for the detection delay is subject to.

$$\mathbb{E}^{(\nu)}(T-\nu)_{\mathbf{1}\{T\geq\nu\}} = \mathbb{E}^{(\nu)}(T-\nu)^{+} \tag{6.11}$$

Accordingly, three related false alarm probability constraints in the ascendant order of stringency are listed as follows:

- For **iid** observations, due to Shiryaev [248], the *Bayesian* view concerns the *mean delay to detection* under the average false alarm

$$P(T<\nu) = \sum_{k=1}^{\infty} \pi_\alpha(k) P_0(T<k) \leq \alpha \tag{6.12}$$

  where π_α is a prior distribution of the change time ν.

- Whereas the **ARL** (*Average Run Length*) [205] to false alarm constraint in a minimax formulation

$$\mathbb{E}_0[T] \geq \gamma > 1 \tag{6.13}$$

is the worst case in Lorden's sense [166], and is no smaller than a given number $\gamma > 1$ when the quality parameter remains fixed θ. The objective is to find the stopping rule that minimizes the worst-case delay subject to an upper bound on the false alarm rate.

- For non-independent observations, Lai proposed a *change-of-measure* argument [155], the most stringent one among the three, to guarantee a lower bound on the *window-limited* stopping time, or the detection delay:

$$\begin{aligned} \sup_{\nu \geq 1} P_0(\nu \leq T < \nu + m_\alpha) &\leq \alpha, \quad \textit{where} \\ \liminf \frac{m_\alpha}{|\log \alpha|} &> I^{-1} \quad \textit{but} \\ \log m_\alpha = o(\log \alpha) \quad &\text{as} \quad \alpha \to 0. \end{aligned} \tag{6.14}$$

The reason we choose the most stringent false alarm constraint, namely Lai's *change-of-measure* argument (6.14) lies in that it meets our desire to have as low as possible false alarm while achieving an asymptotic lower bound for the detection delay.

Correspondingly, as $\alpha \to 0$ for a positive integer I, the asymptotic lower bound for the detection delay is

$$\begin{aligned} \mathbb{E}^{(\nu)}(T-\nu)^+ \geq \{P_0(T \geq \nu)/I + o(1)\}|log\alpha| \\ \text{uniformly in } \nu \geq 1 \quad . \end{aligned} \tag{6.15}$$

6.3 Resilient Estimation

Contaminated Observations with additive outliers Suppose at an *unknown* time ν, the sensor measurement (observation)y_k (6.8) is subject to some *additive outliers* or *anomaly*, formally

$$\tilde{y}_k = y_k + y_{aok}\mathbf{1}\{k \geq \nu\} \tag{6.16}$$

$$= H_k x_k + J_k u_k + \tilde{v}_k \tag{6.17}$$

$$= H_k x_k + J_k u_k + v_k + y_{aok}\mathbf{1}\{k \geq \nu\} \tag{6.18}$$

where $\tilde{y}_k$ is the observed data and the y_{aok} are the additive outliers $\mathbf{1}\{k \geq \nu\}$, either in *isolation* or in *cluster*, $\mathbf{1}\{k \geq \nu\}$ is a compact notion of an indictor function indicating the occurrence of the outliers (anomaly),

$$\mathbf{1} = \begin{cases} 1 & k \geq \nu \\ 0 & k < \nu \end{cases} \tag{6.19}$$

Theorem 1. *A robust state estimate suffices above conditions is optimal in the min-max sense, i.e. having minimum variance over the least favorable contaminating distributions. It can take the*

following form with $\breve{x}_{k|k} \stackrel{\triangle}{=} \mathbb{E}[\tilde{x}_k|\tilde{\mathbf{y}}_\mathbf{k}]$ *, compared to the original Kalman filter.*

$$\begin{aligned}
\breve{x}_{k+1|k} &= F_{k+1}\breve{x}_{k|k} && (6.20)\\
\tilde{P}_{k+1|k} &= F_{k+1}\tilde{P}_{k|k}F_{k+1}^T + Q_{k+1} && (6.21)\\
\tilde{K}_{k+1} &= \tilde{P}_{k+1|k}H_{k+1}^T\tilde{\Sigma}_{k+1}^{-1} && (6.22)\\
\breve{x}_{k+1|k+1} &= \breve{x}_{k+1|k} + \tilde{K}_{k+1}(\tilde{y}_{k+1} - H_{k+1}\breve{x}_{k+1|k} - J_{k+1}u_{k+1}) && (6.23)\\
\tilde{P}_{k+1|k+1} &= (I - \tilde{K}_{k+1}H_{k+1})\tilde{P}_{k+1|k} && (6.24)
\end{aligned}$$

with the robustified (censored) covariance matrix of the innovation (residual) becoming,

$$\tilde{\Sigma}_k = H_k\tilde{P}_{k|k-1}H_k^T + R_k^{\frac{1}{2}}W_kR_k^{\frac{1}{2}} \tag{6.25}$$

where

$$W_k = diag\{w_{1k}, \cdots, w_{mk}\} \tag{6.26}$$

and $w_{1k}, \cdots, w_{mk}$ *would be defined later in the proof.*

Proof: : We first show the result through construction. It is straightforward that the state estimator $\breve{x}_{k|k}$ corresponding to $\hat{x}_{k|k} = \mathbb{E}[x_k|\mathbf{y}_\mathbf{k}, \mathbf{u}_\mathbf{k}]$ of the original Kalman filter can be obtained by minimizing

$$\begin{aligned}
\breve{x}_{k+1|k+1} &= \\
argmin \quad \{ & \quad (\breve{x}_{k+1|k} - x_{k+1})^T(P_{k+1|k})^{-1}(\breve{x}_{k+1|k} - x_{k+1})\\
+ & \quad (\tilde{y}_{k+1} - H_{k+1}x_{k+1} - J_{k+1}u_{k+1})^T(R_k)^{-1}\\
\times & \quad (\tilde{y}_{k+1} - H_{k+1}x_{k+1} - J_{k+1}u_{k+1})\}
\end{aligned} \tag{6.27}$$

with respect to $x_{k+1} \in R^n$, or equivalently

$$\breve{x}_{k|k} = argmin\{\sum_{i=1}^{n}(p_{ik} - a_{ik}x_k)^2 + \sum_{j=1}^{m}(s_{jk} - b_{jk}x_k - q_{jk})^2\} \tag{6.28}$$

where $p_k = (P_{k|k-1})^{-\frac{1}{2}}\breve{x}_{k|k-1}, s_k = (R_k)^{-\frac{1}{2}}\tilde{y}_k, q_k = (R_k)^{-\frac{1}{2}}J_ku_k, a_k = (P_{k|k-1})^{-\frac{1}{2}}, b_k = (R_k)^{-\frac{1}{2}}H_k$, so that p_{ik}, s_{ik} and q_{jk} are the $i-$th component of the vectors $p_k \in R^{n\times 1}$, $s_k \in R^{n\times 1}$ and $q_k \in R^{n\times 1}$ correspondingly; $a_{ik} \in R^{1\times n}$ and $b_{ik} \in R^{1\times n}$ are the $i-$ row vector of the matrix $a_k \in R^{n\times n}$ and $b_k \in R^{n\times n}$ correspondingly. In the case of $M-$estimation, the least squares solution is replaced by

$$\breve{x}_{k|k} = argmin\{\sum_{i=1}^{n}(p_{ik} - a_{ik}x_k)^2 + \sum_{j=1}^{m}\rho_j(s_{jk} - b_{jk}x_k - q_{jk})^2\} \tag{6.29}$$

where the ρ_j are suitable score functions with derivatives, i.e. influence function ψ_j, or psi-function used in robust statistics. One of Huber's psi-function is

$$\psi_H(Z) = \begin{cases} & \text{for} \quad |Z| \leq s \\ s \ sgn(Z) & for \quad |Z| > s \end{cases} \tag{6.30}$$

is often used[5]. It gives robust estimates of location which are optimal in the min-max sense, having minimum variance over the least favorable contaminating distributions.

The normal equations for $\breve{x}_{k|k}$ corresponding to (6.29) have the form

$$\sum_{i=1}^{n} a_{ik}^T (p_{ik} - a_{ik}\breve{x}_{k|k}) + \sum_{j=1}^{m} b_{jk}^T \psi_j(s_{jk} - b_{jk}\breve{x}_{k|k} - q_{jk}) = 0 \tag{6.31}$$

and can be solved explicitly only in some special cases. This is quite pragmatic as well, sensors are normally set with bound values in practice.

Alternatively, one can use the approximated normal equations if we approximate[6] $\breve{x}_{k|k}$ by $\breve{x}_{k|k-1}$ when using the weight function w_{jk} as the following,

$$\sum_{i=1}^{n} a_{ik}^T (p_{ik} - a_{ik}\breve{x}_{k|k}) + \sum_{j=1}^{m} w_{jk} b_{jk}^T (s_{jk} - b_{jk}\breve{x}_{k|k} - q_{jk}) = 0 \tag{6.32}$$

where the weight functions $w_{jk}, j = 1, \ldots, m$ are

$$w_{jk} = \frac{\psi_j(s_{jk} - b_{jk}\breve{x}_{k|k} - q_{jk})}{s_{jk} - b_{jk}\breve{x}_{k|k} - q_{jk}} \tag{6.33}$$

Using (6.32) and some algebra, we obtain robustified (censored) covariance matrix of the innovation (residual),

$$\tilde{\Sigma}_k = H_k \tilde{P}_{k|k-1} H_k^T + R_k^{\frac{1}{2}} W_k R_k^{\frac{1}{2}} \tag{6.34}$$

where $W_k = diag\{w_{1k}, \cdots, w_{mk}\}$

6.4 Robust Outlier Detection

The overall procedure is shown as in Figure 6.3.

6.4.1 System model with outliers contaminated observations

Following the definition of the contaminated measurement $\tilde{y}_k$ (6.16-6.19), the state $\tilde{x}_k$, the estimate $\breve{x}_{k|k}$, and the output residual $\tilde{e}_k$ of the Kalman filter upon the outliers occurred at time ν

[5]The recommended choice of s in (6.30) is $s = u_{1-\varepsilon}$ where u_α is the α-quantile of $N(0,1)$ (e.g., $s = 1.883$ for a 3% contamination of data.

[6]They can be considered as a recursive variant of the normal equations from the *Iterative Weighted Least Squares* IWLS method which is a popular algorithm for numerical calculation of M-estimates.

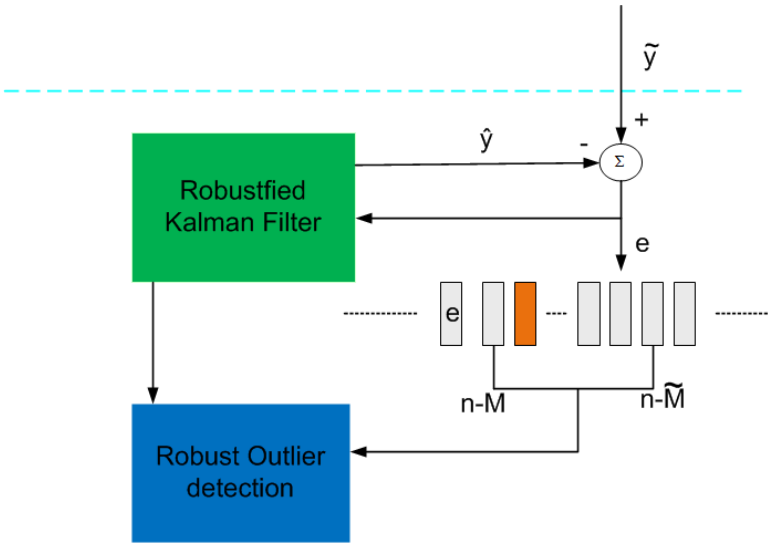


Figure 6.3: Block Diagram of Robust Outlier Detection and Resilient Estimation

can be expressed in the relations of their nominal counterparts, as

$$\begin{array}{rcl} \check{x}_{k|k} & = & \hat{x}_{k|k} + \beta(k,\nu) y_{ao} \\ \tilde{e}_k & = & e_k + \rho(k,\nu) y_{ao} \end{array} \tag{6.35}$$

where the terms $\beta(k,\nu), \rho(k,\nu)$ would be defined later.

Conditioned on the past outputs $\mathbf{y_k}$ and input signals $\mathbf{u_k}$, the innovation e_k has the *conditional mean* $\mathbb{E}[e_k]$. Let's denote $\mu_k = \mathbb{E}[e_k]$, then

$$\mu_k = \mathbb{E}[e_k] = \begin{cases} \rho(k,\nu) y_{aok} & k \geq \nu \\ 0 & k < \nu \end{cases} \tag{6.36}$$

where ν, y_{ao} are unknown. The $\rho(k,t)$ are matrices that can be recursively evaluated after initialization $\rho(t,t) = 0, \beta(t-1,t) = 0$,

$$\beta(k,t) = F_{k-1}\beta(t-1,k) + K_k \rho(k,t) \tag{6.37}$$

$$\rho(k+1,t) = -H_{t+1}F_k\beta(k,t) + I \tag{6.38}$$

where $\beta(k,t)$ and $\rho(k,t)$ are the difference of the estimate $\check{x}_{k|k}$, residual $\tilde{e}_k$ under outliers, comparing with their nominal counterparts as stated in (6.35), to be evaluated recursively in parallel for $k \geq t$ and for every fixed t, one for each t within a moving window $t \in \{n-m, \cdots, n-m'\}$.

Meanwhile, the covariance matrix of the innovation is

$$V_k = \mathbb{E}[(e_k - \mathbb{E}[e_k])(e_k - \mathbb{E}[e_k])^T] \tag{6.39}$$

$$= \begin{cases} \tilde{\Sigma}_k & k \geq \nu \\ \Sigma_k & k < \nu \end{cases} \tag{6.40}$$

$$= \tilde{\Sigma}_k \tag{6.41}$$

It's easy to verify the design purpose, for $k < \nu$ weight functions $w_{jk} = 1$, $\forall j \in [1, m]$ thus $\tilde{\Sigma}_k = \Sigma_k$

6.4.2 Robust Sequential Probability Ratio Tests

According to Huber [119], a statistical procedure is called robust if its performance is insensitive to small deviations of the idealized theoretical model. In terms of the robustness of a test, it shall withstand small arbitrary departures from both the null hypothesis (*robustness of validity*) and the specified alternatives (*robustness of efficiency*) [120]. When encountering deviation, the classical probability ratio test is not robust in the following sense: a single outlying data point thus deviating factor $p_1(x_j)/p_0(x_j)$ equal (or almost equal) to 0 or ∞ may unduely impact the test statistic $T(x) = \prod_1^n p_1(x_j)/p_0(x_j)$ therefore may totally skew the final hypothesis or probability test outcome. By censoring the single factors at some fixed numbers $c' < c''$ for sequential probability ratio test, one can replace the test statistic by $T'(x) = \prod_1^n \pi(x_j)$, where $\pi(x_j) = max\{c', min\{c'', \frac{p_1(x_j)}{p_0(x_j}\}\}$.

Note that we have precisely done so in the stage of resilient estimation that one of the key components of our test statistics, the covariance matrix of the innovation (residual), $\tilde{\Sigma}_k$ (6.34) or V (6.39), has been "censored".

Detection Rules

Without assuming any prior knowledge of parameter η, the RGLR rule maximizes the log likelihood ratio over a window of inputs and decide the time to raise an alarm according to certain rule, which we will state without formally proving as certain steps have showed by Huber [119] and Quang [223] in a sequential testing setting .

Theorem 2. *The following stopping rule is optimal and robust*

$$\begin{aligned} N_G &= \inf\{n : \max_{n-M\le t\le n-M'} \sup_{\eta} \sum_{i=k}^{n} \log[f(\tilde{\Sigma}_i^{-1/2} \\ &\quad \times(e_i-\rho(i,t)\eta))/f(\tilde{\Sigma}_i^{-1/2}e_i)] \ge c_\lambda\} \\ &= \inf\{n : \max_{n-m\le t\le n-m'} (\sum_{i=k}^{n} \rho^T(i,t)\tilde{\Sigma}_i^{-1}e_i)^T \\ &\quad \cdot(\sum_{i=k}^{n} \rho^T(i,t)\tilde{\Sigma}_i^{-1}\rho(i,t))^{-1} \\ &\quad \cdot(\sum_{i=k}^{n} \rho^T(i,t)\tilde{\Sigma}_i^{-1}e_i)/2 \ge c_\lambda\} \end{aligned} \tag{6.42}$$

where $f(y) = e^{\frac{-\|y\|^2}{2}}/(2\pi)^{\zeta/2}$ *denotes the* ζ*-dimensional normal density,* $\zeta = dim(\eta)$*, and* $m'+1 \ge \zeta$ *so that the matrix inversions in (7.28) are valid.*

In essence, we are looking at an **optimal stopping time problem**:not to stop too early to produce a false alarm nor to stop too late to miss a real anomalous event.

Huber [119] showed that in the neighborhoods of the idealized underlying distributions, which is the least favorable situation for both Type I (false alarm) and Type II (miss detection) error probabilities, the so called *censored probability ratio test* is most robust in a well defined minimax sense.

In light that our test statistic has undergone the censoring processing at the robustified estimation stage, so our concerns translate into whether the corresponding sequential testing still are least favorable for errors.

Quang [223] further proved that with the limiting maximum error probabilities less $1/2$, such sequential test is also least favorable for ASN *Average Sample Number* and asymptotically minimax with respect to expected sample sizes.

6.4.3 Threshold and Window size Choice

Note that (7.27) computes $\rho(t,k)$ recursively over the each window. How to optimally choose $M, \tilde{M}$ and c_λ in general is a difficult problem [25] for online practices particularly due to the coupling effect between the threshold and window size on the asymptotical performance of the detection rule. But for off-line operations, the choice of window size is less demanding as all the data set is available, it's only a matter of computation time.

The threshold c in the rule N_W subject to the false alarm probability criterion $P_0(N_W \le m)$ can be computed by using Monte Carlo computation of $P_0(N_W)$ together with the method of successive linear approximation combined with bisection search for iterative solution of the equation $P_0(N_W \le m)$.

With the window size M, we have $M \sim a\log\gamma$ where $E_0(T) \sim \gamma$, and $a > \frac{1}{I(\theta,0)}$. The importance sampling procedure procedure for Monte Carlo computation of $P_0(N_w \le m)$ involves the following steps as shown in Algorithm. 1,

Algorithm 1 Importance Sampling for P_0

while $N \geq 0$ **do** {run N times}
  generate $\nu \in \{1, m\}$ and $\theta \in N(0, p)$
  for $t \leq min(N_W, m)$ **do**
    if $t \leq \nu$ **then**
      $cov_t(e_t) \leftarrow V_t$
      $E_t(e_t) \leftarrow 0$
    else
      $cov_t(e_t) \leftarrow \tilde{V}_t$
      $E_t(e_t) = \rho(t, \nu)\theta$
    end if
    for $1 \leq k \leq i \leq t \leq m$ **do**
      $C_{t,k} \leftarrow I + \Sigma_{i=k}^{t} \rho^T(i,k) V_i^{-1} \rho(i,k)$
      $d_{t,k} \leftarrow \Sigma_{i=k}^{t} \rho^T(i,k) V_i^{-1} e_i$
      $L_t \leftarrow \frac{\Sigma_{k=1}^{t}(detC_{t,k})-1/2exp(d_{t,k}^T C_{t,K}^{-1} d_{t,k}/2)+1-t}{m}$
    end for
  end for
  $N \leftarrow N - 1$
end while
$P_0(T \leq m) \leftarrow \frac{\Sigma^N L_{n,W}^{-1}}{N}$

Note that $E_0(T) \sim \frac{m}{P_0(T \leq m)}, \sim \log \gamma$, thus threshold c in the rule N_W subject to the false alarm probability criterion $P_0(N_W \leq m/\gamma)$ can be computed by using the above procedure for Monte Carlo computation of $P_0(N_W$ together with the method of successive linear approximation combined with bisection search for iterative solution of the equation $P_0(N_W \leq m/\gamma)$.

6.5 Experiments and Evaluation

Currently, we are using synthetic data to conduct experiments. We model the discrete dynamics and two-dimensional measurement of the tracked object as

$$\begin{aligned} x_{t+1} &= A^e x_t + w_t \\ y_{i,t} &= C_i x_t + v_{i,t} \end{aligned} \tag{6.43}$$

where w and v are white Gaussian noises with zero mean and covariance Q^e = diag $(0.15^2, 0.15^2, 0.15^2, 0.15^2)$ and $R_i = R$ = diag$(0.15^2, 0.15^2)$, and $\delta = 0.5$ is the sampling period.

$$A^e = \begin{bmatrix} 1 & 0 & \delta & 0 \\ 0 & 1 & 0 & \delta \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \qquad C_i = C \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}^T \tag{6.44}$$

The reason of employing such examples lies in that

- its multidimensionality suffices the complexity purpose;
- it's generic enough to illustrate the impact of outliers.

6.5.1 Resilient Estimation Performance

As stated in Section 6.2.5, we evaluate the estimation performance in terms of the error variance. Figure 6.4 shows that our resilient estimation scheme performance better than the standard Kalman filter upon randomly injected outliers while maintaining the latter's under nominal conditions.

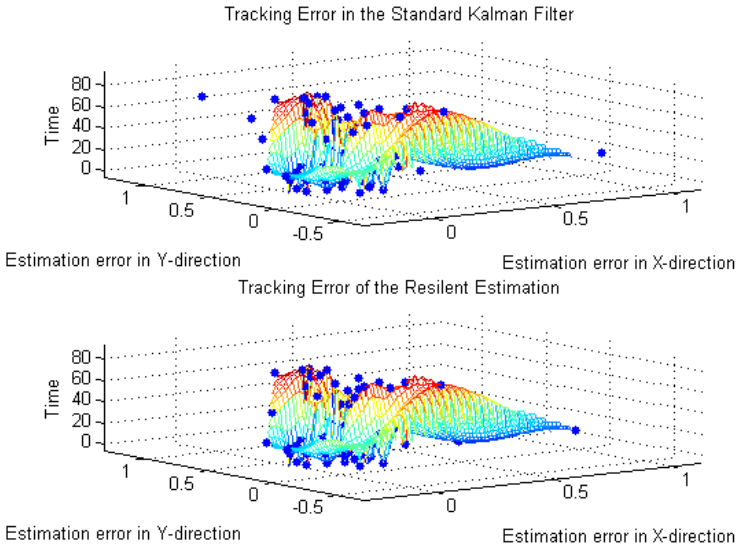


Figure 6.4: Tracking Error Comparison: The lower panel shows the performance of our Resilient Estimation is identical to that of the standard Kalman filter under nominal condition while having much smaller errors upon outliers at time T = 10,30,60.

6.5.2 Robust Outlier Detection Performance

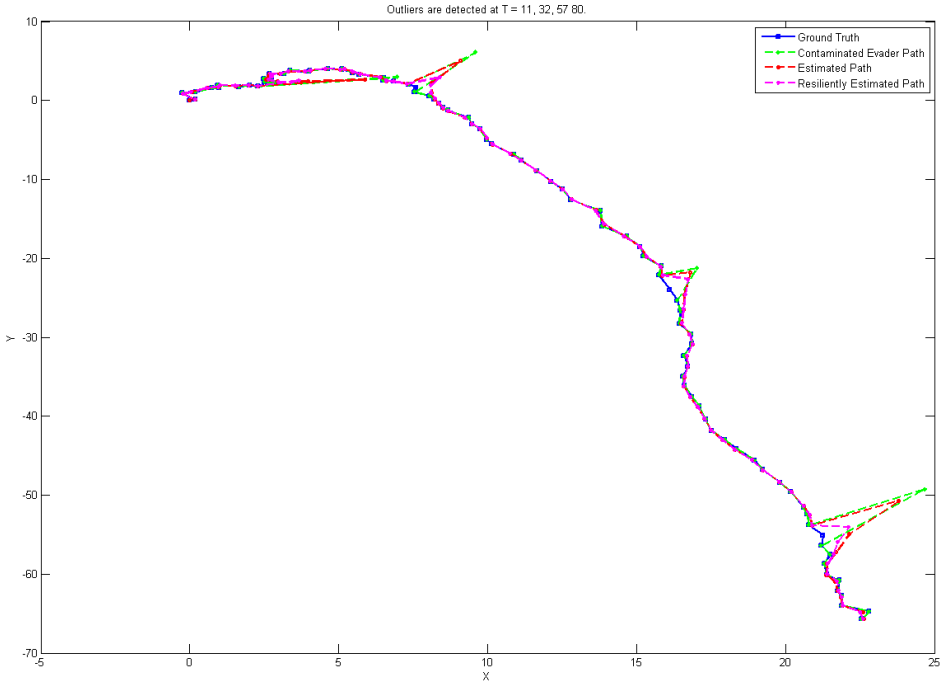
With randomly injected outliers where the false alarm constraint is achieved through Monte Carlo simulation, our approach successfully detects multiple them as shown Figure 6.5.

6.5.3 Limitation and Discussion

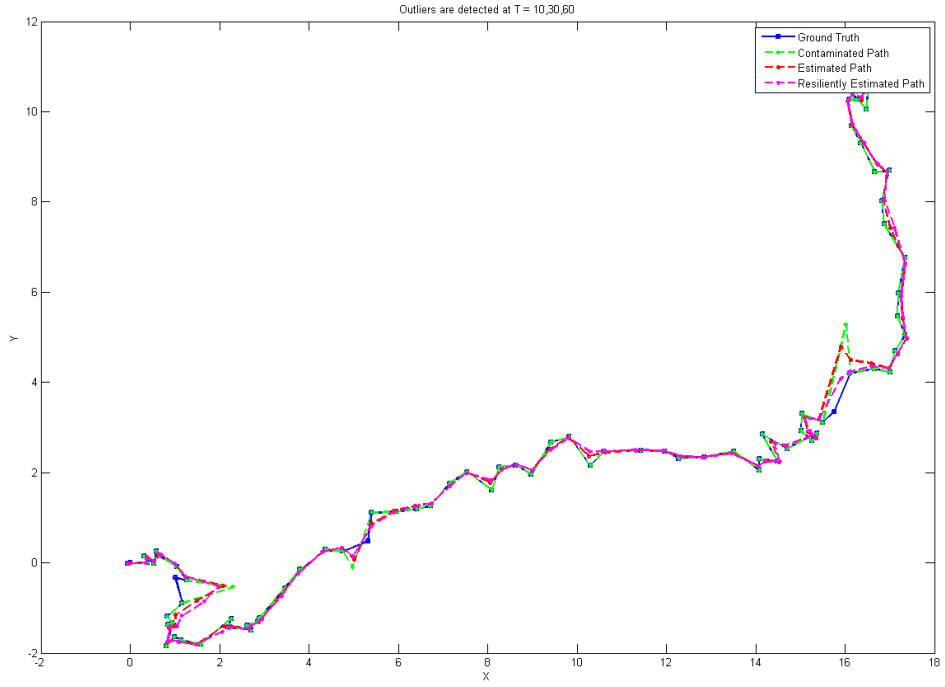
As Pearson discussed in [209], the MT-filter used in this work can be inapplicable when the covariance matrix on which the Kalman filter is based becomes singular. One way to deal with singular covariance matrices for the Kalman filter is to use *Singular Value Decomposition* [61, 283].

6.6 Discussion

The deployment of large-scale WSN profoundly changes the operation of SCADA systems. While such advancement facilities convenience of efficiency, it also exposes SCADA systems and WSN to more potential of uncertainty if the reliability and security aspect is not well addressed. We start the first steps, namely the resilient estimation, towards the concept and realization of the resilient control, which stipulates to maintain the optimality of standard operations under nominal conditions and to adapt abnormal situations through alleviating their impact. We also present an online robust outlier detection scheme that is optimal according to a stringent performance measure. Furthermore, this is accomplished without incurring large overhead. Future work lies in the direction of implement these methodologies on real data.



(a)Detection of 3 outliers



(b)Detection of 4 outliers

Figure 6.5: Detection of Multiple Outliers

Chapter 7

Revisit Dynamic ARIMA-Based Anomaly Detection

A detailed application of RGLRT is given out in this chapter. The time series model of Autoregressive Integrated Moving Average (ARIMA) progress, finds its wide usage in natural, social, economic and network applications. Model building and anomaly detection based on such models are often a first and important step towards monitoring unexpected problems and assuring the soundness and security of those systems being studied. The time variability by the coefficients in those dynamic regression models is particularly relevant and possibly indicative. Thus we introduce a corresponding framework and a novel anomaly detection approach based on the Kalman filter for identifying those dynamic models including their parameters and a General Likelihood Ratio (GLR) test for detecting suspicious changes in the parameters and therefore the models. We illustrate the idea through experiments and show its promising potential in terms of accuracy and robustness.

The most popular time series technique is the Autoregressive Integrated Moving Average (ARIMA) [37, 106, 36, 39] model due to its versatility in capturing dynamics and forecasting predictions. In light that model building lays the foundation for anomaly detection [158], consequently a fair share of the work on machine learning, signal processing and time-series analysis is devoted to detecting outliers or anomalies in time-series and ARIMA to be specific [237]. The existence of anomalies in ARIMA models and their detection arise in a variety of settings including but not limited to natural [108, 184], social [63, 273], economic [197, 73, 163, 8] and network service [281, 288] and network security [151, 291, 284, 91, 231] applications. The time varying structural parameters not only possibly challenge the model fidelity [264] thus undermine the intended effectiveness of its usage but also likely reflect the intrinsic nature of the system that evolves over time [203]. More specifically, any sudden change of these parameters is an indication of some atypical behavior within the system including benign faults [25] and/or malicious attacks [131]. In particular, in the arena of network security, network traffic anomalies may occur due to security threats such as Distributed Denial of Service (DDoS) attacks and network worms.

The work on *network anomogrphy* [291] by Zhang *et al.* inspired our extension. According to their investigation, one of the most successful and robust methods in detecting network traffic anomalies combining Box-Jenkins modeling (ARIMA) with L_1 norm minimization.

CUSUM (Cumulative Summation) method and its variants are widely used for anomaly detection. As pointed out in [25, 254], its major drawback is that it requires *a prior* knowledge on information after change, i.e. the intensity of the anomaly etc. But in practice, such information are not predicable.

We look at the problem through a novel angle and take advantage of the by-product due to the parameter learning and estimation process in the ARIMA model building stage to pre-screen possible anomalies without incurring extra drastic computation burden. It also prevents those anomalies from poisoning the correct model- and baseline-building from the start.

Our goal is to find a quick way to detect such anomalies manifested in the form of change in the system model. The identification and estimation of ARIMA models' parameters is often the first step before any further analysis and often can be achieved through maximum likelihood estimation. The exact likelihood is computed via a state-space representation of the ARIMA process, and the innovations and their variance found by a Kalman filter [139]. We use a *General Likelihood Ratio* (GLR) test [277, 25], which doesn't require any *a prior* knowledge of the anomalies, for detecting suspicious changes in the parameters and therefore the models. Along with the Kalman filter [139], this GLR procedure also adaptively filters the ARIMA parameter estimation in case of missing anomalous observations.

Organization of the paper: We first review the procedure of ARIMA-based anomaly detection in Section 2 with emphasis on the model-building and its transition to a state space model in which the Kalman filter that facilities model estimation and anomaly detection. In Section 3 we describe the GLR test for identifying sudden change in dynamic ARIMA model. Then we illustrate the idea through simulation experiments in Section 4 before conclude in Section 5.

7.1 ARIMA Modeling

While we address the derivation of model-building through a concrete example of anomaly detection on the network level, it's worth pointing the methodology is applicable to other situations.

The link traffic and Origin-Destination (OD) traffic matrix follow

$$b_j = A_j x_j \tag{7.1}$$

where A_j is an $n \times m$ routing matrix, x_j is a length-n vector of unknown OD flow traffic volumes, and b_j is a length-m vector of link loads[1], at time interval j.

If we first assume that the routing matrices A_j are time-invariant and are denoted by A. Then we can combine all t linear systems (7.1) into a single equation

$$B = AX, \tag{7.2}$$

where $B = [b_1 b_2 \cdots, b_t]$ is link traffic data over time t by having b_j as its column vectors, and similarly $X = [x_1 x_2 \cdots, x_t]$.

[1]Note that the link load vector b_j also includes the aggregated traffic at different ingress/egress points; the corresponding rows in A_j encode the OD flows that enter/exit the network at these points.

In the notation introduced by Box and Jenkins [37], models are summarized as ARIMA(p,d,q). A model described as ARIMA(0, 1, 2) means that it contains p = 0 (zero) autoregressive parameters and q = 2 moving-average parameters which were computed for the time series after it was differenced once (d = 1).

7.1.1 Time Series Expression

A general ARIMA model of order (p, d, q) can be expressed as:

$$z_k - \sum_{i=1}^{p} \phi_i z_{k-i} = e_k - \sum_{j=1}^{q} \theta_j z_{k-j} \tag{7.3}$$

where z_k is obtained by differencing the original time series d times (when $d \geq 1$) or by subtracting the mean from the original time series (when $d = 0$), e_k is the forecast error at time k, $\phi_i (i = 1, ..., p)$ and $\theta_j (j = 1, ..., q)$ are the autoregression and movingaverage coefficients, respectively. Let I denote the $t \times t$ identity matrix, ∇ denote the backshift matrix and $\mathbf{1}$ denote the $t \times t$ unity matrix with each entry = 1.

$$Z = \begin{cases} B(I-\nabla)^d \quad \nabla = \begin{smallmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & \dots & & \\ 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \end{smallmatrix}, & d \geq 1 \\ B(I - \frac{1}{t}\mathbf{1}) & d = 0 \end{cases} \tag{7.4}$$

$$E = BT, \text{where the transformation matrix} \tag{7.5}$$

$$T = \tag{7.6}$$

$$\begin{cases} (I-\nabla)^d (I - \sum_{i=1}^{p} \phi_i \nabla^i)(I - \sum_{j=1}^{q} \theta_j \nabla^j) & d \geq 1 \\ (I - \frac{1}{t}\mathbf{1})(I - \sum_{i=1}^{p} \phi_i \nabla^i)(I - \sum_{j=1}^{q} \theta_j \nabla^j) & d = 0 \end{cases}$$

In terms of the classical ARIMA techniques used for anomaly detection , the forecast errors indicate anomalous link traffic, $\tilde{B} = E$. That is, traffic behavior that cannot be well captured by the model is considered anomalous.

7.1.2 State-Space Representation

The discrete time linear dynamical system and measurement model are the following, where i is the index of sensors.

$$x_{t+1} = A_t x_t + w_t \tag{7.7}$$

$$y_t = C_t x_t + v_t \tag{7.8}$$

where $x_t \in \Re^s$ is the state vector, $y_t \in \Re^o$ is the output vector, $w_t \in \Re^s$ is white Gaussian noise with zero mean and covariance $Q > 0$ and v_t's $\in \Re^o$ are white Gaussian noises with covariance

$R_t > 0$. w_t and v_t's are independent. The initial system state x_0 is Gaussian with zero mean and covariance $\sum_0$. We assume x_0 is independent of w_t and v_t's.

7.1.3 The ARIMA(p,d,q) Process in a State-Space Model

Harvey and Pierse [107] derive a state-space representation of a general ARIMA(p,d,q) model with backshift operator L to denote the effect of $(Lz)_k = z_{k-1}$, then

$$\phi(L)\Delta^d y_t = \psi(L)\varepsilon_t$$

Let $r = max(p, q+1)$, the state transition equation can be written as a $(r+d)\times 1$ system

$$\begin{aligned} \mathbf{x}_t &= A\mathbf{x}_{t-1} + B\varepsilon_t \qquad (7.9)\\ &= \begin{bmatrix} \blacksquare & \mathbf{0}_{r\times d} \\ 10\cdots 0 & \delta\cdots\delta \\ \mathbf{0}_{d-1\times r} & \mathbf{I}_{d-1} : 0 \end{bmatrix} \mathbf{x}_{t-1} + \begin{bmatrix} \theta \\ \mathbf{0}_{d\times 1} \end{bmatrix} \varepsilon_t \end{aligned}$$

where

$$\blacksquare = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & 0 & \ddots & 0 \\ \phi_{r-1} & 0 & \cdots & 0 & 1 \\ \phi_r & 0 & \cdots & 0 & 0 \end{bmatrix}, \quad \psi = \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-2} \\ \theta_{r-1} \end{bmatrix}$$

and $-\delta_j$ is the coefficient on L^j in the expansion of $\triangle^d = (\mathbf{1}-L)^d$. This state space representation has $p+q+1$ hyperparameters and a measurement equation given by

$$\begin{aligned} y_t &= Cx_t \qquad (7.10)\\ &= [1\mathbf{0}_{1\times r-1}\delta_1\cdots\delta_d]x_t \qquad (7.11) \end{aligned}$$

7.1.4 Kalman Filter based Exact Maximum Likelihood Estimation of ARIMA

The Kalman filter [139] is a recursive algorithm for generating Minimum Mean Square Error (MMSE) predictions in a state space model. The state space representation is a very general formulation for linear models and it enables the Kalman filter to deal with time varying parameters, measurement errors and missing observations easily. As a by-product, if Gaussian errors are assumed, the filter allows the computation of the log-likelihood function of the state space model. This allows the model parameters to be easily estimated by maximum likelihood methods.

Standard Kalman Filter

$$\begin{aligned}
\hat{x}_{t|t} &= \mathbb{E}[x_t|\mathbf{y_t}] \\
P_{t|t} &= \mathbb{E}[(x_t - \hat{x}_{t|t})(x_t - \hat{x}_{t|t})^{'}|\mathbf{y_t}] \\
\hat{x}_{t+1|t} &= \mathbb{E}[x_{t+1}|\mathbf{y_t}] \\
P_{t+1|t} &= \mathbb{E}[(x_{t+1} - \hat{x}_{t+1|t})(x_{t+1} - \hat{x}_{t+1|t})^{'}|\mathbf{y_t}] \\
\hat{y}_{t+1|t} &= \mathbb{E}[y_{t+1}|\mathbf{y_t}].
\end{aligned}$$

where $P_{t+1|t}$ is the covariance matrix of the estimation.

The Kalman filter comprises two steps.

The **prediction phase** for $\hat{x}_{t+1|t}$ and $P_{t+1|t}$ of the Kalman filter is independent of the observation process with :

$$\hat{x}_{t+1|t} = A\hat{x}_{t|t} \tag{7.12}$$

$$P_{t+1|t} = AP_{t|t}A^{'} + Q \tag{7.13}$$

For the **update phase** of the Kalman filter, given the *residual* or *prediction error*

$$\tilde{e}_t = y_{t+1} - C\hat{x}_{t+1|t} \tag{7.14}$$

and its estimated variance

$$F_t = C_t P_{t+1|t} C_t^{'} + R_t \tag{7.15}$$

$$\begin{aligned}
\hat{x}_{t+1|t+1} &= \hat{x}_{t+1|t} + P_{t+1|t}C^{'}F_t^{-1} \\
&\quad (y_{t+1} - C\hat{x}_{t+1|t}) \qquad (7.16) \\
P_{t+1|t+1} &= AP_tA^{'} + Q - P_{t+1|t}C^{'}F_t^{-1} \\
&\quad CP_{t+1|t} \qquad (7.17)
\end{aligned}$$

7.1.5 The Log-likelihood function

Assuming that the noises are normally distributed, the log-likelihood function for the model can be computed from the residual, prediction error $\tilde{e}_t$ and its associated variance F_t

$$\begin{aligned}
LL &= -\frac{nT}{2}\log(2\pi\sigma^2) - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| \\
&\quad -\frac{1}{2\sigma^2}\sum_{t=1}^{T}(\tilde{e}_t)^{'}F_t^{-1}\tilde{e}_t \qquad (7.18)
\end{aligned}$$

Due to the fact that

$$\begin{aligned}\frac{\partial LL}{\partial \sigma^2} &= \frac{nT}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{t=1}^{T}(\tilde{e}_t)' F_t^{-1}\tilde{e}_t \\ &= 0,\end{aligned}$$

we have

$$\tilde{\sigma}^2 = \sum_{t=1}^{T}\frac{(\tilde{e}_t)' F_t^{-1}\tilde{e}_t}{nT}$$

.

Thus the concentrated log-likelihood function of the model can be maximized with respect to (ϕ, θ) to find the Maximum Likelihood Estimate (MLE) of the hyperparameter θ

$$\begin{aligned}LL^*(\phi,\theta) &= n\log S(\phi,\theta) + \sum_{t=1}^{n}\log f_t \qquad (7.19)\\ &= -\frac{nT}{2}\log(2\pi) - \frac{nT}{2} - \frac{1}{2}\sum_{t=1}^{T}\log|F_t| \\ &\quad -\frac{nT}{2}\log(\sum_{t=1}^{T}\frac{(\tilde{e}_t)' F_t^{-1}\tilde{e}_t}{nT}) \qquad (7.20)\end{aligned}$$

Smoothing. Based on all information available up to time $t-1$, the Kalman filter can function as a smoother with above mentioned recursions work backwards in time to smooth the regression model [106].

7.1.6 Identification of ARIMA and Model Estimation

Let I be the set of indices corresponding to all the ingress points in the link load vectors b_i. The series of subvectors b_i^I will be the input data for model selection and parameter estimation [2].

[2]Note this choice is due to their ready avilability and the fact that ingress traffic is largely invariant to internal topology and routing changes.

Choice of the degree of differencing d^*

Given that the optimal degree of differencing is often the one at which the standard deviation of the differenced series is the lowest [60], we carry out the following steps $\forall d \in \{0,1,2,3,4\}$

$$Z_d = [z_{d,i}]_{i=1}^{t}(1-L)^d[b_i^I]_{i=1}^{t} \tag{7.21}$$

$$E[Z_d] = \frac{1}{t}\sum_{i=1}^{t} z_{d,i} \tag{7.22}$$

$$Var[Z_d] = \frac{1}{t}\sum_{i=1}^{t} |z_{d,i} - E[Z_d]|_2^2 \tag{7.23}$$

$$\text{then} \quad d^* = \underset{d}{argmin}\, Var[Z_d] \tag{7.24}$$

Estimate ϕ and θ given (d^*)

Provided (p,d,q) and input vector series $\{b_k^I\}$, we can estimate the autoregression and moving-average coefficients ϕ_i and θ_j by constructing a state-space model as (7.10) in Section 7.1.3 and then applying the Kalman filter procedure as in Section 7.1.4 to compute the maximum log-likelihood function $LL^*(\phi,\theta)$ (7.20) for each $(p,q) \in 0,1,2,3,4$.

Selection on Model Order (p,q)

Information based criteria are designed to achieve a good balance between model parsimony and low prediction error [39, 60] such as *Akaikefor Information Criterion* (AIC) or *Bayesian information criterion* (BIC). we use AIC as our model selection criterion, which generally is

$$AIC = 2k - 2\ln(LL^*(\phi,\theta)) \tag{7.25}$$

where k is the number of parameters in the statistical model, and $LL^*(\phi,\theta)$ is the maximized value of the likelihood function for the estimated model (7.20). For each $(p,q) \in 0,1,2,3,4$ we estimate ϕ and θ (as in Section 7.1.6) and compute the resulting AIC based on the residuals and the model complexity. We then choose the pair of (p, q) with the lowest AIC.

$$(p,q)^* = \underset{(p,q)\in 0,1,2,3,4}{argmin} \quad AIC \tag{7.26}$$

7.2 Generalized Likelihood Ratio Test for Identifying Sudden Change in Dynamic ARIMA Model

Willsky and Jones (1976) [277] introduced the Window-limited GLR rules in the context of detecting abrupt additive system changes in linear state-space models. Such abstract system changes may occur due to benign environmental changes or unintentional system component faults or malicious activities. The idea is to implement a Kalman filter based on the assumption of no abrupt

system changes, and to monitor the measurement residuals of the filter to determine if a change has occurred and adjusts the filter accordingly.

Recall the state-space stochastic linear dynamical system (7.7) and measurement model (7.8) in Section 7.1.2, if at an unknown time τ the system undergoes additive changes in the sense that $u_t' 1_{\{t \leq \tau\}}$ is added to the right-hand side of (7.7), i.e.

$$x_{t+1} = A_t x_t + w_t + u_t' 1_{\{t \leq \tau\}}$$

then the innovations are still independent Gaussian vectors with covariance matrices F_t, but their means $m_t = E(\tilde{(e)}) = \rho(t,\tau)\eta$ for $t \geq \tau$ instead of the baseline values $m_t = 0$ for $t < \tau$. After the initialization of their associated $\rho(k,k) = 0$, $\alpha(k,k) = 0$, $\beta(k-1,k) = 0$, the matrices $\rho(t,k)$ can be evaluated recursively for $t \geq k$ through the following steps:

$$\begin{aligned} \alpha(t+1,k) &= A_k \alpha(t,k) + I \\ \beta(t,k) &= A_{k-1}\beta(t-1,k) + P_{t|t-1} C_k^T F_k^{-1} \rho(t,k) \\ \rho(t+1,k) &= C_{t+1}(\alpha(t+1,k) - A_t \beta(t,k)) \end{aligned} \tag{7.27}$$

7.2.1 Detection Rules

Without assuming any prior knowledge of parameter η, the GLR rule maximizes the log likelihood ratio over a window of inputs and decide the time to raise an alarm according to the following rule,

$$\begin{aligned} N_G &= \inf\{n : \max_{n-M \leq t \leq n-M'} \sup_{\eta} \sum_{i=k}^{n} \log[f(F_i^{-1/2} \\ &\quad \times (e_i - \rho(i,t)\eta))/f(F_i^{-1/2} e_i)] \geq c_\lambda\} \\ &= \inf\{n : \max_{n-m \leq t \leq n-m'} (\sum_{i=k}^{n} \rho^T(i,t) F_i^{-1} e_i)^T \\ &\quad \cdot (\sum_{i=k}^{n} \rho^T(i,t) F_i^{-1} \rho(i,t))^{-1} \\ &\quad \cdot (\sum_{i=k}^{n} \rho^T(i,t) F_i^{-1} e_i)/2 \geq c_\lambda\} \end{aligned} \tag{7.28}$$

where $f(y) = e^{\frac{-\|y\|^2}{2}}/(2\pi)^{\zeta/2}$ denotes the ζ-dimensional normal density, $\zeta = dim(\eta)$, and $m' + 1 \geq \zeta$ so that the matrix inversions in (7.28) are valid.

In essence, we are looking at an **optimal stopping time problem**:not to stop too early to produce a false alarm nor to stop too late to miss a real anomalous event.

7.2.2 Threshold and Window size Choice

Note that (7.27) computes $\rho(t,k)$ recursively over the each window. How to optimally choose $M, \tilde{M}$ and c_λ in general is a difficult problem [25] for online practices particularly due to the coupling effect between the threshold and window size on the asymptotical performance of the detection rule. But for off-line operations, the choice of window size is less demanding as all the data set is available, it's only a matter of computation time.

The threshold c in the rule N_W subject to the false alarm probability criterion $P_0(N_W \leq m)$ can be computed by using Monte Carlo computation of $P_0(N_W$ together with the method of successive linear approximation combined with bisection search for iterative solution of the equation $P_0(N_W \leq m)$.

7.3 Experiments

Given that ARIMA data sets share the commonality in the perspectives of basic model characteristics and in the interest of time and access, at current stage we've used two small publicly available ARIMA time series datasets [53, 57] besides simulation data and synthetic anomaly generation to test our method.

In order to broaden the scope of anomalies, we inject synthetic ones into the data set in a fashion similar to [254].

- By smoothing the original signal, we extract the long-term statistical trend from the data set.
- Add Gaussian noise to the smoothed signal.
- Add different anomaly combinations in terms of number, time, strength.

As shown in Figure 7.1, the synthetic dataset captures the trend in the original dataset and provides the simulation with more plausibility.

7.3.1 Detection Rates

For the real ARIMA dataset, we adjust the portion of the dataset being investigated by the detection algorithm as a way to control the occurrence of the anomalies. Whereas for the synthetic dataset, the number or size of the anomalies is easily controlled by the dosage of artificial anomalies that we inject into the synthetic dataset. Note that the synesthetic basically is considered anomaly free before any injection as it's a product of smoothing and de-noising of the original dataset. When using the synthetic dataset, each result is based on 1000 simulations.

Sensitivity to Window Size Although theoretically all window sizes can be computed precisely, we still would like to observe how they affect the performance of detection. Without analytically specify a precise window size to achieve the asymptotical optimality, there's a tradeoff between the window size and the detection sensitivity. When window size is too long, the recursive Kalman filtering itself may graduate smooth out the edginess of the anomaly. While window size is too small, the maximization requirement associated with the general likelihood may be met less than sufficiency.

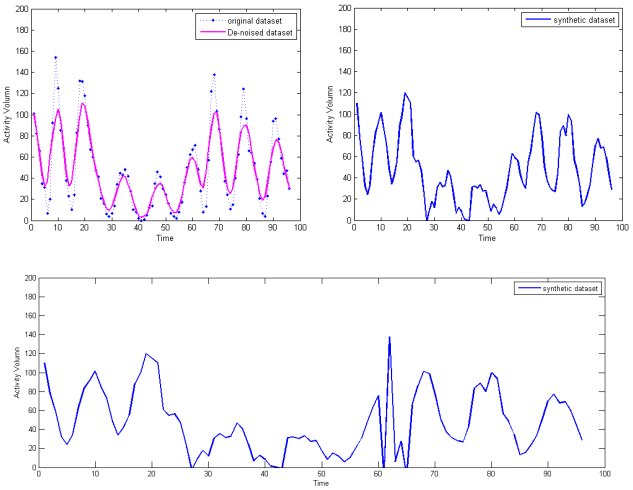


Figure 7.1: Steps for synthetic generation of anomaly where the last panel is the synthetic data with anomaly injected at time period from 60 to 65.

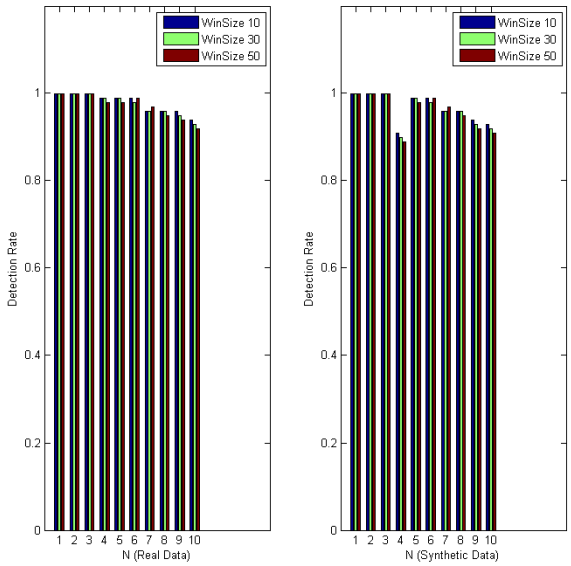


Figure 7.2: Detection Rate (with different window size) in response to the anomaly size N

Note for the synthetic dataset shown in Figure 7.2, when the anomaly size is 4, the detection performance seems to downgrade quite a bit. The likely explanation is that we lump 3 anomalies close together while keep them quite separate in other size cases.

Sensitivity to Threshold

Similarly, it's interesting to verify how sensitive the detection rate can be under the influence of the threshold chosen for the detection rule. As shown in Figure 7.3, we pick an arbitrary threshold

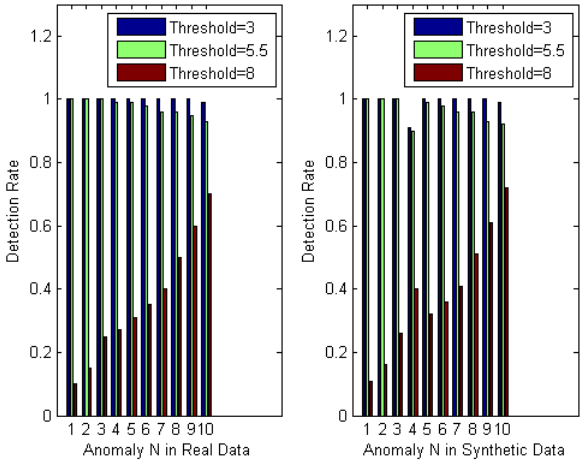


Figure 7.3: Detection Rate (with different threshold) in response to the anomaly size N

at 8 to contrast the other two cases, of which 5.5 is the value calculated through formal derivation and the same value used for testing on detection sensitivity to window size in Section 7.3.1. When threshold is too high (as the case of 8 here), so would miss detection rate. Note that for the case of 3 anomalies close together, it somehow made the high threshold case work better on the synthetic dataset than on the real dataset where the 3 anomalies are rather isolated. Also when the threshold is too low (as the case of 3 here), so would false alarm rate.

7.3.2 Detection Delay

Obviously our method has at least minimum window-length delay in issuing in alarms. This is due to the fact that at every time step, it requires a maximization over window-length data points in order to calculate the generalized likelihood in exchange for not demanding for any *a priori* knowledge of the potential anomalies.

Sensitivity to Anomaly Strength: When using the synthetic dataset with injected anomalies, we notice that the proposed Kalman-GLR scheme has longer mean detection delay (and is more prone to false alarms when detect anomalies using smaller threshold). In Figure 7.4, the mean delay time beyond 100 means it's in fact a miss detection as the magnitude of the anomaly is to weak to be detected.

7.4 Discussion

In this chapter, we describe the comprehensive procedure of building an ARIMA model and propose to identify anomalies during the process of model parameter estimation with the aid from the Kalman filter and GLR test. This approach also prevents such anomalies from poisoning the baseline-building.

Next step we plan to test out the robust methodology developed in [308]. Furthermore we'd

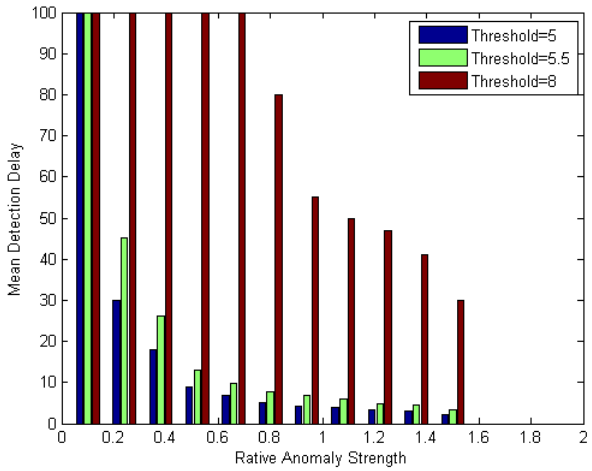


Figure 7.4: Mean Detection Delay (under different threshold) in response to the anomaly size N

like to apply our method to traffic data collected from the Abilene network [1] to study towards the simplification of threshold-setting.

Chapter 8

Anomaly Detection for Clean Energy Resources Prediction and Power Consumption Forecast in the Smart Grid

A tale of two cities

This chapter shows further development of RLRT and its application that is closely related to anomaly detection SCADA systems and smart grids, i.e. anomaly detection for both clean energy resources prediction and power consumption forecase [303] . The advancement in computing and hardware technologies ushers in a new era. While the utilization of clean energy resources including wind and solar power sets to grow from filling the gap of peak hours to taking a larger share in the upcoming smart grid and efficient infrastructure, the price-incentivized electricity consumption shall alleviate peak hours and reduce power outages. But anomalies including both benign faults and malicious attacks threat the reliability and availability of the new grid. To address these duo problems, we aim from the angle of one fundamental technique used. The Autoregressive Integrated Moving Average (ARIMA) time series models play roles at both ends in this new ecosystem: namely, predicting the variable clean energy resource on the supply side and forecasting the flexible load demand on the consume side. Model construction and anomaly detection based on such models are often a first and important step towards monitoring unexpected problems and assuring the soundness and security of those systems being studied. The time variability of the coefficients in those dynamic regression models is particularly relevant and possibly indicative. Thus we introduce a corresponding framework and a novel anomaly detection approach based on a robustified Kalman Filter for identifying those dynamic models including their parameters and a Robust General Likelihood Ratio (RGLR) test for detecting suspicious changes in the parameters and therefore the models. Currently, the effectiveness and robustness of this method is shown through simulation. At two ends of the smart grid, both the clean energy resource supply and electricity power consumption require reliable and accurate predication.

Variable Clean Energy Resources Prediction With the integration of clean energy into electricity grids, it is becoming increasingly important to obtain accurate forecasts. Advancements in

wind and solar forecasting technology aim to make renewable energy reliability a reality. In particular, due to its versatility, building and applying the Autoregressive Integrated Moving Average (ARIMA) time series model enjoys its popularity among industrial and engineering applications such as wind power, solar energy level prediction and power grid load forecasting [113], [84]. For example, Kavasseri et al studied day-ahead wind speed forecasting using f-ARIMA models [141], Nielsen et al built a wind power prediction system that is based on ARIMA [200], [199]; Makarov et al from California *Independent System Operator* (ISO) wind generation and forecasting service deemed ARIMA as the persistence models suitable for the short term wing generation forecasting and real-time dispatch in the Grid Control Centers [170]; Milligan et al applied ARIMA models to both wind speed and wind power output [184]. For a more comprehensive and state-of-art survey on short-term prediction of wind power, interested readers please refer to [84].

ARIMA models also suit the needs of the demand side of smart grid.

Flexible Smart Grid Load Demand Forecast In general, ARIMA models address well the issue of high level short-term hourly load forecasting in traditional power grids [10]. Furthermore, ARIMA modeling techniques show their prowess in capturing the flexible and price-sensitive short-term hourly overall load demand response enabled by the deployment of smart grid [55]. Given that one of the key drivers of the deployment of smart grid, buildings consume approximately 73% of the total electrical energy in the United States [145], it's efficient to monitor down to the building-level electricity consumption. ARIMA models have been applied to building-related applications ranging from modeling building electricity consumption [198] and forecasting and controlling the peak demand in commercial buildings [114], to optimizing the operation of cold storage in a large building [146].

The ubiquitous integration of computers in the smart grid – in the generation, transmission, distribution and metering in homes also introduces malicious security risks besides benign faults throughout the system [143], [68]. Stuxnet [70], one of most sophisticated control system malware known to date, has become the game changer in the field, in terms of demonstrating the severity and therefore raising people's awareness of such issues [1] [274] as described by Falliere et. al at Symantec [70], As of April 21st. 2011, There are more than 50 new Stuxnet-like attacks discovered [194] that beckon threats to the *Supervisory Control and Data Acquisition* SCADA, the underlying control system of the smart grid. The resources of vulnerabilities can be generic and board. Thus our fault and threat model is impact-oriented. We analyze the consequence of their occurrences manifested in the data that would sway the model construction of both the clean energy resource supply and power consumption forecast without excluding the cases where the adversaries purposely poisoning the model construction.

The idea of ARIMA-based anomaly detection is based on whether the data deviate afar from the model predication. Thus the accuracy of the model construction itself is important.

Alternatively, CUSUM (Cumulative Summation) method and its variants are widely used for anomaly detection. As pointed out in [25],[254], its major drawback is that it requires *a priori* knowledge on information after change, i.e. the intensity of the anomaly etc. But in practice, such information are not predicable.

We look the problem through a novel angle and take advantage the by-product due to the pa-

[1] In McAfee's report [18], nearly half of those being surveyed in the electric industry said that they had found Stuxnet on their systems.

rameter learning and estimation process in the ARIMA model building stage to pre-screen possible anomalies without incurring extra drastic computation burden. It also prevents those anomalies from poisoning the correct model- and baseline-building from the start. We take precaution of the skewing and deviating effect of outliers on identifying procedures by applying robustifying measures and integrating a recursive variant of the M-estimator, a Huber function [119], into the Kalman filter [139] via an recursively reweighted least squares implementation. Our Robust General Likelihood Ratio test rectifies and cleans data upon both isolated and patchy outliers while maintain the optimality of the Kalman Filter under the nominal condition. Furthermore it can be theoretically shown that our procedures are of the quickest and optimal detection thus we can achieve the goal of 'nipping it in the bud'. The robust sequential testing bears optimal stopping time, i.e. asymptotically shortest detection delay time while maintaining lowest false alarm rate. For the interest of briefness, readers can refer to Chapter 6 and Chapter 7 for more details.

8.1 Experiments

8.1.1 Data Sets – Real Wind Power Data

The *Transmission Expansion Planning Policy Committee* (TEPPC) of the *Western Electricity Coordinating Council* (WECC) provided us with wind power data. Particularly, we use its CA2 location profile [2]A2 includes Westwind, Antelope and other substations in California) with 3570 MW capacity as of 2006, as shown in Fig. 8.1.

It's easy to identify that the difference order d is 1 as visually its autocorrelation plot shown in Figure 8.2.

Due to the non-stationarity in the raw data series, its mean and variance diverge as time proceeds.

8.1.2 Simulated Data

In order to illustrate the idea of the commonality shared by both the variable clean energy and power consumption in the perspectives of basic model characteristics and in the interest of time and access, without loss of generality, we decide to employ a simulated ARIMA data set as shown in Figure 8.3.

8.1.3 Fogies Attack

An attacker can manipulate the data through means such as protocol defects, social engineering, man-in-the-middle attacks etc. SCADA and smart grid specific attacks [296] to accomplish their goals.

Random outliers are injected into the data set randomly to capture this effect as shown 8.4.

[2]C

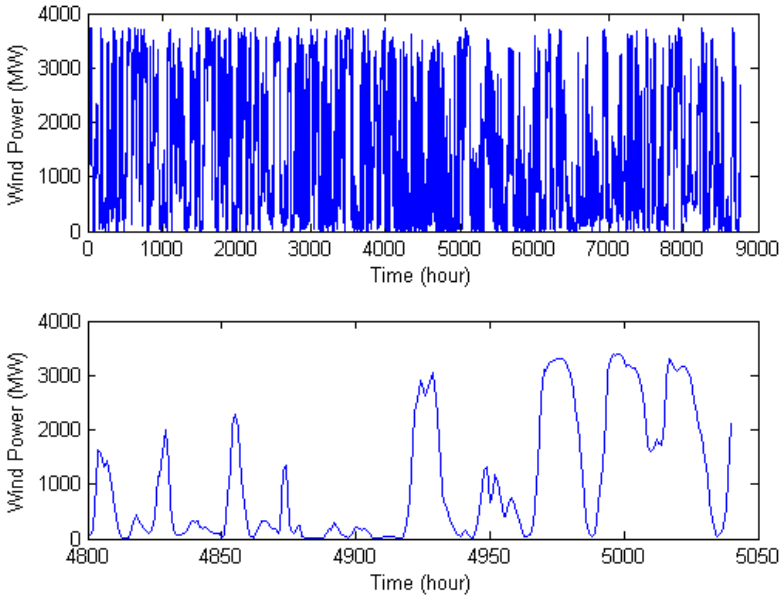


Figure 8.1: Wind Power Hourly Measurements: (Up) 2006 Whole Year, (Bottom) 10 days of Midsummer 2006.

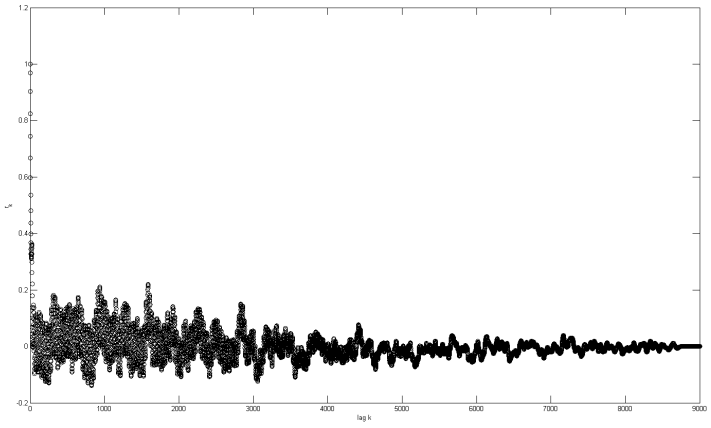


Figure 8.2: The Autocorrelation Plot

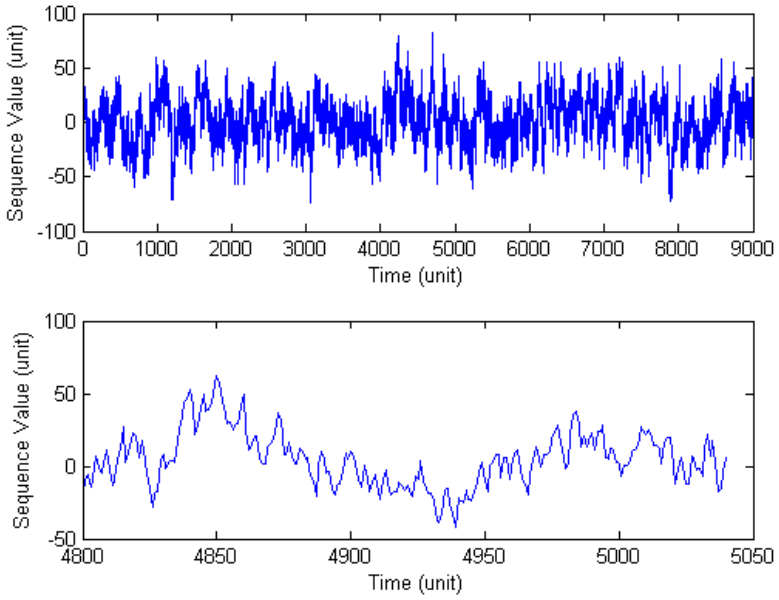


Figure 8.3: Simulated ARIMA Data: (Up) One Year, (Bottom) 10 days of Midsummer .

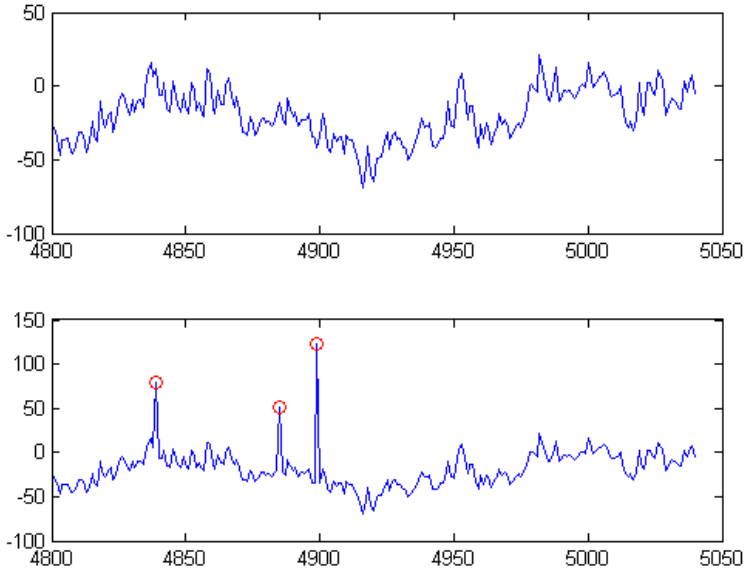


Figure 8.4: Simulated ARIMA Data: (Up) 10 days of Midsummer, (Bottom) With Outliers .

8.1.4 Countermeasure strategy – Parry

In light of the stealthiness of Stuxnet and the long-term hazard of a deviated baseline launched by likely furtive attackers, the main of our work can serve as a prevention measure in the sense that we take precaution during the model-building stage to prevent attackers from landing their intrusions earlier on[3] .

Given that ARIMA data sets share commonality in the perspectives of basic model characteristics and in the interest of time and access, at the current stage we've used two small publicly available ARIMA time series datasets [53, 57] besides simulation data and synthetic anomaly generation to test our method.

With randomly injected outliers where the false alarm constraint is achieved through Monte Carlo simulation, our approach successfully detects them.

8.1.5 Performance Analysis

Comparison with GLR

Given that GLR is based on the standard Kalman filter, assuming the dynamics after change also follow Gaussian. GlR doesn't function well at all when outliers are injected into the raw data sequence.

8.2 Discussion

With the ever rising demand of clean energy and fast increasing deployment of smart grid on the horizon, the generic nature of this study and investigation shows a promising utility in proactively suggesting a feasible solution to anomaly detection including benign faults and malicious attacks for both variable clean energy resource supply and flexible power consumption. Next step we plan to apply it to real wind data in conjunction with simulated user demand sensitive to pricing.

[3] In fencing, the primary function of a *parry* is to prevent an opponent's attack from landing

Chapter 9

Conclusion and Future Plans

In this dissertation, the landscape of cyber attacks and intrusion detection systems for SCADA systems has been clearly outlined. As an initial effort, an in-depth SCADA-specific security solution *Xware* is proposed. A versatile early detection scheme RGLRT along with resilient estimation approach shows its effectiveness in detecting anomalies.

9.1 RGLRT

The strength of RGLRT lies in that it does *not* require *a priori* knowledge of the distributions of the attacks or benign anomalies, i.e., neither their mean nor their variance, which is a clear advantage against *SPRT* in real life. Furthermore its close relation with the state space setting and the Kalman filter gives it a special advantage against *non-parametric CUSUM* in the engineering field. I've explored two main types of its application, namely

- to detect outliers and anomalies through measurements in the Kalman filter when the latter is used for predication and estimation of a dynamical model ;
- to detect outliers and anomalies in the parameters of a model, ARIMA, to be specific, by way of states variables in the Kalman filter when the latter is used to do parameter estimation.

How to expand the application range of the RGLRT is the next step that I am pursuing. Practically, the task of simplifying the window size selection is still worth more consideration.

9.2 Resilient Control

So far, this dissertation works has shown the promise of resilient estimation and the potential of resilient control. Much theory development is needed in the niche of resilient control verse the conventual robust control and minimax approach. With smart grids and the new intelligent infrastructure on the horizon, the concept of resilient control has profound meaning and impact on the development technicality as well.

9.3 Network Intrusion Detection

Network intrusion detection research for SCADA systems to date has been quite limited, with the three most prominent and critical deficiencies being: the lack of a well-considered threat model; the absence of addressing false alarm and false negative (mis-detection) rates; and the need to empirically ground the development of IDS mechanisms in the realities of how such system operate in practice, including the diversity of traffic they manifest and the need to tailor IDS operation to different SCADA environments. To this end, I focus on developing flexible, comprehensive SCADA-oriented IDS analysis; I do not endeavor to provide rigorous, all-encompassing SCADA security.

I will begin with considering how to effectively categorize cyber attacks into taxonomies that illuminate the problem space, considering three distinct dimensions:– how attacks manifest in appearance as seen in network traffic (defense perspective);– how attacks are constructed and the accompanying resources required to realize them (attacker and prevention perspective);– the damage implications of different types of attacks (victim perspective). I next aim to capture the characteristics of a specific SCADA system under study (a segment of the power grid) with full situational awareness, including the dynamics of the physical plant being monitored, its communication patterns, system architecture, network traffic behavior, and specific application-level protocols used, ranging from the dominate Modbus/TCP and DNP3 to newer protocols such as WirelessHART and ISA100.

After study of this SCADA system, I will develop attack trees and derive from it prudent threat models. This will include consideration of evasion mechanisms attackers can employ in light of the applications in use (beyond those already known for TCP/IP). I will derive application-level protocol specifications and implementation specifics and from these construct analyzers for an open-source IDS. At the heart of this effort I envision development of ”normalcy checking,” i.e., a combination of techniques designed to capture two envelopes of possible system activity: (1) definitely safe operations and (2) definitely unsafe operations. When identifiable, the first of these can be safely ignored; the second merits immediate attention/blocking; and the middle ground between the two requires additional analysis. The first technique I will draw upon in this regard is specification-based intrusion detection that constructs the control system’s overall allowable behavior, i.e., as seen from the application level, and reflecting the monitored plant dynamics, including its valid extreme cases. The second uses encodings of misuse signatures and their possible variants. The third draws upon models derived from the control system’s formal dynamics; this aspect is *unique to the problem domain* and holds great promise for refining the scope to which I will apply the analysis. I will draw upon traces of live operation to develop and tune this system. I will incorporate our detection mechanisms into NIDS to realize an operational system, validating its efficacy using, first, commercial SCADA emulation software; then synthesized traffic created in the DETER testbed; then on new traces from the operational environments; followed by live ”shadow” operation. For our testbed, we will construct a test environment consisting of physical PLCs and IEDs to emulate the SCADA system under study, where we inject designed attack traffic along with traffic synthesize from traces separate from those used in developing and tuning the system in order to assess false positive and false negative rates. The final proof, necessarily, will come from prototype in situ deployment, which will require ongoing interactions with the

SCADA system's operational staff.

Bibliography

[1] Abilene network. *http://www.internet2.edu/network/*.

[2] Blaster worm linked to severity of blackout. *Computerworld*, August 2003.

[3] Safegurading scada systems with anomaly detection. *Computer Network Security*, 2276:171–182, 2003.

[4] Sql slammer worm lessons learned for consideration by the electricity sector. *Norh American Electric Reliability Council*, June 2003.

[5] Cybersecurity of pcs/scada networks: Half-baked homeland security, June 2006.

[6] *Contractor Pleads Guilty to SCADA Inrusion and Damage*. http://www.networkworld.com/news/2009/092309-contractor-pleads-guilty-to-scada.html, September 2009.

[7] J. Allen. State of the practice of intrusion detection technologies. Technical report, DTIC Document, 2000.

[8] S. Alnaa and F. Ahiakpor. Arima (autoregressive integrated moving average) approach to predicting inflation in ghana. *Journal of Economics and International Finance*, 3(5):328–336, 2011.

[9] A. Alouani, P. Xia, T. Rice, and W. Blair. On the optimality of two-stage state estimation in the presence of random bias. *Automatic Control, IEEE Transactions on*, 38(8):1279–1283, Aug 1993.

[10] N. Amjady. Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *Power Systems, IEEE Transactions on*, 16(3):498–505, 2001.

[11] B. Anderson and J. Moore. Optimal filtering. *Prentice-Hall Information and System Sciences Series, Englewood Cliffs: Prentice-Hall, 1979*, 1, 1979.

[12] R. Anderson. *Security Engineering: A guide to building dependable distributed systems*. Wiley Publishing, 2008.

[13] M. Aoki. *State Space Modeling of Time Series*. Springer-Verlag Berlin, Heidelberg, 1990.

[14] S. Axelsson. The base-rate fallacy and the difficulty of intrusion detection. *ACM Transactions on Information and System Security (TISSEC)*, 3(3):186–205, 2000.

[15] S. Axelsson. Intrusion detection systems: A survey and taxonomy. Technical report, Department of Computer Engineering, Chalmers University of Technology, Göteborg, Sweden, 2000.

[16] S. Axelsson. A preliminary attempt to apply detection and estimation theory to intrusion detection. Technical report, Department of Computer Engineering, Chalmers University of Technology, Göteborg, Sweden, 2000.

[17] M. Bagshaw and R. Johnson. Sequential procedures for detecting parameter changes in a time-series model. *Journal of the American Statistical Association*, 72(359):593–597, 1977.

[18] S. Baker, N. Filipiak, and K. Timlin. In the Dark Crucial Industries Confront Cyberattacks, year= 2011, publisher=McAfee report, journal=http://www.mcafee.com/us/resources/reports/rp-critical-infrastructure-protection.pdf.

[19] S. Baker, N. Filipiak, and K. Timlin. In the dark crucial industries confront cyberattacks, mcafee report, 2011.

[20] I. Balepin, S. Maltsev, J. Rowe, and K. Levitt. Using specification-based intrusion detection for automated response. In *Recent Advances in Intrusion Detection*, pages 136–154, Pittsburgh, PA, 2003. Springer.

[21] P. Barford, J. Kline, D. Plonka, and A. Ron. A signal analysis of network traffic anomalies. In *Proceedings of the 2nd ACM SIGCOMM Workshop on Internet measurment*, pages 71–82. ACM, 2002.

[22] M. Baron and A. Tartakovsky. Asymptotic optimality of change-point detection schemes in general continuous-time models. *Sequential Analysis*, 25(3):257–296, 2006.

[23] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar. Can machine learning be secure? In *ASIACCS '06: Proceedings of the 2006 ACM Symposium on Information, computer and communications security*, pages 16–25, New York, NY, USA, 2006. ACM.

[24] M. Basseville and A. Benveniste. Design and comparative study of some sequential jump detection algorithms for digital signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-31(3), JUNE.

[25] M. Basseville and I. Nikiforov. Detection of abrupt changes. *Theory and Applications*, page 1993.

[26] M. Basseville and I. Nikiforov. *Detection of Abrupt Changes: Theory and Applications*. Prentice-Hall, 1993.

[27] BCIT. Opc security whitepaper #2 opc exposed, May 2007.

[28] BCIT. Bcit industrial security incident database (isid), 2008.

[29] S. Bellovin. Packets found on an internet. *ACM SIGCOMM Computer Communication Review*, 23(3):26–31, 1993.

[30] I. E. Ben-Gal. *Outlier Detection*. Springer, 2005.

[31] A. Benveniste and M. Basseville. Detection of abrupt changes in signals and dynamical systems: Some statistical aspects. *Analysis and Optimization of Systems – Detection of Changes in Systems*, pages 143–155, 1984.

[32] D. Beresford. The sauce of utter pwnage, January 2011.

[33] G. Bianchi and I. Tinnirello. Kalman filter estimation of the number of competing terminals in an ieee 802.11 network, 2003.

[34] C. Bishop. Novelty detection and neural network validation. In *IEE Proceedings of Vision, Image Signal Process, Appl. Neural Networks*, volume 141, page 217222, 1994.

[35] M. Bowman, S. K. Debray, and L. L. Peterson. Reasoning about naming systems. *ACM Trans. Program. Lang. Syst.*, 15(5):795–825, November 1993.

[36] G. Box and G. Jenkins. *Time series analysis: forecasting and control*. Prentice Hall PTR, 1994.

[37] G. Box, G. Jenkins, and G. Reinsel. *Time series analysis*. Holden-day San Francisco, 1970.

[38] J. Braams. Babel, a multilingual style-option system for use with latex's standard document styles. *TUGboat*, 12(2):291–301, June 1991.

[39] P. Brockwell and R. Davis. *Introduction to time series and forecasting*. Springer Verlag, 2002.

[40] B. Brodsky and B. Darkhovsky. Asymptotically optimal sequential change-point detection under composite hypotheses. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC '05. 44th IEEE Conference on*, pages 7347–7351, December 2005.

[41] R. Brown, J. Durbin, and J. Evans. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 37(2):149–192, 1975.

[42] E. Byres, J. Carter, A. Elramly, and D. Hoffman. Worlds in collision: Ethernet on the plant floor. In *ISA Emerging Technologies Conference, Instrumentation Systems and Automation Society, Chicago*, 2002.

[43] E. Byres, D. Hoffman, and N. Kube. On shaky ground-a study of security vulnerabilities in control protocols. In *5th American Nuclear Society International Topical Meeting on Nuclear Plant Instrumentation, Controls, and Human Machine Interface Technology, American Nuclear Society, Albuquerque, USA*, 2006.

[44] E. Byres and J. Lowe. The myths and facts behind cyber security risks for industrial control systems. *VDE Congress*, 2004.

[45] I. S. C37.1-1994. Ieee standard definition, specification, and analysis of systems used for supervisory control, data acquisition, and automatic control. *The Institute of Electrical and Electronics Engineers, Inc*, 1994.

[46] F. Callier and C. Desoer. *Linear system theory*. Springer, 1991.

[47] H. Caussinus and A. Roiz. Interesting projections of multidimensional data by means of generalized component analyis. In *Proceedings in Computational Statistics*, pages 121–126. Physica-Heidelberg, 1990.

[48] P. Chen and et. al. Experiments in instrumenting wireless sensor networks for real-time surveillance. In *IEEE ICRA Video and Poster*, 2006.

[49] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes. Using model-based intrusion detection for scada networks. In *Proceedings of the SCADA Security Scientific Symposium*, pages 127–134. Citeseer, 2007.

[50] S. Cheung, B. Dutertre, M. Fong, U. Lindqvist, K. Skinner, and A. Valdes. Using model-based intrusion detection for scada networks. In *SCADA Security Scientific Symposium*, Miami Beach, Florida, January 2007.

[51] L. Chiang, E. Russell, and R. Braatz. *Fault detection and diagnosis in industrial systems*. Springer Verlag, 2001.

[52] M. Clark. Post congress tristesse. In *TeX90 Conference Proceedings*, pages 84–89. TeX Users Group, March 1991.

[53] G. Cobb. The problem of the nile: conditional solution to a changepoint problem. *Biometrika*, 65(2):243, 1978.

[54] P. Comon. Independent component analysis – a new concept? *Signal Processing*, (36):287–314, 1994.

[55] A. Conejo, J. Morales, and L. Baringo. Real-time demand response model. *Smart Grid, IEEE Transactions on*, 1(3):236–242, 2010.

[56] L. Davies and U. Gather. The identification of multiple outliers. *Journal of Amercian Statistical Association*, 88, 1993.

[57] P. De Jong and J. Penzer. Diagnosing shocks in time series. *Journal of the American Statistical Association*, 93(442):796–806, 1998.

[58] S. Dharmapurikar and V. Paxson. Robust tcp stream reassembly in the presence of adversaries. In *Proceedings of the 14th conference on USENIX Security Symposium-Volume 14*, pages 5–5. USENIX Association, 2005.

[59] DHS. Scada systems and the terrorist threat : protecting the nation's critical control systems : joint hearing before the subcommittee on economic security, infrastructure protection, and cybersecurity with the subcommittee on emergency preparedness, science, and technology of the committee on homeland security. *United States. Congress. House. Committee on Homeland Security. Subcommittee on Economic Security, Infrastructure Protection, and Cybersecurity*, (109-45), October 2005.

[60] F. Diebold. *Elements of forecasting*. Thomson, 2004.

[61] G. Dimitriu. Using singular value decomposition in conjection with data assimilation procedures. In *NMA 2006 LNCS 4310*, pages 435–442, Verlin, Heidelberg, 2007. Springer Verlag.

[62] P. Doe. This is a test test entry of type @MISC, June 2009.

[63] A. Dubrawski. Detection of events in multiple streams of surveillance data. *Infectious Disease Informatics and Biosurveillance*, pages 145–171, 2011.

[64] D. Dzung, M. Naedele, T. V. Hoff, and M. Crevatin. Security for industrial communication systems. *Proceedings of the IEEE*, 93(6):1152 – 1177, June 2005.

[65] B. Edelman. Netscape 8's "trust rating" system - screenshots. http://www.benedelman.org/spyware/ns8/, June 2005.

[66] C. Endorf, E. Schultz, and J. Mellander. *Intrusion detection & prevention*. McGraw-Hill Osborne Media, 2004.

[67] EPRI. Anomaly-based intrusion detection in scada (supervisory command and data acquisition.

[68] G. Ericsson. Cyber security and power system communicationessential parts of a smart grid infrastructure. *Power Delivery, IEEE Transactions on*, 25(3):1501–1507, 2010.

[69] A. N. Evgueni Gordienko and E. Zaitseva. Stability estimating in optimal sequential hypotheses testing. *Kybernetika, The Journal of the Czech Society for Cybernetics and Information Sciences*, 45(2):3 3 1 3 4 4, 2009.

[70] N. Falliere, L. O. Murchu, and E. Chien. W32. stuxnet dossier. *http://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf*, February 2011.

[71] S. Fan and R. Hyndman. Forecast short-term electricity demand using semi-parametric additive model. In *Universities Power Engineering Conference (AUPEC), 2010 20th Australasian*, pages 1–6. IEEE.

[72] W. Fan, M. Miller, S. Stolfo, W. Lee, and P. Chan. Using artificial anomalies to detect unknown and known network intrusions. In *Knowl. Inf. Syst.*, pages 507–527. Published by the IEEE Computer Society, 2004.

[73] D. Findley, B. Monsell, W. Bell, M. Otto, and B. Chen. New capabilities and methods of the x-12-arima seasonal-adjustment program. *Journal of Business & Economic Statistics*, 16(2):127–152, 1998.

[74] R. Fitzgerald. Divergence of the kalman filter. *Automatic Control, IEEE Transactions on*, 16(6):736 – 747, dec 1971.

[75] I. N. Fovino, A. Coletta, and M. Masera. Taxonomy of security solutions for the scada sector deliverable: D 2.2, version: 1.1. *A European Network For The Security Of Control And Real Time Systems*, March 2010.

[76] Q. Gan and C. Harris. Comparison of two measurement fusion methods for kalman-filter-based multisensor data fusion. *IEEE Transactions on Aerospace and Electronic Systems*, 37(1):273–279, January 2001.

[77] S. Ganeriwal, L. K. Balzano, and M.Srivastava. Reputation-based framework for high integrity sensor networks. *ACM Transactions on Sensor Networks*, V(N), September 2007.

[78] GAO. Critical infrastructure protection: Challenges and efforts to secure control systems. *No.: GAO-04-354*, page 47, 2004.

[79] GAO. Critical infrastructure protection challenges and efforts to secure control systems report to congressional requesters. Technical Report GAO-04354, United States Government Accountability Office, March 2004.

[80] GAO. Department of homeland securitys (dhss) role in critical infrastructure protection (cip) cybersecurity. *No.: GAO-05-434*, May 2005).

[81] GAO. Critical infrastructure protection multiple efforts to secure control systems are under way, but challenges remain, report to congressional requesters. *No.: GAO-07-1036*, September 2007.

[82] O. N.-T. Garcia. Security in embedded systems challenges and oportunities. In *International Conference on Emerging Security Information, Systems and Technologies, Securware07*, 2007.

[83] H. A. Gardner, G. and G. Phillips. An algorithm for exact maximum likelihood estimation of autoregressive-moving average models by means of kaiman filtering. *Applied Statistics*, 29(3):311–322, 1980.

[84] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, and C. Draxl. The state-of-the-art in short-term prediction of wind power: A literature overview. Technical report, ANEMOS. plus, 2011.

[85] M. A. Girshick and H. Rubin. A bayes approach to a quality control model. *The Annals of Mathmatical Statistics*, 23(1):114–125, 1952.

[86] D. Gizopoulos, M. Psarakis, and A. Paschalis. Robust sequential fault testing of iterative logic arrays. In *VLSI Test Symposium, 15th IEEE*, pages 238–244, AprMay 1997.

[87] J. Glaz, J. Naus, and S. Wallenstein. *Scan Statistics*. Springer, 2001.

[88] E. Grant and R. Leavenworth. *Statistical Quality Control*. McGraw-Hill, 1996.

[89] L. Greenemeier. Robots arrive at fukushima nuclear site with unclear mission. *Scientific American*, 2011.

[90] M. Grimes. Scada exposed, 2005.

[91] B. Gupta, M. Moorthy, and B. M.R. Analyzing data mining algorithms in sql server. *International Journal of Research and Reviews in Computer Science*, 2(3):670–675, 2011.

[92] A. S. Hadi. Identifying multiple outliers in multivariate data. *the Royal Statistical Society Series B (Methodological)*, 54(3):761–771, 1992.

[93] A. S. Hadi, A. H. M. R. Imon, and M. Werner. Detection of outliers–overview. *Computational Statistics*, 1(1), July/August 2009.

[94] U. Hammes. Robust positioning algorithms for wireless networks. 2010.

[95] U. R. Hammes. *Robust Positioning Algorithms for Wireless Networks*. PhD thesis, TU Darmstadt, February 2010.

[96] F. Hampel. Contributions to the theory of robust estimation. *Ph.D. dissertation, Dept. of Statistics, Univ. of California*, 1968.

[97] F. Hampel. A general qualitative definition of robusteness. *Annals of Mathmatics Statistics*, 42:1887–1896, 1971.

[98] F. Hampel. A general qualitative definition of robustness. *Annnals of Mathmatics Statics*, 42, 1971.

[99] F. Hampel. The influence curve and its role in robust estimation. *American Statistical Association*, 69:382–393, 1974.

[100] F. Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69, 1974.

[101] D. Han and F.Tsung. The optimal stopping time for detecting changes in discrete time markov processes. *Sequential Analysis*, (28):115–135, 2009.

[102] D. Han and F. Tsung.

[103] M. Handley, V. Paxson, and C. Kreibich. Network intrusion detection: Evasion, traffic normalization, and end-to-end protocol semantics. In *Proceedings of the 10th conference on USENIX Security Symposium-Volume 10*, pages 9–9. USENIX Association, 2001.

[104] S. Hansman and R. Hunt. A taxonomy of network and computer attacks. *Computers & Security*, 24(1):31–43, 2005.

[105] S. Harmeling, G. Dornhege, D. M. J. Tax, F. C. Meinecke, and K.-R. M"uller. From outliers to prototypes: Ordering data. *Neurocomputing*, 69(13-15):1608–1618, 2006.

[106] A. Harvey. *Forecasting, structural time series models and the Kalman filter*. Cambridge Univ Press, 1991.

[107] A. Harvey and R. Pierse. Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79(385):125–131, 1984.

[108] J. Haslett and A. Raftery. Space-time modelling with long-memory dependence: assessing ireland's wind power resource. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 38(1):1–50, 1989.

[109] D. Hawikins. *Identification of Outliers*. Chapman and Hall, London, 1980.

[110] D. Hawkins, Q. PEIHUA, and W. CHANG. The changepoint model for statistical process control. *Journal of Quality Technology*, 35(4):355–366, 2003.

[111] M. Herlihy. A methodology for implementing highly concurrent data objects. *ACM Trans. Program. Lang. Syst.*, 15(5):745–770, November 1993.

[112] D. V. Hinkley. Inference about the change-point from cumulaive sum tests. *Biometrika*, 58:509–523, 1971.

[113] S. Ho and M. Xie. The use of arima models for reliability forecasting and analysis. *Computers & industrial engineering*, 35(1-2):213–216, 1998.

[114] A. Hoffman. Peak demand control in commercial buildings with target peak adjustment based on load forecasting. In *Control Applications, 1998. Proceedings of the 1998 IEEE International Conference on*, volume 2, pages 1292–1296. IEEE, 1998.

[115] J. Howard. An analysis of security incidents on the internet 1989-1995. Technical report, DTIC Document, 1997.

[116] P. Huber. Robust estimation. *Selected Statistical Papers*, pages 3–25, 1968.

[117] P. Huber. Robust statistics: a review. *Ann. Math. Statist*, 43(3):1041–1067, 1972.

[118] P. Huber. *Robust Statistics*. Wiley, Hoboken, NJ, 2004.

[119] P. J. Huber. A robust version of the probability ratio test. *The Annals of Mathematical Statistics*, 36(6):1753–1758, December 1965.

[120] P. J. Huber and E. M. Ronchetti. *Robust statistics*. Wiley, Hoboken, N.J, 2nd edition, 2009.

[121] L. Hutwagner, E. Maloney, N. Bean, L. Slutsker, and S. Martin. Using laboratory-based surveillance data for prevention: an algorithm for detecting salmonella outbreaks. *Emerging Infectious Diseases*, 3(3):395, 1997.

[122] A. Hyvrinen. Fast and robust fixed-point algorithms for independent component analysis. *IEEE Transactions on Neural Networks*, 10(3):626–634, 1999.

[123] A. Hyvrinen, J. Karhunen, and E. Oja. *Independent Component Analysis*. John Wiley & Sons, 2001.

[124] A. Hyvrinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9:1483–1492, 1997.

[125] A. Hyvrinen, J. Srel, J. Ssrels, and R. Vigrio. Spikes and bumps: Artefacts generated by independent component analysis with insufficient sample size, 1999.

[126] IEC61850. Iec ts 61850: Power systems management and associated information exchange data and communications security. *Power systems management and associated information exchange–Data and communications security–Part*, 1 9, 2004.

[127] IEC62531. Iec ts 62351: Power systems management and associated information exchange data and communications security. *Power systems management and associated information exchange–Data and communications security–Part*, 1, 2007.

[128] M. Ignagni. Separate bias kalman estimator with bias state noise. *Automatic Control, IEEE Transactions on*, 35(3):338–341, Mar 1990.

[129] V. Igure, S. Laughtera, and R. Williams. Security issues in scada networks. *Computers &Security*, 25(7):498–506, October 2006.

[130] R. Isermann. Model-based fault detection and diagnosis-status and applications. In *IFAC*, 2004.

[131] R. Jana and S. Dey. Change detection in teletraffic models. *Signal Processing, IEEE Transactions on*, 48(3):846–853, 2000.

[132] J. Jeong and S. Lee. Outlier elimination method for robust visual servo control in complex environment. In *Robotics and Biomimetics (ROBIO), 2010 IEEE International Conference on*, pages 938 –943, dec. 2010.

[133] W. Jiang, S. Han, K. Tsui, and W. Woodall. Spatiotemporal surveillance methods in the presence of spatial correlation. *Statistics in Medicine*, 30(5):569–583, 2011.

[134] H. joo Lee and S. J. Roberts. On-line novelty detection using the kalman filter and extreme value theory. In *Proceeding of ICPR*, 2008.

[135] A. Jøsang and R. Ismail. The beta reputation system. 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy, June 2002.

[136] J. Jung. Real-time detection of malicious network activity using stochastic models. Technical report, 2006.

[137] J. Jung, V. Paxson, A. W. Berger, and H. Balakrishnan. Fast portscan detection using sequential hypothesis testing. In *Proceedings of the IEEE Symposium on Security and Privacy*, May 2004.

[138] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

[139] R. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, 1960.

[140] G. Kariniotakis, G. Stavrakakis, and E. Nogaret. Wind power forecasting using advanced neural networks models. *Energy Conversion, IEEE Transactions on*, 11(4):762–767, 1996.

[141] R. Kavasseri and K. Seetharaman. Day-ahead wind speed forecasting using f-arima models. *Renewable Energy*, 34(5):1388–1393, 2009.

[142] A. Kharin. *AUSTRIAN JOURNAL OF STATISTICS*, 37(1):5160, 2008.

[143] H. Khurana, M. Hadley, N. Lu, and D. Frincke. Smart-grid security issues. *Security & Privacy, IEEE*, 8(1):81–85, 2010.

[144] K. Killourhy, R. Maxion, and K. Tan. A defense-centric taxonomy based on attack manifestations. In *Dependable Systems and Networks, 2004 International Conference on*, pages 102–111. IEEE, 2004.

[145] Y. Kim, T. Schmid, M. Srivastava, and Y. Wang. Challenges in resource monitoring for residential spaces. In *Proceedings of the First ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 1–6. ACM, 2009.

[146] A. Kimbara, S. Kurosu, R. Endo, K. Kamimura, T. Matsuba, and A. Yamada. On-line prediction for load profile of an air-conditioning system. *ASHRAE TRANS*, 101:198–207, 1995.

[147] R. Klump, R. E. Wilson, and K. E. Martin. Visualizing real-time security threats using hybrid scada / pmu measurement displays. *Hawaii International Conference on System Sciences*, 2:55c, 2005.

[148] C. Ko, M. Ruschitzka, and K. Levitt. Execution monitoring of security-critical programs in distributed systems: A specification-based approach. *sp*, page 0175, 1997.

[149] C. Kose and R. Wesel. Robustness of likelihood ratio tests: hypothesis testing under incorrect models. In *Signals, Systems and Computers, 2001. Conference Record of the Thirty-Fifth Asilomar Conference on*, volume 2, pages 1738–1742, 2001.

[150] D. Kravets. Feds: Hacker disabled offshore oil platforms' leak-detection system, April 2009.

[151] B. Krishnamurthy, S. Sen, Y. Zhang, and Y. Chen. Sketch-based change detection: methods, evaluation, and applications. In *Proceedings of the 3rd ACM SIGCOMM Conference on Internet Measurement*, pages 234–247. ACM, 2003.

[152] C. Kruegel, D. Mutz, W. Robertson, and F. Valeur. Bayesian event classification for intrusion detection. In *Proceedings of the 19th Annual Computer Security Applications Conference (ACSAC '03)*. Published by the IEEE Computer Society, 2003.

[153] R. Krutz. *Securing SCADA Systems*. Wiley, Indianapolis, IN, 2006.

[154] T. Lai. Sequential analysis: some classical problems and new challenges. *Statistica Sinica*, 11(2):303–350, 2001.

[155] T. L. Lai. Information bounds and quick detection of parameter changes in stochastic systems. *Information Theory, IEEE Transactions on*, 44(7):2917–2929, Nov 1998.

[156] T. L. Lai. Sequential multiple hypothesis testing and efficient fault detection-isolation in stochastic systems. *Information Theory, IEEE Transactions on*, 46(2):595–608, Mar 2000.

[157] L. Lamport. *LaTeX User's Guide and Document Reference Manual*. Addison-Wesley Publishing Company, Reading, Massachusetts, 1986.

[158] W. Lee and D. Xiang. Information-theoretic measures for anomaly detection. In *Security and Privacy, 2001. S&P 2001. Proceedings. 2001 IEEE Symposium on*, pages 130–143. IEEE, 2001.

[159] E. Lehmann and J. Romano. *Testing statistical hypotheses*. Springer Verlag, 2005.

[160] L. Lewis and D. Peterson. Scada honeynet results from the pcsf annual meeting.

[161] T. Lewis. *Critical infrastructure protection in homeland security: defending a networked nation*. LibreDigital, 2006.

[162] Z. Li, A. Das, and J. Zhou. Usaid: Unifying signature-based and anomaly-based intrusion detection. *Advances in Knowledge Discovery and Data Mining*, pages 702–712, 2005.

[163] X. Li-ming, H. Yun-bing, Y. Xu, and H. Guang. Chinese energy consumption structure prediction by application of arima. *China Mining Magazine*, 2011.

[164] H. Liu, S. Shah, and W. Jiang. On-line outlier detection and data cleaning. *Computers and Chemical Engineering*, (28):1635–1647, 2004.

[165] Y. Liu, P. Ning, and M. Reiter. False data injection attacks against state estimation in electric power grids. In *Proceedings of 16th ACM Conference on Computer and Communications Security*, November 2009.

[166] G. Lorden. Procedures for reacting to a change in distribution. *The Annals of Mathmatical Statistics*, 42(6):1897–1908, 1971.

[167] D. Lough. *A taxonomy of computer attacks with applications to wireless networks*. PhD thesis, Virginia Polytechnic Institut, 2001.

[168] J. Lucas and R. Crosier. Fast initial response for cusum quality-control schemes: give your cusum a head start. *Technometrics*, pages 199–205, 1982.

[169] M. Mahoney and P. Chan. An analysis of the 1999 darpa/lincoln laboratory evaluation data for network anomaly detection. In *Recent Advances in Intrusion Detection*, pages 220–237. Springer, 2003.

[170] Y. Makarov, D. Hawkins, E. Leuze, and J. Vidov. California iso wind generation forecasting service design and experience. In *Proc. of the 2002 AWEA Windpower Conference, Portland, Oregon*, 2002.

[171] M. Man and V. Wei. A taxonomy for attacks on mobile agent. In *EUROCON'2001, Trends in Communications, International Conference on.*, volume 2, pages 385–388. IEEE, 2001.

[172] M. Mandjes and P. Zuraniewski. M/g/[infinity] transience, and its applications to overload detection. *Performance Evaluation*, 2011.

[173] R. Martin and C. Masreliez. Robust estimation via stochastic approximation. *IEEE Transaction on Infromation Theory*, IT-21, 1975.

[174] R. D. Martin and D. J. Thomson. Robust-resistant spectrum estimation. In *Proceeding of The IEEE*, volume 70, pages 1097–1115, September 1982.

[175] R. D. Martin and D. J. Thomson. Robust-resistant spectrum estimation. In *Proceeding of The IEEE*, volume 70, pages 1097–1115, September 1982.

[176] J. McHugh. Testing intrusion detection systems: A critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory. *ACM Transactions on Information and System Security*, 3(4):262–294, 2000.

[177] J. McHugh. Intrusion and intrusion detection. *International Journal of Information Security*, 1(1):14–35, 2001.

[178] A. McKinnon, K. Dorow, T. Damania, O. Haugan, W. Lawrence, D. Bakken, and J. Shovic. A configurable middleware framework with multiple quality of service properties for small embedded systems. In *Network Computing and Applications, Second IEEE International Symposium on*, pages 197–204. IEEE, 2003.

[179] R. Mehra. On the identification of variances and adaptive kalman filtering. *Automatic Control, IEEE Transactions on*, 15(2):175 – 184, apr 1970.

[180] Y. Mei. Is average run length to false alarm always an informative criterion?

[181] F. C. Meinecke, S. Harmeling, and K. robert Mller. Robust ica for super-gaussian sources. In *Proc. Int. Workshop on Independent Component Analysis and Blind Signal Separation (ICA2004)*, 2004.

[182] Metasploit. Metasploit blog, August 2010.

[183] A. Metke and R. Ekl. Security technology for smart grid networks. *Smart Grid, IEEE Transactions on*, 1(1):99–107, 2010.

[184] M. Milligan, M. Schwartz, and Y. Wan. Statistical wind power forecasting models: results for us wind farms. *National Renewable Energy Laboratory, Golden, CO*, 2003.

[185] S. Mitter, İ. Schick, M. I. of Technology. Laboratory for Information, and D. Systems. *Point estimation, stochastic approximation, and robust Kalman filtering*. Massachusetts Institute of Technology, Laboratory for Information and Decision Systems, 1993.

[186] M.Mosallae, S. K., and K. Amanian.

[187] I. Modbus. Modbus application protocol specification v1. 1a. *North Grafton, Massachusetts (www. modbus. org/specs. php)*, 2004.

[188] D. C. Montgomery. *Introduction to Statistical Quality Control*. John Wiley & Sons, Inc., Hoboken, N.J, 2009.

[189] B. Moore. Principal component analysis in linear systems: Controllability, observability, and model reduction. *IEEE TRANSACTIONS ON AUTOMATIC CONTROL*, AC-26(1), FEBRUARY 1981.

[190] H. Moore. Fun with vxworks, 2010.

[191] B. Moran and R. Belisle. Modeling flow information and other control system behavior to detect anomalies. In *Proceedings of the SCADA Security Scientific Symposium 2008*, 2008.

[192] G. Moustakides. Optimal stopping times for detecting changes in distributions. *The Annals of Statistics*, pages 1379–1387, 1986.

[193] G. Moustakides and J. Thomas. Optimum detection of a weak signal with minimal knowledge of dependency. *Information Theory, IEEE Transactions on*, 32(1):97 – 102, jan 1986.

[194] P. Muncaster. Stuxnet-like attacks beckon as 50 new scada threats discovered. *http://www.v3.co.uk/v3-uk/news/2045556/stuxnet-attacks-beckon-scada-threats-discovered*, Apr. 2011.

[195] E. Naess, D. Frincke, A. McKinnon, and D. Bakken. Configurable middleware-level intrusion detection for embedded systems. In *Distributed Computing Systems Workshops, 2005. 25th IEEE International Conference on*, pages 144–151. IEEE, 2005.

[196] T. Nas. *Cost-benefit analysis: Theory and application*. Sage Publications, Inc, 1996.

[197] C. Nelson. The prediction performance of the frb-mit-penn model of the us economy. *The American Economic Review*, 62(5):902–917, 1972.

[198] G. Newsham and B. Birt. Building-level occupancy data to improve arima-based electricity use forecasts. In *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*, pages 13–18. ACM, 2010.

[199] T. Nielsen, H. Madsen, H. Nielsen, L. Landberg, and G. Giebel. Zephyr–the prediction models. In *Proceedings of the European Wind Energy Conference*, pages 868–871, 2002.

[200] T. Nielsen, H. Madsen, and J. Tofting. Experiences with statistical methods for wind power prediction. *Proc. EWEC99*, pages 1066–1069, 1999.

[201] I. V. Nikiforov. Optimal sequential detection and isolation of changes in stochastic systems. *IRISA*, 1993.

[202] R. Novoselov, S. Herman, S. Gadaleta, and A. Poore. Mitigating the effects of residual biases with schmidt-kalman filtering. In *The 8th International Conference on Information Fusion*, volume 1, pages 358–365, July 2005.

[203] J. Nyblom. Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84(405):223–230, 1989.

[204] P. Oman and M. Phillips. Intrusion detection and event monitoring in scada networks. *Critical Infrastructure Protection*, pages 161–173, 2007.

[205] E. S. . Page. Continuous inspection schemes. *Biometrika*, 41:100–114, 1954.

[206] T. Paukatong. Scada security: A new concerning issue of an in-house egat-scada. In *Transmission and Distribution Conference and Exhibition: Asia and Pacific, 2005 IEEE/PES*, pages 1–5. IEEE, 2005.

[207] V. Paxson. Bro: a system for detecting network intruders in real-time* 1. *Computer networks*, 31(23-24):2435–2463, 1999.

[208] V. Paxson. Topics in network intrusion detection. In *Tutorial, 8th ACM Conference on Computer and Communications Security (CCS-8)*, November 2001.

[209] R. Pearson. Outliers in process modeling and identification. 10(1):55–63, January 2002.

[210] D. Peterson. Digital bond: Securing the critical infrastructrue, 2008.

[211] C. Pfleeger and S. Pfleeger. *Security in computing*, volume 604. Prentice Hall, 2007.

[212] P. Pingree. The deep impact test benches &# 8211; two spacecraft, twice the fun. In *Aerospace Conference, IEEE*, pages 1–9. IEEE, 2006.

[213] P.J.Huber. Roubst estimation of a location parameter. *Annnals of Mathmatics Statics*, 35, 1974.

[214] P.J.Huber. *Robust Statistics*. Wiley, New York, 1981.

[215] M. Pollak. Optimal detection of a change in distribution. *Annals of Statistics*, (13):206–227, 1985.

[216] M. Pollak and A. G. Tartakovsky. On optimality properties of the shiryaev-roberts procedure. Oct 2007.

[217] H. V. Poor and O. Hadjiliadis.

[218] G. Pottie and W. Kaiser. Wireless integrated network sensors. *Communications of the ACM*, 43(5):51–58, 2000.

[219] E. Price and V. VandeLinde. Robust estimation using the robbins-monro stochastic approximation algorithm. *IEEE Transaction on Infromation Theory*, 25, 1979.

[220] N. Provos and T. Holz. *Virtual honeypots: from botnet tracking to intrusion detection*. Addison-Wesley Professional, 2007.

[221] P.S.Maybeck. Stochastic models, estimation, and control. *Mathematics in Science and Engineering*, 141, 1979.

[222] T. Ptacek. Insertion, evasion, and denial of service: Eluding network intrusion detection. Technical report, DTIC Document, 1998.

[223] P. X. Quang. Robust sequential testing. *The Annals of Statistics*, 13(2):638–649, June 1985.

[224] E. Rakaczky. Intrusion insights adapting intrusion prevention functionality for process control/scada systems, 2006.

[225] S. Ravi, A. Raghunathan, P. Kocher, and S. Hattangady. Security in embedded systems: Design challenges. *ACM Transactions on Embedded Computing Systems (TECS)*, 3(3):461–491, 2004.

[226] Y. Ritov. Decision theoretic optimality of the cusum procedure. (18):1464–1469, 1990.

[227] R.K.Pearson. Exploring process data. *Journal of Process Control*, (11):179–194, 2001.

[228] S. W. Roberts. Control chart tests based on geometric moving averages. *Technometrics*, pages 239–250, 1959.

[229] S. W. Roberts. A comparison of some control chart procedures. *Technometrics*, (8):411–430, 1966.

[230] E. Robinson, B. Woodworth, and R. Pawlowski. Security-hardened attack-resistant platform (sharp), 2008.

[231] A. Rodriguez and M. de los Mozos. Improving network security through traffic log anomaly detection using time series analysis. *Computational Intelligence in Security for Information Systems 2010*, pages 125–133, 2010.

[232] P. J. Rousseeuw and K. V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1998.

[233] J. Rrushi, R. Campbell, and U. di Milano. Detecting attacks in power plant interfacing substations through probabilistic validation of attack-effect bindings. In *SCADA Security Scientific Symposium*, 2008.

[234] B. I. Rubinstein, B. Nelson, L. Huang, A. D. Joseph, S.-H. Lau, N. Taft, and J. D. Tygar. Evading anomaly detection through variance injection attacks on pca. In *RAID '08: Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection*, pages 394–395, Berlin, Heidelberg, 2008. Springer-Verlag.

[235] T. Ryan. *Statistical methods for quality improvement*, volume 840. Wiley, 2011.

[236] S. Salas and E. Hille. *Calculus: One and Several Variable*. John Wiley and Sons, New York, 1978.

[237] M. Sánchez and D. Peña. The identification of multiple outliers in arima models. *Communications in Statistics-Theory and Methods*, 32(6):1265–1287, 2003.

[238] S. Sangsuk-Iam and T. Bullock. Analysis of discrete-time kalman filtering under incorrect noise covariances. *Automatic Control, IEEE Transactions on*, 35(12):1304 –1309, dec 1990.

[239] C. Santos-Pereira and A. Pires. Detection of outliers in multivariate data- a method based on clustering and robust estimators. In *Proceedings in Computational Statistics*, page 291296. Physica-Verlag, 2002.

[240] K. Scarfone and P. Mell. Guide to intrusion detection and prevention systems (idps). *NIST Special Publication*, 800(2007):94, 2007.

[241] I. Schick and S. Mitter. Robust recursive estimation in the presence of heavy-tailed observation noise. *The Annals of Statistics*, pages 1045–1080, 1994.

[242] B. Schneier. *Beyond fear: Thinking sensibly about security in an uncertain world*. Springer Us, 2003.

[243] R. Sekar, A. Gupta, J. Frullo, T. Shanbhag, A. Tiwari, H. Yang, and S. Zhou. Specification-based anomaly detection: a new approach for detecting network intrusions. In *Proceedings of the 9th ACM conference on Computer and communications security*, pages 265–274. ACM, 2002.

[244] M. Severo and J. ao. Gama. Change detection with kalman filter and cusum. *In Discovery Science*, page 243254, 2006.

[245] D. Sexton. Isa sp100.11a overview, 2007.

[246] Z. Shao, Q. Zhuge, Y. He, and E. Sha. Defending embedded systems against buffer overflow via hardware/software. 2003.

[247] W. A. Shewhart. *The Economic Control of Quality of a Manufactured Product*. Van Nostrand, Princeton, 1931.

[248] A. Shiryaev. On optimum methods in quickest detection problems. *Theory of Probability and Its Applications*, (8):2246, 1963.

[249] A. N. Shiryaev. *Optimal Stopping Rules*. Springer, New York, 1978.

[250] A. N. Shiryaev. *Optimal Stopping Rules*. Springer, New York, 2nd edition, 2008.

[251] A. Silberschatz, P. Galvin, and G. Gagne. *Operating System Concepts*. John Wiley & Sons, Inc., 7th edition, 2007.

[252] V. Siris and F. Papagalou. Application of anomaly detection algorithms for detecting syn flooding attacks. *Computer communications*, 29(9):1433–1442, 2006.

[253] R. Snyder. Robust time series analysis. *European Journal of Operational Research*, (9):168–172, 1982.

[254] A. Soule, K. Salamatian, and N. Taft. Combining filtering and statistical methods for anomaly detection. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 31–31. USENIX Association, 2005.

[255] S. H. Steiner. Exponentially weighted moving average control charts with time-varying control limits and fast initial response. *Journal of Quality Technology*, 31:75, 1999.

[256] K. Stouffer, J. Falco, and K. Kent. Guide to supervisory control and data acquisition (scada) and industrial control systems security – recommendations of the national institute of standards and technology. Technical report, September 2006.

[257] S.ZACKS and Z.Kander. Test procedures for possible changes in parameters of statistical distributions occuring at unkown time points. *The Annals of Mathmatical Statistics*, 37:1196–1210, 1966.

[258] T.Cipra and R.Romera. Kalman filter with outliers and missing observations. *TEST*, 6(2), December 1997.

[259] J.-A. Ting, E. Theodorou, and S. Schaal. Learning an outlier-robust kalman filter. Technical Report TR-CLMC-2007-1.

[260] J.-A. Ting, E. Theodorou, and S. Schaal. A kalman filter for robust outlier detection. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 1514 –1519, 29 2007-nov. 2 2007.

[261] J.-A. Ting, E. Theodorou, and S. Schaal. A kalman filter for robust outlier detection. In *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1514–1519, San Diego, CA, Oct 29 - Nov 2 2007.

[262] C. Tsang and S. Kwong. Multi-agent intrusion detection system in industrial network using ant colony clustering approach and unsupervised feature extraction. In *Industrial Technology, 2005. ICIT 2005. IEEE International Conference on*, pages 51–56. IEEE, 2005.

[263] P. Tsang and S. Smith. Yasir: A low-latency, high-integrity security retrofit for legacy scada systems. In *Proceedings of The Ifip Tc 11 23 rd International Information Security Conference*, pages 445–459. Springer, 2008.

[264] R. Tsay. Outliers, level shifts, and variance changes in time series. *Journal of Forecasting*, 7(1):1–20, 1988.

[265] J. Tukey. A survey of sampling from contaminated distributions. *Contributions to Probability and Statistics Essays in Honor of Harold Hotelling*, pages 448–485, 1960.

[266] J. Tukey. *Exploratory Data Analysis*. Addision-Wesley, Reading,MA, 1977.

[267] A. Tversky and D. Kahneman. Loss aversion in riskless choice: A reference-dependent model. *The Quarterly Journal of Economics*, 106(4):1039, 1991.

[268] US-Cert. Vulnerability note vu#190617 livedata iccp server heap buffer overflow vulnerability, 2006.

[269] A. Valdes and S. Cheung. Intrusion monitoring in process control systems. In *Proceedings of the* 42^{nd} *Hawaii International Conference on System Sciences*, Big Island, Hawaii, Jan. 5–8, 2009.

[270] A. Wald. *Sequential Analysis*. J. Wiley & Sons, New York, 1947.

[271] P. C. C. Wang. Robust asymptotic tests of statistical hypotheses involving nuisance parameters. *The Annals of Statistics*, 9(5):1096–1106, September 1981.

[272] Z. Wang, J. Lam, and X. Liu. Robust filtering for discrete-time markovian jump delay systems. *Signal Processing Letters, IEEE*, 11(8):659 – 662, aug. 2004.

[273] K. Wangdi, P. Singhasivanon, T. Silawan, S. Lawpoolsri, N. White, and J. Kaewkungwal. Development of temporal modelling for forecasting and prediction of malaria infections using time-series and arimax analyses: A case study in endemic districts of bhutan. *Malaria Journal*, 9(1):251, 2010.

[274] Y. Wangdi, D. Veal, and S. Maj. Critical infrastructure cyber threat–a case study. *IJCSNS*, 11(6):20, 2011.

[275] K. Whisnant, K. Gross, and N. Lingurovska. Proactive fault monitoring in enterprise servers. *Proc. IEEE Int. Multiconf. Comput. Sci. Comput. Eng*, pages 3–10, 2005.

[276] A. Willis. Design of a modified sequential probability ratio test (sprt) for pipeline leak detection. *Computers & Chemical Engineering*, 35(1):127–131, 2011.

[277] A. Willsky and H. Jones. A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *Automatic Control, IEEE Transactions on*, 21(1):108–112, 1976.

[278] A. S. Willsky. A survey of design methods for failure detection in dynamic systems. *NASA STI/Recon Technical Report N*, 76:11347–+, Nov. 1975.

[279] A. S. Willsky. Detection of abrupt changes in dynamic systems. In *Detection of Abrupt Changes in Signals and Dynamical Systems, number 77 in Lecture Notes in Control and Information Sciences*, pages 27–49. Springer-Verlag, 1986.

[280] A. S. . Willsky and H. L. Jones. A generalized likelihood ratio approach to state estimation in linear systems subject to abrupt changes, NOV 1974.

[281] R. Wolski. Dynamically forecasting network performance using the network weather service. *Cluster Computing*, 1(1):119–132, 1998.

[282] K. Xiao, N. Chen, S. Ren, L. Shen, X. Sun, K. Kwiat, and M. Macalik. A workflow-based non-intrusive approach for enhancing the survivability of critical infrastructures in cyber environment. In *Proceedings of Third International Workshop on Software Engineering for Secure Systems (SESS'07)*. IEEE Computer Society, 2007.

[283] L. Xu, K. Tomsovic, and A. Bose. Topology error identification using a two-stage dc state estimator. *Electric Power Systems Research*, 74:167–175, April 2005.

[284] A. Yaacob, I. Tan, S. Chien, and H. Tan. Arima based network anomaly detection. In *2010 Second International Conference on Communication Software and Networks*, pages 205–209. IEEE, 2010.

[285] B. Yakir. Optimal detection of a change in distribution when the observations form a markov chain with a finite state space. *Change-point problems: Papers from the AMS-IMS-SIAM Summer Research Conference held at Mt. Holyoke College 1992*, 1994), PAGES=346-358.

[286] S. A. Yamamoto, D. Asahara, T. Itao, and T. S. Tanaka. Distributed pagerank: A distributed reputation model for open peer-to-peer networks. In *SAINTW04: Proceedings of the 2the 2004 International Symposium on Applications and the Internet Workshops*, pages 66–77, New York, NY, USA, 2004. ACM.

[287] D. Yang, A. Usynin, and J. Hines. Anomaly-based intrusion detection for scada systems. In *5th Intl. Topical Meeting on Nuclear Plant Instrumentation, Control and Human Machine Interface Technologies (NPIC&HMIT 05)*, pages 12–16. Citeseer, 2005.

[288] M. Yang, X. Li, H. Chen, and N. Rao. Predicting internet end-to-end delay: an overview. In *System Theory, 2004. Proceedings of the Thirty-Sixth Southeastern Symposium on*, pages 210–214. IEEE, 2004.

[289] J. Zachary, J. McEachen, and D. Ettlich. Conversation exchange dynamics for real-time network monitoring and anomaly detection. 2004.

[290] S. Zanero. Behavioral intrusion detection. *Computer and Information Sciences-ISCIS 2004*, pages 657–666, 2004.

[291] Y. Zhang, Z. Ge, A. Greenberg, and M. Roughan. Network anomography. In *Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement*, pages 30–30. USENIX Association, 2005.

[292] T. Zheng, A. A. Girgis, and E. B. Makram. A hybrid wavelet-kalman filter method for load forecasting. 54(1):11–17, April 2000.

[293] M. Zhong, Q. Ding, and P. Shi. Parity space-based fault detection for markovian jump systems. *Intern. J. Syst. Sci.*, 40:421–428, April 2009.

[294] F. Zhou, T. Tang, and C. Wen. A new multi-scale estimation scheme for dynamic system. In *Proceedings of the 26th Chinese Control Conference*, volume 26, pages 396–399. IEEE, June 2007.

[295] B. Zhu. Tradeoffs in estimation over wireless sensor network. *Master Thesis*, May 2007.

[296] B. Zhu, A. Joseph, and S. Sastry. Taxonomy of cyber attacks on scada systems. In *Proceedings of the 2011 IEEE International Conference on Cyber, Physical, and Social Computing (CPSCom 2011)*. IEEE Computer Society.

[297] B. Zhu and S. Sastry. 'beaver' the architecture of an intrusion tolerant scada system. *presented at the IEEE S & P Oakland WIP session, work-in-progress*, May 2008.

[298] B. Zhu and S. Sastry. Data fusion assurance for the kalman filter in uncertain networks. In *Proceedings of the 4th International Conference on Information Assurance and Security*, Washington DC,USA, September 2008. IEEE Computer Society.

[299] B. Zhu and S. Sastry. Scada-specific intrusion and prevention systems: A survey and taxonomy. Technical report, May 2008.

[300] B. Zhu and S. Sastry. The cyber-physical security implication and countermeasure of scada protocols in power grids. *work-in-progress*, November 2009.

[301] B. Zhu and S. Sastry. Look into the "noise" early detection of abnormal signals in cyber-physical systems. *work-in-progress*, November 2009.

[302] B. Zhu and S. Sastry. Scada-specific intrusion detection/prevention systems: A survey and taxonomy. In *Proceedings of the First Workshop on Secure Control Systems (SCS'10)*, Stockholm, Sweden, 2010.

[303] B. Zhu and S. Sastry. Anomaly detection for clean energy resources prediction and power consumption forecast in the smart grid. *to appear in ICMLA 2011*, 2011.

[304] B. Zhu and S. Sastry. Jie: A viable intrusion detection system for scada systems, 2011.

[305] B. Zhu and S. Sastry. Resilient control and early detection for critical infrastructures, 2011.

[306] B. Zhu and S. Sastry. Revisit dynamic arima based anomaly detection. MIT,Boston, MA, 2011.

[307] B. Zhu and S. Sastry. Robust discovering and tracking in challenging environments. In *Proceedings of the 2011 International Symposium on Safety, Security, and Rescue Robotics (SSRR 2011)*, Kyoto, Japan, 2011.

[308] B. Zhu and S. Sastry. Robust estimation and intrusion detection for scada systems. *Technical Report EECS, UC Berkeley*, 2011.

[309] B. Zhu, B. Sinopoli, K. Poola, and S. S. Sastry. Estimation in wireless sensor network. In *Proceedings of the 26th American Control Conference*, July 2007.

[310] X. Zhu, Y. Soh, and L. Xie. Robust kalman filter design for discrete time-delay systems. *Circuits, systems, and signal processing*, 21(3):319–335, 2002.

[311] X. Zhu, Y. C. Soh, and L. Xie. Robust kalman filter design. In *Decision and Control, 2000. Proceedings of the 39th IEEE Conference on*, volume 4, pages 3813 –3818 vol.4, 2000.