CYBER SECURITY OF SUBSTATION AUTOMATION SYSTEMS

By

JUNHO HONG

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
School of Electrical Engineering and Computer Science

AUGUST 2014

To the Faculty of Washington State University:

The members of the Committee appointed to examine the dissertation of JUNHO HONG find it satisfactory and recommend that it be accepted.

Chen-Ching Liu, Ph.D., Chair

Carl Hauser, Ph.D.

Anurag K. Srivastava, Ph.D.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my academic and research advisor Prof. Chen-Ching Liu for providing me the opportunity to work with him. He helped me and provided me with constant support during my entire Ph. D. study. His profound knowledge, dedication to research, tirelessness and determination always inspire me. I would also like to express my appreciation to the members of my committee, Prof. Carl Hauser, and Prof. Anurag K. Srivastava, for all their support and feedback during my research. Their lectures on power systems and smart grid cyber security laid a solid foundation for my research and knowledge. I also owe special appreciation to Prof. Manimaran Govindarasu at Iowa State University, Prof. Ying Chen at Tsinghua University, Prof. Pavel Gladyshev at University College Dublin, and Prof. Chee-Wooi Ten at Michigan Technological University for their help and guidance during the research.

I want to thank all the our power group members, Dr. Guanqun (Steven) Wang, Yazhou (Leo) Jiang, Zijie Lin, Chih-Che (Ryan) Sun, Dr. Yin Xu, Haosen Guo, and Energy Systems Innovation (ESI) Center staff members, Jody Opheim and Heather Flodin. Special thanks go to my office mates, Tianying (Lily) Wu, Lin Zhang, Hyojong Lee, Jeong-Hun Kim and my friends at UCD, Talieh ZarabZadeh, Shinn-Shyan (Jasper) Wu, Alexandru Stefanov, Line He, Shane Ryan, and Jing Xie.

I owe my sincere appreciation to my parents, my wife Woorim Seon, my baby Terry Hong, and relatives who have been a constant source of support and encouragement. This work was partially supported by the National Science Foundation (NSF) and Science Foundation Ireland (SFI). I would like to acknowledge the financial support extended to this project by the sponsors.

CYBER SECURITY OF SUBSTATION AUTOMATION SYSTEMS

Abstract

by Junho Hong, Ph.D.
Washington State University
August 2014

Chair: Chen-Ching Liu

Cyber intrusions into substations of a power grid are a source of vulnerability since most substations are unmanned and with limited protection of the cyber and physical security. In the worst case, simultaneous cyber intrusions into multiple substations can lead to severe cascading events, causing catastrophic power outages. In addition, substation communication protocols do not include cyber security features in their original standard. Generic Object Oriented Substation Event (GOOSE) contains the circuit breaker trip command whereas Sampled Measured Value (SMV) includes measured analog values such as currents and voltages. Due to the importance of substation automation multicast messages, IEC 62351 standards proposed the authentication method as a primary security measure for GOOSE and SMV messages since they required fast transmission time. However, performance testing for the application of the authentication method to GOOSE and SMV is in an early stage, and there is presently no solution to detection of the GOOSE and SMV related error, anomaly and intrusion. Cyber security technologies for anomaly detection at a substation are in an early stage of development. Technologies to detect anomalies for substation automation multicast protocols and applications are critically needed. This dissertation is concerned with anomaly detection in the computer network environment of a substation. The proposed integrated Anomaly Detection System (ADS) contains host- and network-based anomaly detection systems for the substations, and simultaneous anomaly detection for multiple substations. Potential

scenarios of simultaneous intrusions into the substations have been simulated using a substation automation testbed based on the IEEE 39 and modified IEEE 118-bus systems. The host-based anomaly detection considers temporal anomalies in the substation facilities. The malicious behaviors of substation automation based on multicast messages are incorporated in the proposed network-based anomaly detection. The proposed impact evaluation method can help operators find the most critical substation among the anomaly detected substations. In addition, the proposed simultaneous intrusion detection method is able to identify the same type of attacks at multiple substations and their locations. The result is a new integrated tool for detection and mitigation of cyber intrusions at a single substation or multiple substations of a power grid.

Keywords: Cyber security of Substations, Anomaly Detection, Network Security, GOOSE Anomaly Detection, SMV Anomaly Detection and Intrusion Detection

PUBLICATIONS

Journal and Magazines

[1] **J. Hong**, C.-C. Liu, and M. Govindarasu, "Integrated anomaly detection for cyber security of the substations," Accepted for publication in *IEEE Trans. Smart Grid*, 2014.

[2] C.-C. Liu, A. Stefanov, **J. Hong**, and P. Panciatici, "Intruders in the grid," *IEEE Power and Energy Magazine*, vol. 10, no. 1, pp. 58-66, Jan.-Feb. 2012.

[3] C.-W. Ten, **J. Hong**, and C.-C. Liu, "Anomaly detection for cybersecurity of the substations," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 865-873, Dec. 2011.

Conference Proceedings

[1] **J. Hong**, C.-C. Liu, and M. Govindarasu, "Detection of cyber intrusions using network-based multicast messages for substation automation," *Proc. IEEE PES Innov. SmartGrid Technol. (ISGT)*, Feb. 2014.

[2] **J. Hong** and C.-C. Liu, "Cyber-physical security of a power grid based on events at substations," *The 12th International Workshop on Electric Power Control Centers (EPCC)*, June 2013. [Online]. Available: http://www.epcc-workshop.net/assets/downloads/hong-cyber-physical-security.pdf

[3] **J. Hong**, A. Stefanov, C.-C. Liu, and M. Govindarasu, "Cyber-physical security in a substation," *IEEE Power and Energy Society General Meeting*, July 2012.

[4] **J. Hong**, S.-S. Wu, A. Stefanov, A. Fshosha, C.-C. Liu, P. Gladyshev, and M. Govindarasu, "An intrusion and defense testbed in a cyber-power system environment," *IEEE Power and Energy Society General Meeting*, July 2011.

Book Chapters

[1] **J. Hong**, Y. Chen, C.-C. Liu, and M. Govindarasu, "Cyber-physical security of substations in a power grid," in *Cyber Physical Systems Approach to Smart Electric Power Grid*, S.-K. Khaitan (Editor), Springer-Verlag Inc, 2014, under review.

[2] **J. Hong** and C.-C. Liu, "Distribution automation in a smart grid environment," in *Advanced Solutions in Power Systems: HVDC, FACTS, and Artificial Intelligence Techniques*, M. Eremia (Editor), IEEE, 2013, under review.

Technical Reports

[1] C.-C. Liu, **J. Hong**, Y. Jiang, and S. Liu, Optimal Blackstart Capability (OBC) for Power System Restoration. Annual Report, EPRI, Palo Alto, CA. 2013.

[2] C.-C. Liu, Y. Jiang, **J. Hong**, and S. Liu, Optimal Blackstart Capability (OBC) User's Guide. EPRI, Palo Alto, CA. 2013.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1. Introduction

1.1 Motivation

Power grids are complex cyber and physical systems. The physical system of power grids includes power plants, substations, and transmission and distribution systems. Electric power is produced by generators, while substations convert Alternating Current (AC) voltage from a voltage level to another for delivery from power plants to the load. Transmission systems deliver electric power to distribution substations through transmission networks. Distribution systems transport electric energy to customers. The physical system of power grids relies on the cyber system for monitoring, control, and operation. The cyber system of power grids is formed by the Information and Communications Technology (ICT) at the substations and the Supervisory Control And Data Acquisition (SCADA) system at the control center. Therefore, a power grid is a critical infrastructure that relies on ICT and SCADA systems for monitoring, control and operation. On top of the power infrastructure resides layers of information and communications technology (ICT) that are interconnected with electric grids. The cyber and power infrastructures together constitute a large and complex cyber-physical system. The SCADA system acquires analog and status data needed for dispatchers in a control center to perform economic and power system security functions with support from an Energy Management System (EMS). At substations, advanced IT systems have been installed with communication layers based on industry standards. As the electricity industry evolves into a market environment, more and more information is exchanged between EMS and other entities, e.g., electricity markets and other interconnected grids.

A blackout of a power grid has a significant impact on the society and economy. These catastrophic outages can be caused by human errors, equipment failures and natural disasters [1]. Research has been conducted on the mitigation of these outages, e.g., methods to identify and isolate the faulted area(s) and restore unaffected areas by self-healing technologies [2]. However, power outages and blackouts can also be induced by cyber attacks. As a result, cyber security of the ICT for power grids has become a critical issue. With the increasing deployment of information and communications technology (ICT), power grids need to incorporate the cyber intrusion as a major threat since well organized cyber attacks at multiple substations may trigger a sequence of cascading events, leading to a blackout [3, 4]. It is important to model the cyber-power system as one integrated complex structure. For instance, what are the consequences and impact of a cyber attack on the information and communications technology on a power systems? Along with targeted attacks, such as sniffing or malicious alterations of data packets, cyber attacks based on denial of service (DoS) mechanisms and the use of viruses and worms can cause serious disruption of services. A DoS attack prevents legitimate users of the facilities from performing regular or emergency services. An aggressive attack is the combination of denial of control and denial of view, where the controller is no longer in control and can not recognize the loss of capability. This type of attack destroys the capability of control systems or operators to operate the system by reducing observability and/or controllability of the cyber-physical system. The following three examples show the reported cyber intrusions and demonstration that are aimed at critical infrastructures.

(1) The widely publicized cyber attack on industrial control systems is the Stuxnet worm, a malware targeting SCADA systems. According to Symantec infection statistics (September 29, 2010), Stuxnet has infected approximately 50,000 to 100,000 computers in a number of countries. The objective is to reprogram industrial control systems by modifying code on Programmable Logic

Controllers (PLCs) and turn them into the attacker's agents. Stuxnet searches for a specific type of PLC and waits for a certain condition before it takes control. Although the target has no connection to Internet, it is highly vulnerable as the infection is initiated by a simple flash memory. Following successful infections, Stuxnet updates itself using peer-to-peer communications among infected computers. Media suggested that Stuxnet's targets were nuclear plant. However, with modifications, it can become a serious threat to power grids.

(2) A demonstration of a targeted cyber attack was provided by the US Department of Energy's Idaho National Laboratory, in March 2007, for a project named "Aurora." A previously classified video was produced and released to the press in September 2007, to demonstrate the vulnerabilities of the electric power grids. The attack was launched remotely on the control system of an electric generator. The cyber attack induces mechanical effects that drive the generator out of control, the rotor hits the stator and the windings are shredded. The project demonstrates how a cyber attack is translated into damages on physical devices. Coordinated simultaneous attacks on multiple power plants with the objective of damaging a large number of generators are serious threats to national security.

(3) In February 2011, McAfee published a white paper on "Global Energy Cyber attacks: Night Dragon," stating that coordinated and targeted cyber attacks have been conducted against global oil, energy, and petrochemical companies by the use of remote administration tools (RAT) and special network techniques. Remote administration tools are used by administrators or hackers to manage systems or the victims' computers. The attacks were launched from several countries to obtain proprietary and confidential information. First, the extranet web servers were compromised, then access was gained to internal servers and desktops, usernames and passwords were acquired, and

direct communications from infected machines to the Internet was enabled. As a result, security was breached and private documents were accessed.

In order to mitigate cyber attacks, a firewall is widely adopted as an access control method against hackers. However, firewalls do not guarantee cyber security. It has been reported that companies' firewalls have been mis-configured and, even if the configuration of firewalls is correct, it has vulnerabilities because firewall is not able to detect insider attacks and connection from the trusted side. Hence, solutions based solely on firewalls can be inadequate.

Protection relays in the substations are critical devices for system protection. Conventional relays have only local access using a serial cable connection. As ICTs evolve, remote access is enabled for Ethernet based networks, allowing site engineers, operators and vendor personnel to access remotely. Remote access to Intelligent Electronic Devices (IEDs) from within a substation, corporate office, or locations external to the grid, is a common practice for control and maintenance purposes. Dial-up, Virtual Private Network (VPN), and wireless are available mechanisms between remote access points and the substation Local Area Network. These access points are potential cyber vulnerabilities of the substations. When remote access points are compromised by intruders, malicious attacks to operate circuit breakers and/or to access critical information, such as Substation Configuration Description (SCD), can be launched. Furthermore, IEDs may have a web server to allow a remote configuration change and control.

International standard protocols have been developed for power system data communication by International Electrotechnical Commission (IEC) Technical Committee (TC) 57. These protocols, e.g., DNP3.0, IEC 60870-5, IEC 60870-6 and IEC 61850, are widely used for power equipment,

EMS, SCADA, and distribution automation. However, these standard protocols have vulnerabilities and open standards can be easy to access. Intruders can analyze protocols and most of these protocols are not equipped with cyber security methods since cyber security has emerged in recent years as a serious concern. Therefore, IEC technical committee (TC) 57 published the cyber security standards, IEC 62351, for power systems management and associated information exchange [5]. IEC 62351 for information security of power system control operations now has within its scope the above mentioned protocols the protective measures of packet encryption, authentication, and network & system management methods. Nevertheless, this standard is not able to cover all cyber intrusions, e.g., compromising firewalls and intrusion attempts to substation user interface or Intelligent Electronic Devices (IEDs).

Substation automation based on IEC 61850 is a key element to achieve interoperability in a smart grid [6]. The concept of IEC 61850 is adopted in distribution automation and the deployment of distributed energy resources (DERs). Cyber-physical security of substations is a critical issue for the smart grid as substations play an important role in monitoring and control of the power grids. However, as explained above, the substation automation standard, IEC 61850, does not include cyber and information security features for substations. IEC 62351 standards proposed the authentication method as a primary security measure for GOOSE and sampled value messages since they required fast transmission time (less than 4 ms). However, performance testing for the application of the authentication method to GOOSE and SV is in an early stage. Cyber intrusions related to these protocols may cause serious damages to a power grid. Intruder(s) may modify GOOSE control messages and operate circuit breakers in a substation. They can also send fabricated (and improper) protection coordination messages to other substations. A SV message attack can generate fabricated analog values to a control center, leading to undesirable operations.

1.2 Literature Survey

One way to address above mentioned issues is to develop new technologies to detect and disrupt malicious activities across the networks. An anomaly detection system is an early warning mechanism to extract relevant cyber security events from substations and correlate these events. In the literature, methods for event correlations, such as alarm processing, fault diagnosis, and security assessment for power systems have been proposed [7, 8, 9]. The work of [10] explains the concept of cyber-physical security in four steps: (1) modeling of the cyber-net, (2) simulation of the physical behaviors of a power grid, (3) development of a vulnerability index for the cyber-physical system, and (4) determination of mitigation measures. In order to mitigate the cyber attacks related to substation automation, an intrusion detection system for IEC 61850 based substation automation system was proposed [11]. The work of [12] proposed a retrofit data logger solution and an intrusion detection system for serial communication based MODBUS and DNP3 in the substations. Temporal anomaly detection in a substation has been developed in work of [13]. The vulnerabilities of critical infrastructures have been reported by National Institute of Standards and Technology (NIST) and discussed at the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) Workshop on CIP 002-009 [14]. NIST also identified key attributes of the logical design for intrusion-based attacks on power equipment that is critical to standardization and modeling [15], [16]. However, none of them proposed the cyber security measures to detect cyber threats for substation multicast protocols such as GOOSE and SV. Therefore, technologies to detect anomalies and intrusions for multicast messages of substation automation protocols are critically needed.

Several testbeds for cyber-physical security of power systems have been developed by a number of institutions. Idaho National Laboratory (INL) developed a National SCADA Testbed (NSTB) that can be used to identify and mitigate existing vulnerabilities [17, 18, 19]. The Virtual Control System Environment (VCSE) is developed by Sandia National Laboratory (SNL) that can be used to model and simulate cyber-physical system security [20], [21]. Iowa State University stablished the PowerCyber testbed using Real Time Digital Simulators (RTDS), and ISEAGE WAN emulation [22]. The Virtual Power System Testbed (VPST) is developed by the University of Illinois with the PowerWorld power system simulator and a Real-Time Immersive Network Simulation Environment (RINSE) [23]. The work of [24] proposes anomaly-based intrusion detection on the SCADA Control Systems (TASSCS) at the University of Arizona. The CRUTIAL testbeds are proposed to analyze the ICT resilience of power control systems in Europe [25], [26]. The testbed at the University College Dublin (UCD) has the capability to simulate cyber attacks and its impact on the power grids. This testbed is based on the commercial EMS and DIgSILENT power system simulator [27]. Royal Melbourne Institute of Technology (RMIT) developed the SCADASim testbed for testing of different attack and security solutions on actual devices and applications using a simulated environment [28].

1.3 Objectives and Contributions

This dissertation is concerned with anomaly detection at a substation. An integrated method for host-based and network-based anomaly detection schemes is proposed. The host-based anomaly detection uses a systematic extraction technique for intrusion footprints that can be used to identify credible intrusion events within a substation, e.g., firewall, user-interface, IEDs, and circuit breakers. The network-based anomaly detection is focused on multicast messages in a substation network; it also detects, in a real-time environment, anomalies that demonstrate abnormal behaviors. The main

contribution of this dissertation is a new method for (1) an integrated anomaly detection system for protection of IEC 61850 based substation automation system, e.g., IEDs, user-interface and firewall, (2) a network-based anomaly detection algorithm that can be used to detect malicious activities of IEC 61850 based multicast protocols, e.g., GOOSE and SMV, across the substation network, (3) an impact evaluation method is proposed based on the detected anomalies, and (4) simultaneous anomaly detection among multiple substations using anomaly detection system data. Anomaly detection for multicast messages in a substation automation network is a new field of research for the power grids. In this research, a cyber security testbed has been developed and used to validate the proposed anomaly detection algorithms. Cyber intrusions are simulated using the testbed including protective IEDs. The test results demonstrate that proposed anomaly detection algorithms are effective for the detection of simulated attacks.

1.4 Organization of This Dissertation

This dissertation includes six chapters. Chapter I introduces the motivations, literature survey, objectives and contributions of this dissertation. Chapter II describes a substation automation system that includes IEC 61850 standard, multicast message, vulnerabilities, and intrusion scenarios of the substations. Single substation attacks and simultaneous attacks to multiple substations will be explained using the testbed and attack tree. Chapter III illustrates the proposed temporal event based anomaly detection algorithm, RAIM framework, impact analysis, and detection of simultaneous attacks. The proposed anomaly detection algorithm in this chapter uses the system and security logs that are generated from user-interface, IEDs, firewalls and circuit breakers. Therefore, the anomaly detection algorithm will rely on data logs at the substation level networks. RAIM is the main framework of this chapter; it stands for Real-time monitoring, Anomaly detection, Impact analysis, and Mitigation strategies. The impact factor shows how close a system is to a collapse and identifies

the most critical substation among the substations where anomalies are detected. The proposed methodology for evaluation of the impact of cyber intrusions at a substation level is validated using the modified IEEE 118-bus system model. The integrated anomaly detection that contains host and network-based detection algorithm is described in Chapter IV. The proposed host-based anomaly detection uses a systematic extraction technique for intrusion footprints that can be used to identify credible intrusion events within a substation. The network-based anomaly detection is focused on multicast messages in a substation network that can be used to detect anomalies or abnormal behaviors in a real-time. This chapter also proposes an attack similarity method which can be used to calculate a similarity coefficient among the substations where anomalies are detected. The conclusions and recommendations for the future work are given in Chapter V.

Chapter 2. Substation Automation System

The concept and design of substation automation system was proposed by the International Electrotechnical Commission (IEC) Technical Committee (TC) 57, Working Group (WG) 10. IEC TC 57 published IEC 61850 which is a standard for the design of substation automation system. The main purposes of IEC 61850 standard can be divided into four parts, (1) Lower configuration and installation cost, (2) Multi-vendor interoperability, (3) Long term stability, and (4) Minimal impact to the existing system. The installation and engineering cost of IEC 61850 based devices are drastically reduced since all hardwired connections from CTs and VTs to relays are changed to Ethernet based communications using Sampled Measured Value (SMV) messages which contain sampled data of currents and voltages. The Generic Object Oriented Substation Event (GOOSE) enables IEC 61850 based devices to quickly exchange critical data (e.g., a trip signal to a circuit breaker)., i.e., less than 4 [msec], over the Ethernet based communication. This also significantly reduces the cost of wire installation. The Substation Configuration Language (SCL) contains device configuration information. Therefore, IEC 61850 based devices do not need any manual configurations, they import the configured SCL file through the ICT network. Standardized communication protocols and logical nodes enhance multi-vendor interoperability. Therefore, substation operators can use IEDs and user-interfaces from different vendors in a substation. The concept of IEC 61850 is extended to distributed energy resources (DERs) and distribution automation. Hence, IEC 61850 enables devices from different manufacturers to exchange information in the substation level as well as system level [29]. The ICT technologies have been fast evolving over the last decade and the trend is continuing. However, the evolving cycle of power substation functions and software applications are slow compared to that of ICTs. The long term stability allows upgrading of ICT at a substation without re-engineering of the entire substation system. Since multi-vendor interoperability significantly reduced the gaps of device configuration

between different vendors, substation engineers can add or remove existing devices at a lower cost. For instance, substation engineers can set up new devices and applications in a substation by sending SCL files via the ICT network [30].

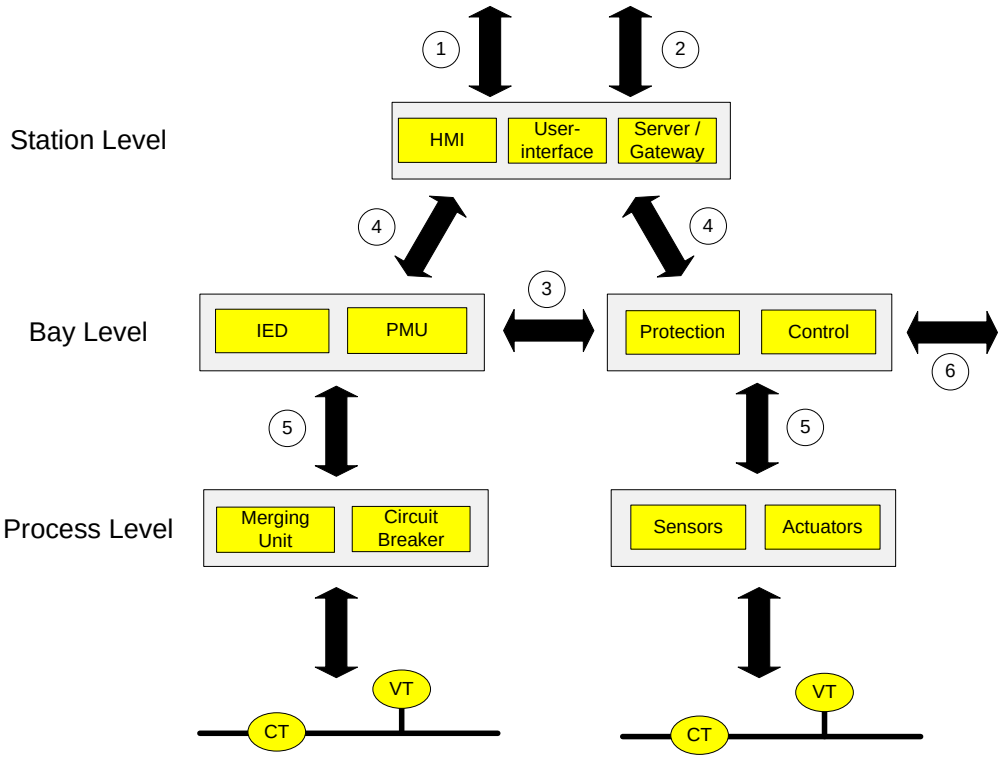

Fig. 2.1 Communication topology of the substation automation system (cyber system)

Fig. 2.1 shows the three levels of the substation automation system, i.e., the station, bay, and process levels. The station level is where the user-interface, Human Machine Interface (HMI), substation server and gateway are located. The server and gateway exchange data coming from/to substation, e.g., remote access points (interface 1), control centers (interface 2) using Distributed Network Protocol (DNP) 3.0 or IEC 60870-5 [31]. The protective devices exchange critical data, e.g., interlocking (interface 3), between bays using GOOSE messages. Control and protection data

are exchanged between the station and bay level using Manufacturing Message Specification (MMS) message (interface 4). Measurements such as currents and voltages are sent to the station level from the process level to bay level whereas control data are sent from the bay level to process level (interface 5) using SMV and GOOSE, respectively. Interface 6 shows the remote control and protection features between substations [32].

A substation includes various types of critical physical equipment, e.g., transformers, circuit breakers (52), bus bars, disconnect switches, and feeders, as shown in Fig. 2.2. The substation in Fig. 2.2 has two main transformers, and single busbars. When a fault occurs at a transformer or a busbar, the faulted area can be isolated by switching actions. The substation equipment will be protected by different types of protective relays. For instance, the transformer and busbar are protected by differential relays while the feeder is protected by overcurrent relays.

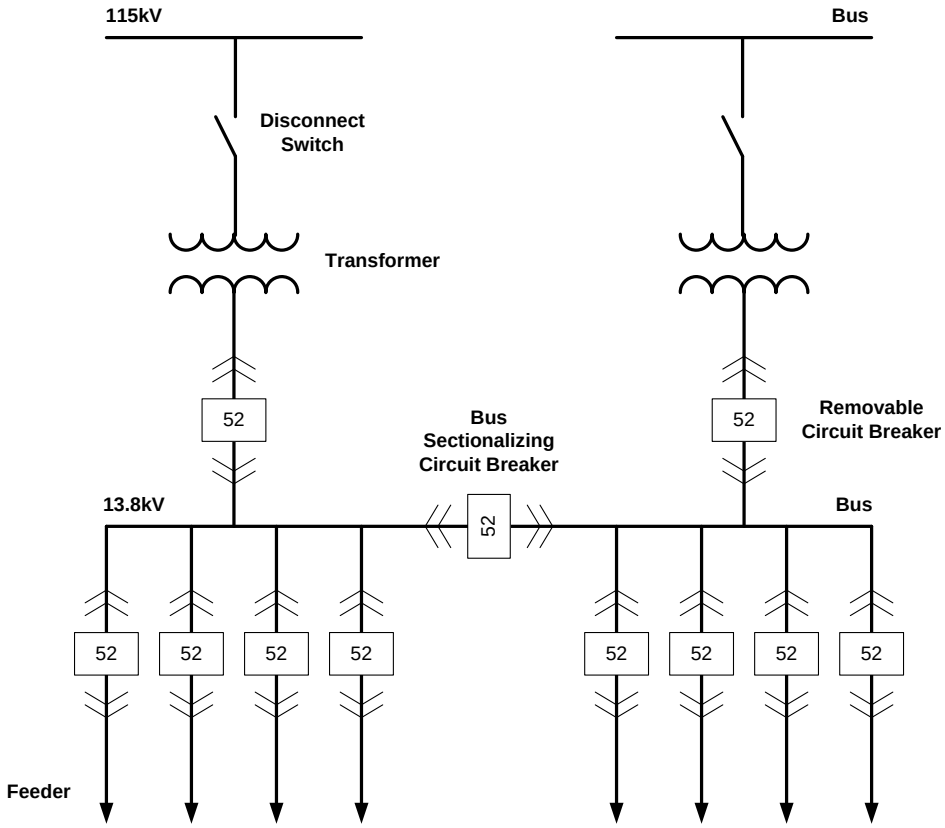

Fig. 2.2 The one line diagram of a substation (physical system) [33]

2.1 IEC 61850 Standard

The IEC 61850 is divided into 10 sections and 7 sub-sections as shown in Table 2.1. Part 1 is an overview of the IEC 61850 standard series, basic interface and reference model of a substation automation system. Part 2 provides an explanation of the abbreviations and terms that are used in IEC 61850 series. Part 3 describes the general requirements of the ICT networks and guidelines for environmental conditions and recommendations. Part 4 is concerned with the system and project management with respect to the engineering process, life cycle of the overall system and supporting tools for engineering and testing. The scope of part 5 covers the communication requirements of the functions that are performed in the substation automation system. It also explains the Logical Nodes

(LNs) for each function, e.g., PTOC is an AC time overcurrent relay that is able to trip the circuit breaker when the input current exceeds the predetermined threshold. The IED related configuration languages are shown in part 6, e.g., SCL, IED Capability Description (ICD), System Exchange Description (SED), Instantiated IED Description (IID), System Specification Description (SSD) and Configured IED Description (CID) that are based on the Extensible Markup Language (XML). Part 7 deals with the basic communication structure for substation and feeder equipment. Part 7-1 explains the principles of the modeling method, communication and information models that are used in IEC 61850-7-x. The definition and structure of Abstract Communication Service Interface (ACSI) communication in substations are introduced in part 7-2. Part 7-3 provides details of the layered substation communication architecture. The ICT models of functions and devices that are related to substation automation are described in part 7-4. Specially, this part of the standard includes details of logical node names and data names for communication between substation devices, e.g., IEDs and user-interfaces. Part 8-1 describes a method for data exchange between ACSI and MMS communication. Finally, part 9-1 and part 9-2 explain the structure and mapping of the SMV. Part 10 covers the subject of conformance testing for IEC 61850 systems.

Table 2.1: Sections of IEC 61850 standards

Section	Title
IEC 61850-1	Introduction and overview
IEC 61850-2	Glossary
IEC 61850-3	General requirements
IEC 61850-4	System and project management
IEC 61850-5	Communication requirements for functions and device models
IEC 61850-6	Configuration language for communication in electrical substations related to IEDs
IEC 61850-7	Basic communication structure for substation and feeder equipment
├ IEC 61850-7-1	├ Principles and models
├ IEC 61850-7-2	├ Abstract communication service interface (ACSI)
├ IEC 61850-7-3	├ Common Data Classes
└ IEC 61850-7-4	└ Compatible logical node classes and data classes
IEC 61850-8	Specific communication service mapping (SCSM)
└ IEC 61850-8-1	└ Mappings to MMS (ISO/IEC9506-1 and ISO/IEC 9506-2)
IEC 61850-9	Specific communication service mapping (SCSM)
├ IEC 61850-9-1	├ Sampled values over serial unidirectional multidrop point to point link
└ IEC 61850-9-2	└ Sampled values over ISO/IEC 8802-3
IEC 61850-10	Conformance testing

2.2 Multicast Message in a Substation Automation System

The communication protocols in IEC 61850 can be classified into seven types. Due to the requirement of type 1, 1A and 4 messages, e.g., GOOSE and SV, they use three communication stacks, i.e., physical, data link and application layer as shown in Fig. 2.3. GOOSE supports critical data exchange such as interlocking between IEDs, trip messages from IED to circuit breakers or the status of circuit breakers to IED. The basic concept of information exchange is that a publisher writes values in a GOOSE packet and subscriber receives and reads the values from the GOOSE

packet. GOOSE uses Media Access Control (MAC) address for the multicast[1] scheme. Due to the real-time requirement, GOOSE applies a re-transmission[2] scheme in order to achieve the appropriate level of communication speed and reliability. As shown in Fig. 2.1, the merging unit receives voltage and current values from CT and VT through the hard wire. Then the merging unit sends measured current and voltage values to protection IEDs using SMV messages. A merging unit can send SMV messages to multiple IEDs since SMV supports the multicast scheme. There are three types of resolution (bits) amplitude for SMV messages such as bits (P1 class), 16 bits (P2 class) and 32 bits (P3 class) [34].

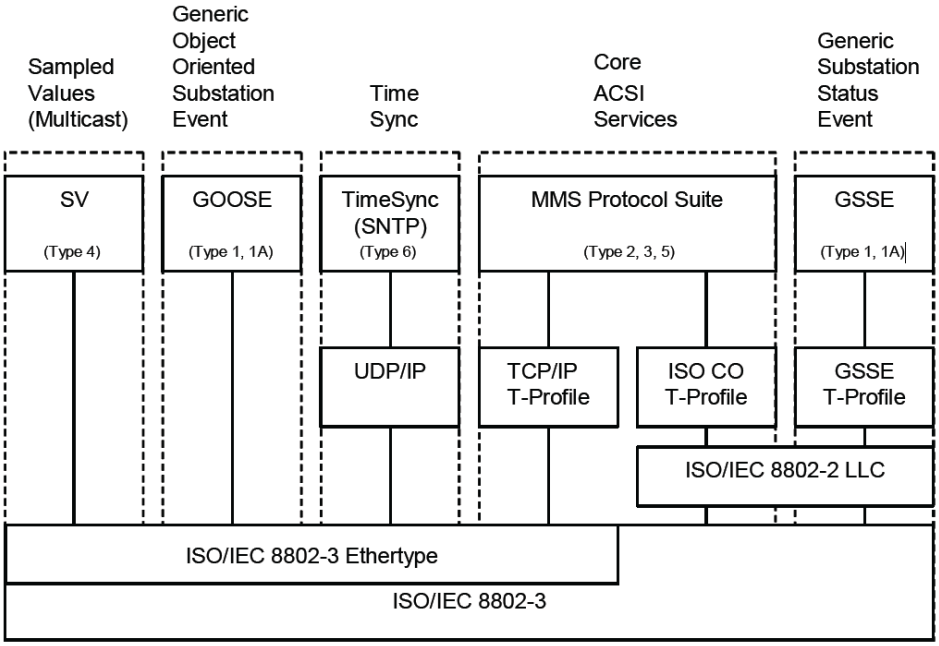

Fig. 2.3 Communication protocols in IEC 61850 [35]

[1] Multicast is the delivery of data or information in a single host to multiple receivers simultaneously.

[2] The receiver does not send any response to the sender.

- Type 1: Fast messages
- Type 1A: Trip
- Type 2: Medium speed messages
- Type 3: Low speed messages
- Type 4: Raw data messages
- Type 5: File transfer functions
- Type 6: Time synchronization messages

2.3 Vulnerabilities and Intrusion Scenarios of the Substations

The cyber security of substations has been recognized as a critical issue since it consists of various types of critical physical and cyber devices as explained in previous Section. They can be physically or electrically connected, e.g., a protection and control unit of a transformer is connected to user-interface via the substation local area network. The remote access to substation networks, e.g., IED or user-interface, is a common way for maintenance of the substation facilities. However, there are many potential cyber security issues, such as: (1) Well-trained intruder(s) compromise the remote access points for cyber attacks, (2) Standardized communication protocols allow intruders to analyze the substation communications, (3) Unencryptable multicast messages (e.g., GOOSE and SMV) due to the requirements, (4) Mis-configured firewalls, and (5) IEDs and user-interfaces with default passwords.

2.3.1 Substation Vulnerabilities

2.3.1.1 Unsecured Industrial Protocols

Communication protocol is an important element for the operation of a power grid. The protocol must not be modified, fabricated or monitored except by system operators. Despite their importance, cyber security features are not included in most industrial protocols since cyber security was not a major concern when industrial communication protocols were published, e.g., DNP 3.0, IEC 61850, IEC 60870-5 and Inter-Control Centre Communication Protocol (ICCP). Therefore, IEC TC 57 WG 15 established the IEC 62351 standard. The primary objective is to develop standards for security of the communication protocols defined by IEC TC 57. The GOOSE and SMV messages contain critical information and use the multicast scheme. The multicast scheme has potential cyber vulnerabilities, e.g., group access control and group center trust. Most encryption schemes or other cyber security features that delay the transmission time are not applicable for these protocols since the performance requirement of GOOSE and SMV messages is within 4 [msec]. Therefore, IEC 62351 standard recommends an authentication scheme with a digital signature using Hash-based Message Authentication Code (HMAC) for GOOSE and SMV. However, the performance test to apply the authentication scheme to GOOSE and SMV is yet to be performed. The existing intrusion and anomaly detection systems do not normally support IEC 61850 based protocols since they are more focused on general cyber intrusions such as Distributed Denial of Service attack (DDoS). In order to mitigate the communication based cyber attacks to substation automation networks, the work of [11] proposed an Intrusion Detection System (IDS) for IEC 61850 based substation automation system. An intrusion detection system for serial communication based MODBUS and DNP3 in the substations is proposed in [12]. Reference [13] proposes a temporal anomaly detection

method and [36] reports an integrated anomaly detection method for detecting malicious activities of IEC 61850 based multicast protocols (e.g., GOOSE and SMV) in the substation ICT network.

2.3.1.2 Remote Access Points

Power system components are located in wide-spread and remote sites. Remote access to substation networks using Virtual Private Network (VPN), dial-up or wireless is a common way to monitor and maintain the substation. The main problem of the remote access point is that remote access points may not be installed with adequate security features, e.g., poorly configured firewall, weak ID and password policy, bad key management for cryptography, and use of un-secured external memory (e.g., USB flash drive). Therefore, substation security managers have to consider the following actions in order to enhance the cyber security: (a) Check firewall policies and logs periodically to identify security breaches, (b) Change ID and password frequently and enhance the password policy (e.g., including numerical digits and special characters), (c) Enhance security of the key server against attacker(s), and (d) Provide security practice education for operators.

2.3.1.3 Default Password and Built-in Web Server

A typical substation may have a number of IEDs and it is difficult to manage the different passwords for each IED. Therefore, substation operators may use the default or same password for all IEDs. In addition, some IEDs and user interfaces have a built-in web server and hence it may be vulnerable to cyber intrusions, e.g., remote configuration change and control with default passwords. Substation security managers have to check the security and system logs of IEDs and user-interface to detect unauthorized access.

2.3.2 Hypothesized Intrusion Scenarios to Substations

Security threats to the substation automation system can be divided into two parts based on the physical and cyber assets. The physical assets are the hardware components, e.g., GPS (A4), IED (A5) and circuit breaker (A8), whereas cyber assets include physical and cyber resources, e.g., firewall (A2), communication network (A3) and software applications in the user-interface (A6), as illustrated in Fig. 2.4. Mitigation actions against security threats have to consider both physical and cyber intrusions.

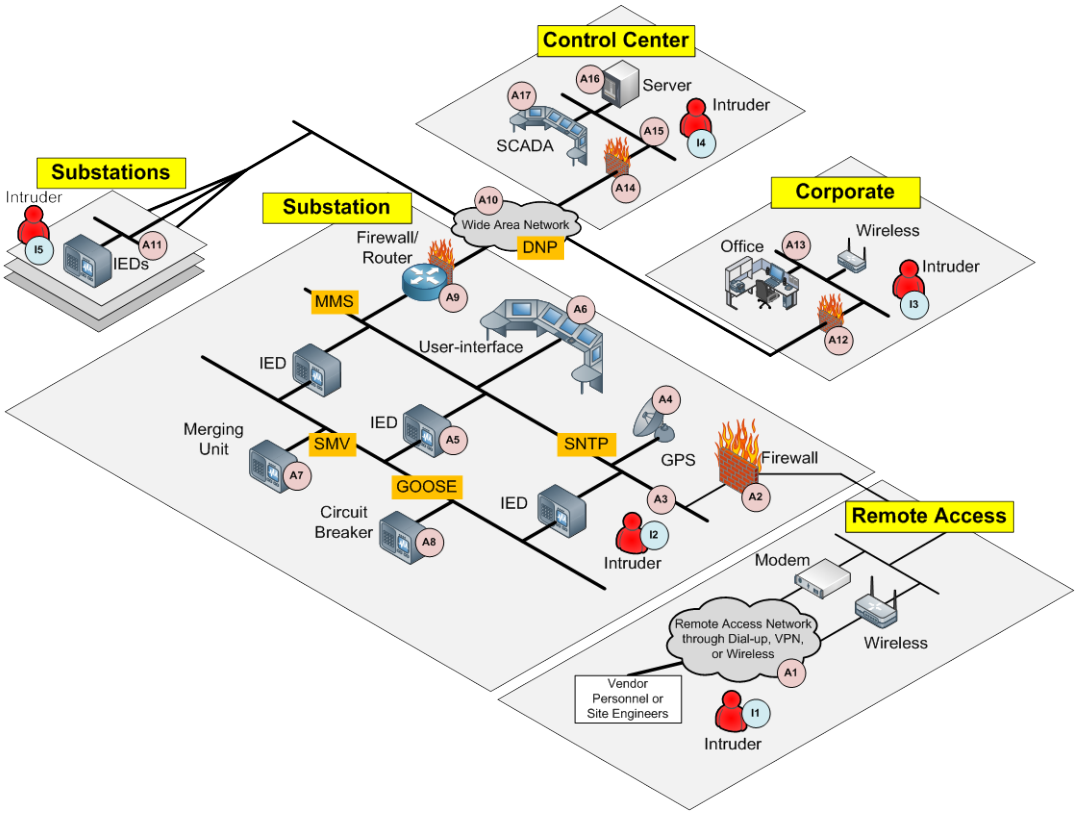

Fig. 2.4 Overview of substation ICT network diagram and security threats

Security threats to substations can be inadvertent events as well as deliberate attacks. Inadvertent events include animal intrusions, equipment failures and natural disasters [5]. Animal intrusion is a major concern for substation operators [37]. A significant amount of research has been undertaken over the last decade concerning monitoring of the health condition for substation components. Natural disasters such as flood, volcanic eruption, earthquake and tsunami, are rare but, in a severe scenario, can lead to cascading events and catastrophic outages. The work of [38] proposes weather-related power outages and enhancement of the system resiliency. Deliberate threats can be caused by disgruntled employees, cyber attackers, and malwares. Disgruntled employees can be threats for the substation security as they are familiar with the substation systems. The threats of cyber attacks are higher than before since substations need remote access connections for maintenance. Stuxnet is a relevant example of cyber threats (malwares) that are aimed at control systems of critical power infrastructure [39]. .

2.3.2.1 Single Substation Attack

As shown in Fig. 2.4, potential cyber security threats and locations of intruders in a substation automation network include:

A1: Compromise remote access points (e.g., dial-up, VPN and wireless)

A2/A9/A12/A14: Compromise firewall

A3: Gain access to substation network

A4: Interrupt GPS time synchronization

A5: Gain access to bay level devices or change protective device settings

A6: Gain access to user-interface

A7: Compromise process level devices (e.g., merging unit)

A8: Change the status of circuit breaker (e.g., close to open or vices versa)

A10: Gain access to wide area network (e.g., DNP 3.0)

A11: Gain access to neighbor substation network

A13: Gain access to corporate network

A15: Gain access to control center network

A16: Compromise the server in a control center

A17: Compromise the user-interface in a control center

I1: Intruder from outside of substation network via remote access points

I2: Intruder from inside of substation network

I3: Intruder from outside of substation network via corporate network

I4: Intruder from outside of substation network via control center network

I5: Intruder from outside of substation network via neighbor substation network

As depicted in Fig. 2.4, possible intrusions to the substation local area network can originate from outside or inside a substation network.

The following combinations represent the possible intrusion paths from outside to a local area network at a substation. Intrusions can originate from remote access points (A1) or neighbor substation network (A11) or corporate network (A13) or control center network (A15) all the way to the substation local area network (A3), e.g.,

from A1-A2-A3;

from A11-A10-A9-A3;

from A13-A12-A10-A9-A3;

from A15-A14-A10-A9-A3

Cyber attacks from inside the substation can originate from the substation network (A3) or user-interface (A6) then gain access to other facilities in the substation. Inside attack can be performed by social engineering [40]. One of the realistic examples of this attack is that intruder(s) send an email to substation operators that appears to comes from a credible source. However, this email contains a fabricated website link or malware software so once operators open this email, their PCs or laptops will be infected. After that, this malware will infect the external flash drive that plugged into compromised devices. Finally, operator(s) may use the infected flash drive at the substation network to copy documentation. Then this malware will find a path to external communication, and send all information to intruder(s) or change the setting of the protection devices (e.g., IEDs).

It is crucial to protect the substation automation ICT network against cyber attacks as a successful cyber intrusion can cause significant damages on the power grid. Once an intruder can access the substation communication network, (s)he can access other facilities in the substation. For instance, the result of cyber attack, A4, may disrupt time synchronization of all communication protocols in the substation ICT network, and operators may lose the availability of substation communications. Upon successfully cracking an ID and a password and gaining an access to the user-interface (A6), the intruder may control or modify the settings of the IEDs (A5). Then they can operate circuit breakers through the connection of IEDs. Another possibility is to gain access to the ICT network of a neighbor substation, e.g., from A9-A10-A11, then multiple cyber attacks can be carried out. More details about simultaneous cyber attacks to the multiple substations will be discussed in the following Section.

2.3.2.2 Simultaneous Attacks to Multiple Substations

Each substation has a different level of importance in a power grid. Since generally, a high voltage substation carries more power. The level of cyber security is also different at each substation. For instance, substation A uses firewall, IDS and cryptography features for cyber security mitigation whereas substation B only uses firewalls. In this example, the security level of substation A is higher than substation B whereas the cost of security implementation at substation B is lower. By analyzing the security level of each substation and importance in a power grid, an intruder may find the optimal combination (considering cost-benefit model) of target substation(s) that can trigger a sequence of cascading events, leading to a system blackout. Therefore, the impact of simultaneous cyber attacks to multiple substations can be much higher than that of a single substation attack.

2.3.2.3 Attack Tree

In the field of computer science and information technology, attack trees have been used to analyze potential threats and attack paths against cyber attacks [41, 42, 43]. However, the concept of attack trees is broadened and applied to other systems, e.g., cyber security of power systems [44, 45]. Although there are numerous concepts and definitions of attack trees, the most commonly occurring concepts are nodes (root or leaf), edges, connectors and attributes [46]. Fig. 2.5 shows a simplified attack tree for the substation automation system. Root node (T1) is the ultimate goal (i.e., open circuit breakers) with combinations of leaf nodes (T3) that do not have any predecessor. Leaf nodes (T3) contain sub goals or steps to archive the final goal (T1). Edges (T2) are connectors for all nodes. There are two types of connectors (T4) in Fig. 2.5, “AND” and “OR.” AND connector shows different steps (nodes) toward achieving the same goal. For instance, an intruder has to complete two steps, *Social Engineering* and *Compromise Operator Laptop*, in order to achieve *Obtain ID and Password*. Attributes represent features or properties relevant for numerical analysis

of security models, e.g., attack probability and cost of an attack. Fig. 2.5 shows an example model of cost of an attack. If the first priority is to minimize the attack cost, the combination of (9)-(10)-(5)-(2) is the best way to achieve the final goal. However, if the priority of attack is to minimize attack steps, (4)-(1) is the best way to open circuit breakers.

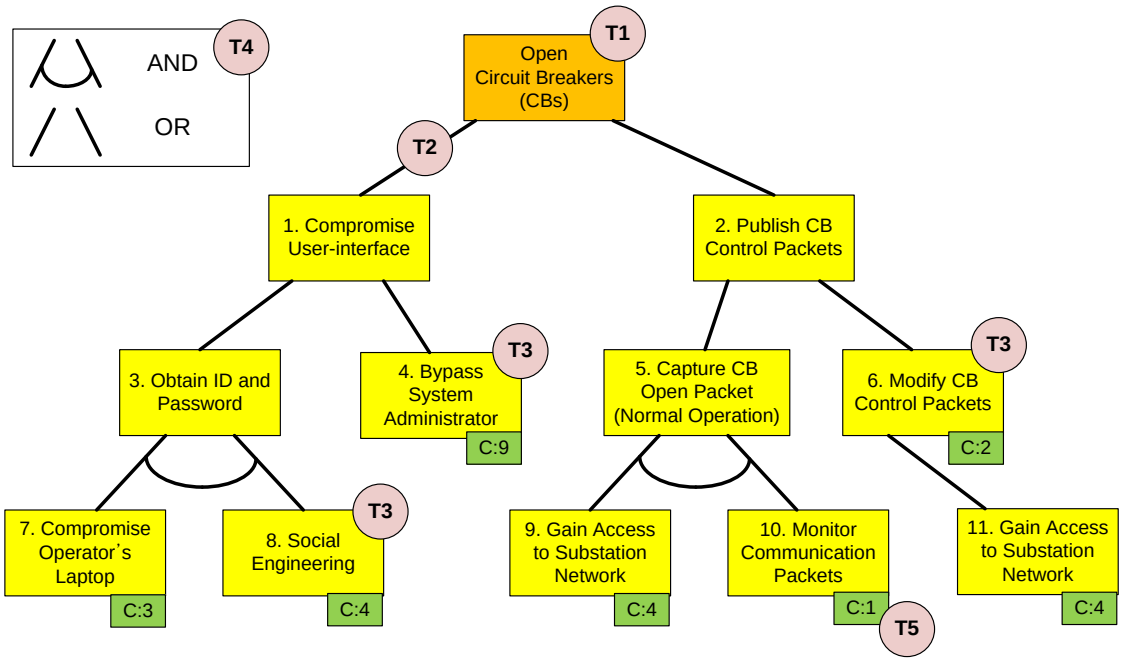

Fig. 2.5 Attack tree diagram for substation automation systems

Chapter 3. Anomaly Detection for Cyber Security of the Substations

3.1 Introduction

A power grid can become vulnerable with respect to electronic intrusions that are launched to manipulate critical cyber assets for the purpose of a cyber attack. The complexity of cascading events triggered through the substation level control systems can de-energize power system components and aggravate operating conditions by causing overloading and instability. An analytical method has been proposed to model the attack upon substations that may initiate cascading failures [3]. Cyber security of Intelligent Electronic Devices (IEDs) in the substations has been recognized as a critical issue for the smart grid [16]. One way to address these issues is to develop new technologies to detect and disrupt malicious activities across the networks. An anomaly detection system is an early warning mechanism to extract relevant cyber security events from substations and correlate these events. In the literature, methods for event correlations, such as alarm processing, fault diagnosis and security assessment for power systems have been proposed [7, 8, 9]. A survey of the important issues related to cascading events has been reported [47]. Cyber attack events may be discovered but details of such incidents are usually not publicly available. Some reports described penetration testing conducted by private companies to try to connect from an external network to internal critical cyber assets, e.g., programmable electronic devices and communication networks. It is shown that cyber assets are accessible from remote access points, e.g., modem over a landline, wireless technology, or Virtual Private Network (VPN) using a routable [48]. This dissertation is concerned with the sources of vulnerabilities due to cyber intrusions at the substations of a power grid. These vulnerabilities have been reported by National Institute of Standards and Technology (NIST) and discussed at the North American Electric

Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) Workshop on CIP 002-009 [14]. NIST also identified key attributes of the logical design for intrusion-based attacks on power equipment that is critical to standardization and modeling [15], [16]. The hypothesized intrusion scenarios in this dissertation are constructed based on the aforementioned critical access points of a typical substation setup and intrusion-based attacks.

While the information and communications technology of the power system control center infrastructure has evolved into a highly connected network environment [49], [50], technologies for detection of intrusion-based anomalies are not yet available. Intrusion detection models have been developed to monitor the system's security audit records to identify abnormal usages if there are security violations [51]. Trust-based security mechanisms have been designed to suppress cyber attacks or other malicious events for event logging, analysis, or blocking power system operations [52], [53]. Data objects for intrusion detection are categorized in the IEC 62351; however, research about the cyber system and anomaly correlations from cyber-power interactions is in an early stage [54]. A challenge on information extraction is to efficiently detect and identify the relevant events from a power system control network. Reduction of high to lower dimensional data vectors for computational efficiency is desirable. Fast information assimilation and anomaly detection models have been proposed to incorporate high dimensional data vectors from various data sources [55], [56].

Inferring anomaly requires event construction using temporal detection. Correlation techniques by a temporal approach can be used to learn from characteristics of events. A combination of transaction-based models with combined hidden Markov model and feature-aided tracking has been proposed to detect asymmetric patterns [57]. Enhancement of the previous framework is required

for two reasons [54], [58]: (1) Cyber-infrastructure can be accessible by multiple users at different locations and there remains the possibility of simultaneous attacks upon multiple substations; (2) There may be other combinations of cyber attacks upon substations and the resulting impact is not captured or observed.

In this dissertation, the proposed anomaly detection method is based on systematic extraction of intrusion footprints that can be inferred from credible intrusion events across the computer network within a substation. The contribution of this dissertation is a new substation anomaly detection algorithm that can be used to systematically extract malicious "footprints" of intrusion-based steps across substation networks. An impact factor is used to evaluate how substation outages impact the entire power system. The conceptual design of RAIM was reported [54]. The focus of [58] is on the Supervisory Control And Data Acquisition (SCADA) system, which incorporates the entire communication and control systems in the control center and substations. The concept of impact factor has been reported in [58]. This dissertation contributes to the state-of-the-art of cyber security of power grids in two areas: (1) an anomaly detection and correlation algorithm is developed, and (2) an impact evaluation method is proposed based on the detected anomalies. The result is a new monitoring mechanism for cyber security of the computer network at multiple substations in order to enhance resilience of the power grid.

Section 3.2 provides a generalized intrusion scenario. Section 3.3 describes a prototype design for Real-time monitoring, Anomaly detection, Impact analysis, and Mitigation strategies (RAIM) [54]. Section 3.4 is concerned with anomaly construction based on temporal events. Section 3.5 provides an attack analysis with the identification of classes of contingencies with different levels of

complexity. Section 3.6 is a case study of simultaneous impact evaluations based on an anomaly set. Section 3.7 provides the conclusion and recommendations for the future work.

3.2 Hypothesized Intrusion Scenarios

As depicted in Fig. 3.1, the following combinations represent the possible intrusion paths through remote connections to a local area network at a substation.

- Any point of (A1, A2, A3)-B1-B2
- Any point of (A1, A2, A3)-B3-B1-B2

Each combination includes connections through remote dial-up or VPN to substation level networks targeted on substation user interface or IEDs. Once the local network is penetrated, a cyber attack can be launched through

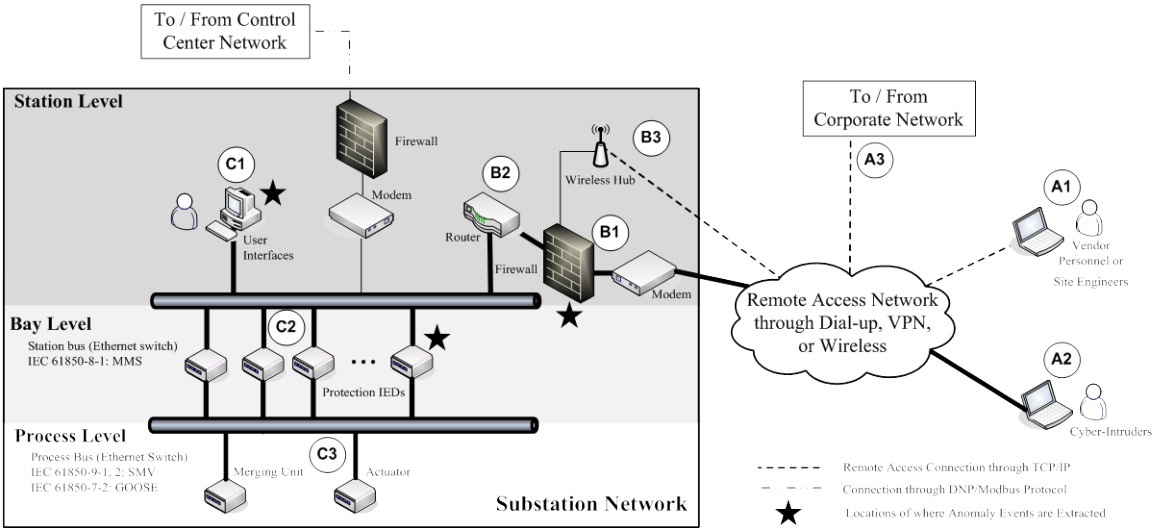

Fig. 3.1 Path combinations of intrusion scenarios to substation level networks (Bold lines)

- User interface, C1,
- Direct IED connection, C2, or
- Eavesdropping and data packet modification, C3.

Note that a CIGRE survey on the wireless security has been conducted [59]. Discussion on intrusion scenarios through local wireless connection is outside the scope of this dissertation. The following subsection describes the steps to execute the two possibilities through C1 or C2 to cause a disruption:

1) Accessing Substation User Interface: The user interface provides a direct access to the substation communication. Upon successful penetration into the user interface with the highest access privilege, an intruder would be able to utilize the console and explore information by enumerating switching devices in the local network. Breaker opening commands can be sent once the local

controllable parameters are identified.

2) Accessing Substation IEDs: Upon successfully cracking a password and gaining access to an IED, an intruder can access the Substation Configuration Description (SCD) file which contains the one-line diagram of the substation, communication network, composition of IED and data flow based on IEC 61850 [60], [61]. Once the required information is identified, actions to operate circuit breakers can be launched through direct IED connection.

3.3 Prototype of RAIM

Network and Security Management (NSM) abstract data objects have been proposed in IEC 62351, which mainly describe anomaly properties based on (1) Unauthorized access, (2) Communication protocol monitoring, and (3) System health [62]. This information can be used for event constructions to extract evidences for intrusions. Fig. 3.2 describes the data object model abstraction for the RAIM framework [54].

The current dissertation provides an anomaly detection algorithm for substation-based intrusions. The proposed substation RAIM model in Fig. 3.2 is divided into 3 data object models, *i.e.*, *RAIM-SStationComp*, *RAIM-BayIED*, and *RAIM-SStation ConnType*.

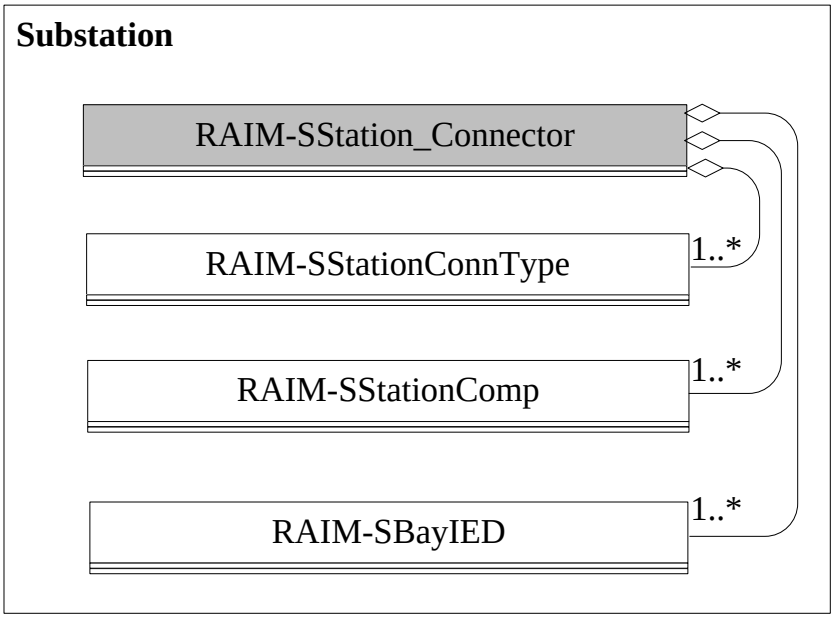

Fig. 3.2 The object modeling of RAIM for substation

A *RAIM-SStationComp* consists of functions of status, security, extractor, alarm, and log instrumenting features. A status determines the status of substation computers and running applications. It also defines maximum numbers of connections on the user interface, as well as determining the response time of each computer. This can be used to verify the source IP address of established connections and timeframe for each connection. A security method uses encryption, authentication, and compression that creates, distributes, and decrypts used in the function [63]. The failed logon feature is used to identify credible intrusion threat with respect to the time and frequency, *e.g.*, consecutive failed logon attempts within a short period of time. A list of user privilege on Operating System (OS) and substation applications are maintained. An alarm attribute is the accumulative violation messages that are set in a system to infer a credible list of potential cyber intrusion. A *RAIM-BayIED* encompasses similar features of *RAIM-SStationComp* except with an additional flag indicator to validate if parameter settings have been changed with a time stamp on

each change. This applies to both Abstract Communication Service Interface (ACSI) and Substation Configuration Language (SCL) [64].

The *RAIM-SStationConnType* is the communication type perimeter settings that will constantly update the type of connection, *e.g.*, either through dial-up network or VPN. Anomaly detection is the discovery of symptoms resulting from malicious attempts that can be inferred based on their footprints. Detection is performed based on repeated failed password logon, increased file size or additional executable files, and undesirable changes of critical settings to local machines that operate the physical equipment. An anomaly inference system of RAIM prototypes is designed based on the hypothesized intrusion scenarios with the following attributes.

1) Failed logon statistics upon local user interface compute- rs or IEDs
2) The changes of file systems on local user interface
3) The changes of IED critical settings that may mis-operate the system operations.

These attributes are illustrated in Fig. 3.2. The intrusion attempts on each device or computer in substation level network are included in RAIM-SStationSBayIED.attempts and RAIM-SStationComp.attempts, respectively. The file changes and updates are part of the RAIM-SStationSBayIED. CriticalSettings, where the update of the IED critical settings is described in RAIM-SStationComp.FileSystems.

3.4 Temporal Event Constructions

If an attacker does not know the login information for user interface or IED, (s)he may attempt to find it. Hence, failed logon attempts are recorded, and the device will be locked down if and when the number of failure attempts exceeds a preset threshold. Upon a successful electronic intrusion, an attacker is able to control the user interface or IEDs in a substation. The attacker can tamper with the authentication to keep legitimate users from logging into the user interface. The attacker then performs malicious actions, *e.g.*, once the attacker changes the tap settings on main transformers (MTRs). If the MTRs are operated in a parallel mode, closing the sectionalizing circuit breakers (CBs) between them can cause a damaging circulating current flow between the MTRs. As an attacker successfully opens CBs, a further action may be the all-trip condition to all MTR CBs. Then, all transmission or distribution lines connected with the substation will be disconnected.

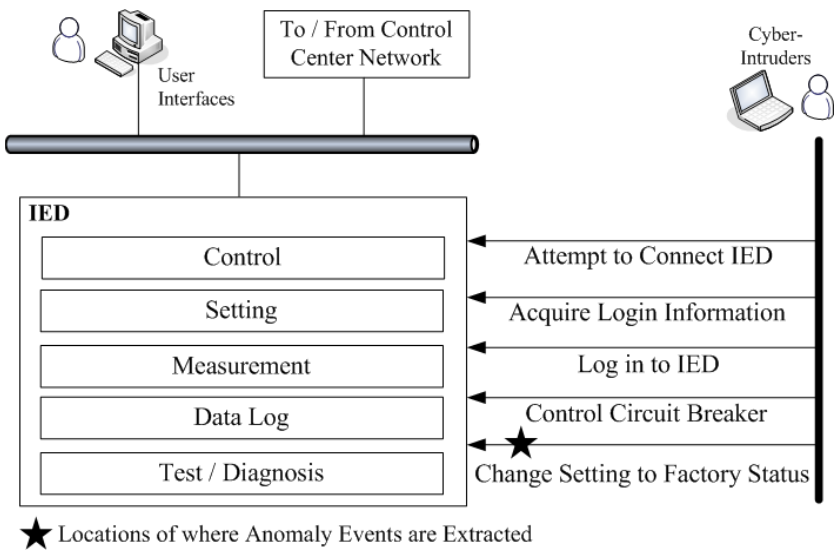

Fig. 3.3 Anomaly detection of a cyber attack at the IED level

As shown in Fig. 3.3, the proposed anomaly detection algorithm is to detect intrusions when unexpected actions are being taken by one or more attackers. A successful logon to IED will allow the attacker to execute functions to restore the IED settings back to its factory status from the 'Test / Diagnosis' menu, and IED will lose all user configuration files that are crucial for system operation. When the attacker attempts to execute this function, the proposed anomaly detection algorithm detects an attempt to change a setting without authorization. The step in which the attempt is detected is marked with a star in Fig. 3.3. Although the operator will recognize the loss of an IED connection after this intrusion, details of the condition may not be known until a further investigation is conducted.

Two domain-specific cyber attacks are highly relevant for power infrastructure control systems: (i) *Night Dragon* [65] and (ii) *Stuxnet* [66]. The steps of these cyber attacks and their malicious characteristics are based on: (1) intrusion attempts, (2) change of the file system, (3) change of target system's setting, and (4) change of target system's status. These 4 parameters capture the malicious intrusion behaviors across all substation-level networks and are key attributes for improving situational awareness of cyber intrusions.

As described in Section 3.3, anomaly detection will rely on data logs at the substation level networks, including IEDs. Several of binary (0, 1) status indicators are defined here: π^{a}, indicates the detection of intrusion attempts upon computers or IEDs, π^{fs} represents a change of the file system, π^{cs} denotes a change of IED critical settings, and π^{o} is for a change of status on switches. The weight factors associated with each status indicator can be represented by a row vector, *i.e.*,

$$\boldsymbol{\pi}_{(1\times k)} = \begin{pmatrix} 1 & \alpha\pi^{a}_{(T\times L)} & \beta\pi^{fs}_{(T\times M)} & \delta\pi^{cs}_{(T\times N)} & \varepsilon\pi^{o}_{(T\times O)} \end{pmatrix} \tag{3-1}$$

The first element of the row vector, 1, is an assigned value to avoid singularity of a zero vector in further calculations using the row vector. Each element of the vector (except the first element) carries a weighting factor. The weighting factors, α, β, δ and ε are associated with the existence of intrusion attempts, file system, IED settings, and switching actions, respectively. The values assigned to each of the weighting factors are based on the relationship of $\alpha < \beta < \delta < \varepsilon$. The symbols, L, M, N, and O, denote the size of each element in Eq. (3-1). T is the number of records of anomaly for each period of time. An example is given in Eq. (3-2) as an example of the attributes described in Eq. (3-1) with L = 1, M = 1, N = 1, and O =1 and weighting factor parameters, $\alpha = 1$, $\beta = 5$, $\delta = 10$, $\varepsilon = 20$. Note that changes of IED critical settings and status of switching devices are given higher weights. An example generated randomly shows that an intrusion attempt exists and that there is an indication that IED critical settings may be changed. That is,

$$\boldsymbol{\pi}_{(1\times5)} = (1 \quad 1 \quad 0 \quad 10 \quad 0) \tag{3-2}$$

This is the metric to determine the anomaly between two periods of snapshots. If a discrepancy exists between two different periods, the value of Δ_{ta} is a number between 0 and 1. A value of 0 implies no difference whereas 1 indicates the maximal discrepancy. A significant value of Δ_{ta} serves to indicate an anomaly. The example below describes a temporal anomaly for a sequence of 7 time instance, *i.e.*,

$$\boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 20 \\ 1 & 1 & 0 & 0 & 20 \\ 1 & 1 & 0 & 0 & 20 \\ 1 & 1 & 5 & 0 & 20 \\ 1 & 1 & 5 & 0 & 20 \\ 1 & 1 & 5 & 10 & 20 \\ 1 & 1 & 5 & 10 & 20 \end{bmatrix} \tag{3-3}$$

The matrix $\boldsymbol{\Pi}$ contains a number of row vectors for the same substation. This will be used for detection of temporal anomalies by comparing several row vectors representing a consecutive sequence of time instants. Normalization is conducted row by row for matrix $\boldsymbol{\Pi}$. The vector norm of a row vector $\boldsymbol{\pi}$ is defined by $\|\boldsymbol{\pi}\|_2 = \sqrt{\sum_{i=1}^{K} \pi_i^2}$ where K=1+L+M+N+O is the dimension of the vector. That is, for each row the normalized vector is

$$\hat{\boldsymbol{\pi}} = \frac{\boldsymbol{\pi}}{\|\boldsymbol{\pi}\|_2} \tag{3-4}$$

The resulting matrix is denoted by $\hat{\Pi}$, i.e.,

$$\hat{\boldsymbol{\Pi}} = \begin{bmatrix} 0.0499 & 0 & 0 & 0 & 0.9988 \\ 0.0499 & 0.0499 & 0 & 0 & 0.9975 \\ 0.0499 & 0.0499 & 0 & 0 & 0.9975 \\ 0.0484 & 0.0484 & 0.2420 & 0 & 0.9679 \\ 0.0484 & 0.0484 & 0.2420 & 0 & 0.9679 \\ 0.0436 & 0.0436 & 0.2178 & 0.4356 & 0.8712 \\ 0.0436 & 0.0436 & 0.2178 & 0.4356 & 0.8712 \end{bmatrix} \tag{3-5}$$

Temporal anomaly is determined by the two vectors that occur at two different time instants [67]. The anomaly that occurred between the two time instants is determined by the normalized row vectors. A scalar parameter for the temporal anomaly is defined as

$$\Delta_{\text{ta}} \;=\; 1 - \frac{\hat{\boldsymbol{\pi}} \cdot \hat{\boldsymbol{\pi}}_{-1}^{\text{T}}}{\|\hat{\boldsymbol{\pi}}\|_2 \cdot \|\hat{\boldsymbol{\pi}}_{-1}\|_2} \tag{3-6}$$

The resulting matrix is denoted by $\hat{\boldsymbol{\Pi}}$. Based on Eq. (3-5) one can obtain a column vector for temporal anomaly that provides irregularities of events over a certain time period. *i.e.*,

$$\Delta_{\text{ta}}^{\text{T}} = (0 \;\; 0.0012 \;\; 0 \;\; 0.0297 \;\; 0 \;\; 0.0999 \;\; 0) \tag{3-7}$$

The first vector of $\Delta_{\text{ta}}^{\text{T}}$ is 0 because there is nothing to compare to for the first row of $\hat{\boldsymbol{\Pi}}$ as this procedures starts. After the first element of Eq. (3-7), the second element is the value resulting from the calculation based on the first and second rows. Other elements are obtained in a similar manner. For a given substation, a matrix $\hat{\boldsymbol{\Pi}}$ is formed by normalizing the matrix $\boldsymbol{\Pi}$ as illustrated in Eq. (3-6). The rank of $\hat{\boldsymbol{\Pi}}$ for this substation is used to determine an index ζ, *i.e.*,

$$\zeta \;= \text{rank}\left(\hat{\boldsymbol{\Pi}}\right) - 1 \tag{3-8}$$

Based on Eq. (3-8), $\zeta = 0$ implies that there is no anomaly event on this substation. If the rank of ζ for a substation is greater than or equal to 1, the substation will be included in the credible list that will be considered for further evaluations. This will be correlated with other substations.

In order to quantify the likelihood of an anomaly, a vector **p** is used to denote a column that is calculated from the anomaly corresponding to the weighted anomaly entities from $\widehat{\mathbf{\Pi}}$ and the temporal anomaly Δ_{ta}. B is used to represent an n × m matrix with all elements equal to 1. That is,

$$\mathbf{p} = \begin{cases} 0 & \text{if } \zeta = 0 \\ \widehat{\mathbf{\Pi}} \cdot \mathbf{B} \cdot \Delta_{ta} & \text{Otherwise} \end{cases} \tag{3-9}$$

The observation over a certain time period can be made throughout all substations using Eq. (3-9). For the example in Eq. (3-5), $\zeta = 4 - 1 = 3$. Hence, $\mathbf{p}^T$ = (0.1372, 0.1435, 0.1435, 0.1709, 0.1709, 0.2109, 0.2109). To include a substation in the evaluation list, the value of an index for intrusion credibility, denoted by ϱ, is evaluated based on the difference between maximum and average values of **p**, *i.e.*,

$$\varrho = \max \mathbf{p} - \bar{\mathbf{p}} \tag{3-10}$$

The intrusion credibility index ϱ is used to identify the substation candidates to be included in the credible list, and is only included in the list when $\varrho > \varrho*$, where $\varrho*$ is the threshold value. This continues for all substations on the operational list until all substations in the list are examined. The value of ϱ for the given example is 0.0812.

3.5 Simultaneous Attack Events

A combination of cyber attack events upon multiple substations is determined based on credible high impact threats from anomaly inference. The complexity of scenarios for cyber attacks on 1

substation or 2 is lower than that of 3 substations or more (simultaneously). For the use in vulnerability assessment, three categories of scenarios are proposed, *i.e.*,

1) Critical Substations: This list contains critical substations of a power system. A substation is included in this list if its removal (de-energization) from the grid under a normal operating condition results in a non-convergent power flow computation. Such a non-divergent condition is an indication of an infeasible operating condition, *e.g.*, voltage collapse.

2) Single and Double Substations: This category of substati- ns does not include the critical substations above. If credible malicious activities are detected, evaluations with respect to credibility and the resulting impact will be performed. Ranking for each event will be sorted in a descending order. The selection of evaluations is considered for intrusion of one or two substations, *i.e.*, $k \leq 2$.

3) Multiple Substations: The event of simultaneous cyber attacks on 3 or more substations is more complex. This list evaluates the impact by removing the k substations that result in a higher impact and serves as a message to power dispatchers.

The steady-state and dynamic behaviors of a power system under a cyber attack can be studied using power flow simulation tools. Evaluation of a power system under cyber attacks can be performed by de-energizing substation(s). As mentioned earlier, a failure to obtain a steady-state power flow solution is an indication of a major impact that may lead to a power system collapse. The impact of isolating a substation in the overall system is measured by an impact factor corresponding to the substation. This measure represents a worst case analysis for the impact of a

cyber intrusion such as a single-, double- or multiple substation scenarios. It is assumed that a cyber attack is intended to utilize the direct control or functions embedded in the control network in order to disconnect the substation components from the power system. The impact factor introduced in the work of [58] is applicable for analysis of cyber attacks on substations. The proposed impact factor is determined by a ratio and a loading level, L, where Loss of Load (LOL) is the total loss of load as a result of the hypothesized cyber attack on the substation(s). To evaluate the impact factor, the ratio of the loss of load and the total load is determined first. Then an exponent L starts at 1 with an increasing incremental step, *e.g.*, 0.01. Each step is validated with a power flow calculation. The impact factor is a measure of how close the power system is to a collapse, which is indicated by non-convergence of the power flow computation. As the LOL due to isolation of a substation increases, the importance of the substation also increases. If the exponent L is higher than 1, it means that the power system is further away from a collapse. When it is 1, the net exponent is 0 and it is an indication that the system is closer to the collapse point. This process continues until it fails to obtain a power flow solution. The parameter L* denotes the value of L where power flow divergence occurs. The impact factor γ is defined by

$$\gamma = \left(\frac{P_{LOL}}{P_{Total}}\right)^{L^*-1} \tag{3-11}$$

Note that L* = 1 leads to an impact factor $\gamma = 1$, which represents the highest level of impact.

By identifying the vulnerability at k substations, the overall system wide vulnerability index can be obtained by

$$V_{sub}(S_1, \cdots, S_k) = \gamma_{1,\cdots,k} \cdot \max(\varrho_1, \cdots, \varrho_k) \quad (3\text{-}12)$$

Where 1,..k represent the hypothesized outage involving k substations and $\gamma_{1,\cdots,k}$ is the corresponding impact of the k-substation outage. Since ϱ for various substations may be different, the maximum value is selected for the worst case among the list, which determines the overall vulnerability index $V_{sub}(S_1, \cdots, S_k)$.

Determining the vulnerability indices and ranking will help to identify the most vulnerable cases. The overall vulnerability depends on the intrusion credibility indices and the impact factor as a result of the cyber attack. The values of impact factors are determined by power flow evaluations. The proposed method uses an algorithm to estimate the point of failure to converge. The credibility of intrusion depends on the anomaly detection data logs from the k substations. It is assumed that data logs are accessible for the purpose of anomaly detection. Communication between substations is not required for the proposed method. The temporal anomaly feature enables the proposed detection scheme to be performed continuously.

3.6 Simulation Results

The proposed methodology for evaluation of the impact of cyber intrusions at a substation level is validated using the modified IEEE 118-bus system model. This research provides a method to identify critical substations or vulnerable areas of the power system. The impact analysis can be performed through (1) randomly selected substations, or (2) user selected substations. In this modified 118-bus model, buses retain the same numbering convention except for those substations with more than one bus. A diagram for this modified system is shown in Fig. 3.4.

The IEEE 118-test system, as described in Fig. 3.4, consists of 109 substations with a total load of 4,266 MW. The hypothesis here is that an intrusion into a substation will lead to operation of

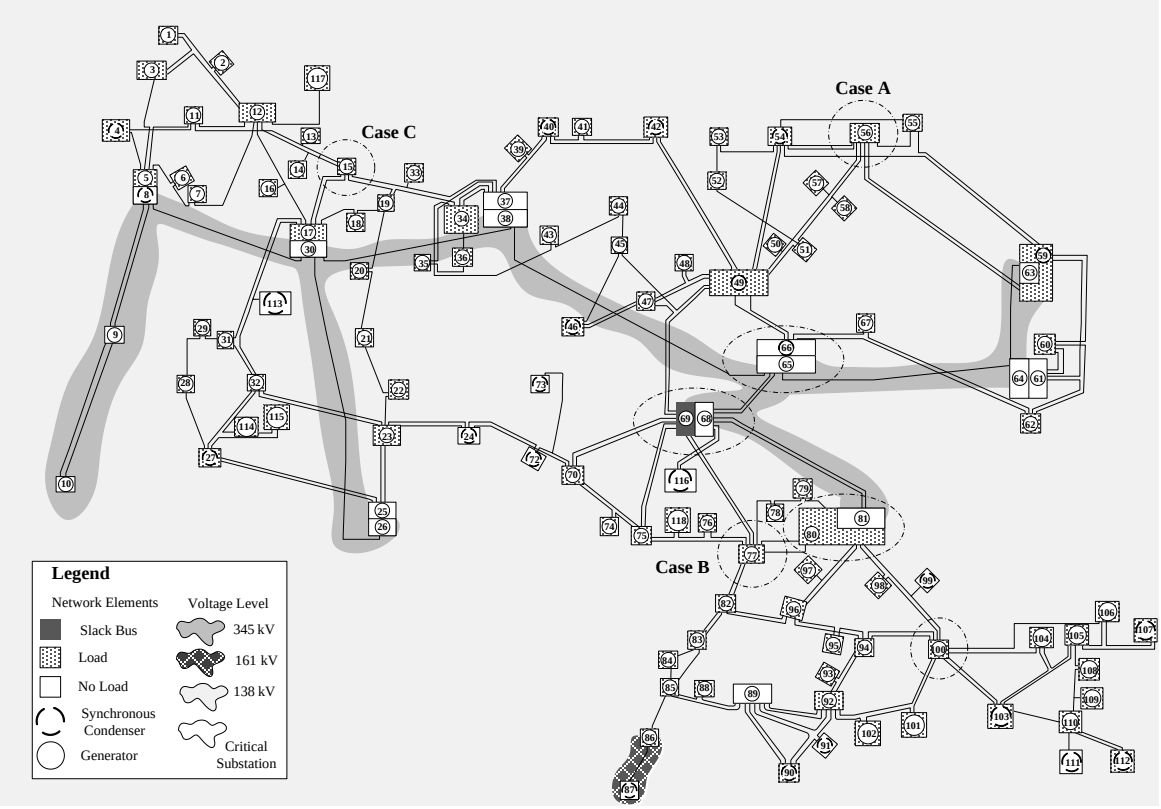

Fig. 3.4 IEEE 118-test system

physical circuit breakers and will isolate the chosen substation from the power grid. This intrusion simulation is to operate all breakers in the substation, which is a worst case scenario for the impact analysis. All intrusion logs are captured by the proposed anomaly detection algorithm.

There are 3 cases for the simulation results. Case I shows a simultaneous attack that leads to a more severe outcome relative to a single attach, *i.e.*, non-convergent power flow result. The data logs are obtained from an IED. The matrix **Π** in Eq. (3-5) is generated from IED logs, *i.e.*, unauthorized setting changes and open commands in the IED. Case II is concerned with the evaluation of vulnerability indices for 345-kV substations, the highest voltage substations for the IEEE 118-Bus system. As a result of vulnerability index calculation, the chosen scenarios, which lead to isolation of 345-kV substations, have the highest vulnerability level. Case III shows how the proposed method can identify the worst cases where cyber vulnerability improvements are most desirable.

3.6.1 Case Study I: Simultaneous Attack

Table 3.1 provides the results of intrusion simulations for a single- and a double-substation cases. For the single-substation case, the loading level L* of substation 49 is obtained at 1.6629 after running power flow 47 times. The impact factor obtained is 7.574718×10^{-2}. The second case in Table 3.1 is an example of cyber attacks on 2 substations. The impact caused by the intrusion is the removal of substations 49 and 2526. Since the power flow fails to converge, the loading level L*=1.0, which results in the highest impact.

In case attackers already know the username and password of a substation user interface or an IED, it will simply bypass the step, *i.e.*, there is no intrusion attempt, which will result in 0 for the password attempt attributes. The following case assumes that cyber attackers already know the

username and password of a substation user interface or an IED. As a result, they are able to execute a switching action on the circuit breaker without making a password attempt log.

Tables 3.2 and 3.3 report sample IED logs of substations 49 and 2526, respectively. Table 3.2 includes messages indicating an intrusion into substation 49, leading to a change of settings. It represents an unauthorized change of settings for a protective device when cyber attackers know the password for IED control software. Table 3.3 provides the unauthorized commands to open 3 circuit breakers and 2 disconnect switches in substation 2526, assuming that cyber attackers know the substation logon credentials.

Table 3.1: Hypothesized cyber attack upon single and multiple substation(s)

	Cyber Attack upon a Substation	Simultaneous Attack upon Two Substations
Substation(s) Time Elapsed Loading Level Loss of Load	49 8.975412×10^{-1} s 1.662900 87 MW	49 and 2526 8.398672×10^{-2} s 1.0 378 MW
Impact Factor	7.574718×10^{-2}	1.0

Table 3.2: IED logs of substation 49

Substation 49				
No.	**Date**	**Time**	**Contents**	**Issue**
47	15.09.2010	10:28:59,609	50	Unauthorized Setting Change
48	15.09.2010	10:29:57,629	51	Unauthorized Setting Change
49	15.09.2010	10:30:02,368	87	Unauthorized Setting Change
50	15.09.2010	10:31:21,523	87T	Unauthorized Setting Change
51	15.09.2010	10:32:20,594	21	Unauthorized Setting Change

Table 3.3: IED logs of substation 2526

Substation 2526				
No.	**Date**	**Time**	**Contents**	**Issue**
45	15.09.2010	10:28:33,560	Breaker 1	Unauthorized Open command
46	15.09.2010	10:29:43,159	Breaker 2	Unauthorized Open command
47	15.09.2010	10:30:04,368	Disconnect Switch 1	Unauthorized Open command
48	15.09.2010	10:31:14,270	Breaker 3	Unauthorized Open command
49	15.09.2010	10:32:23,237	Disconnect Switch 2	Unauthorized Open command

Based on the logs in Table 3.2 and 3.3, matrix **Π** is constructed for each of the two substations, 49 and 2526. The results are

$$\mathbf{\Pi}_{\text{sub }49} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \end{bmatrix} \tag{3-13}$$

$$\mathbf{\Pi}_{\text{sub }2526} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \end{bmatrix} \tag{3-14}$$

The impact factor of cyber attack on single substation 49 is 7.574718×10^{-2} and coordinated cyber attack on two substations 49 and 2526 is 1, as shown in Table 3.1. The intrusion credibility index of cyber attack on single substation 49 is 0.7504 and coordinated cyber attack on two substations 49

and 2526 is 0.7917. The indices are calculated by Eq. (3-1) ~ (3-10) based on the IED logs in Eq. (3-13) and (3-14), *i.e.*, $\Pi_{\text{sub }49}$, $\Pi_{\text{sub }2526}$, respectively. Therefore, by Eq. (3-12), the vulnerability index of cyber attack on single substation 49 is 0.05684 and coordinated cyber attack on two substations 49 and 2526 is 0.7917 since $\max(\varrho_{49}, \varrho_{2526})$ is 0.7917.

3.6.2 Case Study II: Most Critical Substation

The system given in Fig. 3.4 has four important substations, the removal of each of them will result in non-convergent power flow. They are substations 100, 6566, 6869, and 8081 that are enclosed in an oval or a circle in Fig. 3.5. Hence, these 4 cases are categorized as critical.

In order to identify other critical cases, the remaining 345 kV substations are included in a list for vulnerability assessment using the proposed method.

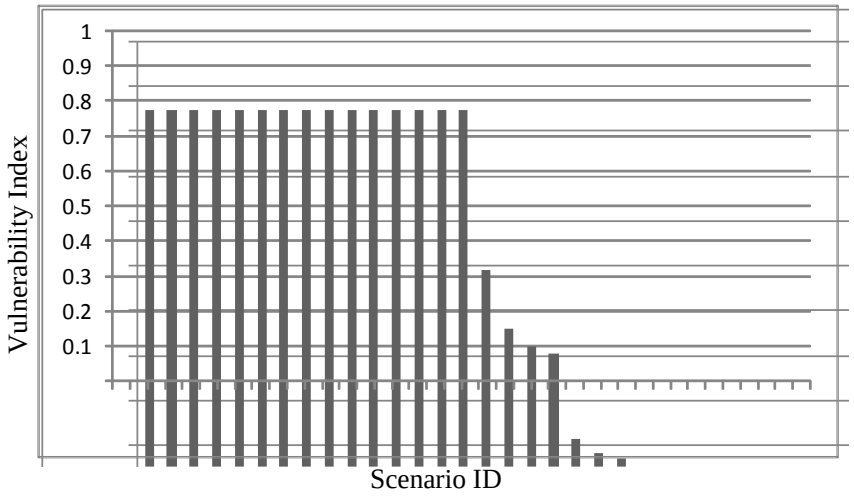

Fig. 3.5 Vulnerability ranking of enumerating credible events for $\mathbf{Subs_{case}}$

Table 3.4: Critical scenario for $\mathbf{Subs_{case}}$

	Scenario Number	Max ϱ	Impact Factor	Vulnerability Index
1	9, 6164	6164	1	0.8477
2	5899, 6164	6164	1	0.8477
3	2526, 3738, 6164	6164	1	0.8477
4	2526, 5963, 6164	6164	1	0.8477
5	1730, 3738, 6164	6164	1	0.8477
6	10, 2526	10	1	0.8468
7	9, 3738	3738	1	0.8464
8	5899, 3738	3738	1	0.8464
9	2526, 3738, 5963	3738	1	0.8464
10	9, 2526	2526	1	0.8453
11	5899, 2526	2526	1	0.8453
12	9, 5963	9	1	0.8446
13	9, 1730	9	1	0.8446
14	5899, 1730	1730	1	0.8441
15	5899, 5963	5899	1	0.8437

The critical anomaly list is substations $\mathrm{Subs}_{\mathrm{case}}$ = (5899, 9, 10, 2526, 1730, 3738, 5963, 6164). All of them are 345 kV substations. There are a total of 40 combinations of substation outages. The vulnerability index for each scenario, 1...40, is shown in Fig. 3.5. Fig. 3.5 shows the highest impact scenarios of single- and double- and multiple-substations in the set. Scenarios with ID 1…15 are highly vulnerable combinations of substation outages. They are listed in Table 3.4. Among the 15 scenarios 4 combinations involve more than 2 substations, *i.e.*, (2526, 3738, 5963), (2526, 3738, 6164), (2526, 5963, 6164) and (1730, 3738, 6164). The remaining 11 scenarios represent single-

and double-contingencies: (9, 1730), (5899, 1730), (5899, 2526), (5899, 3738), (5899, 5963), (5899, 6164), (9, 2526), (9, 3738), (9, 5963), (9, 6164), (10, 2526). Note that the impact factor for all these 15 scenarios is 1., as shown in Table 3.4.

Fig. 3.5 shows 40 scenarios in single, double, and multiple substations, the first 18 are above the mean value of vulnerability index ($\overline{V}$ = .31264). The case is ranked based on the vulnerability level of each combination. The mean value helps to identify the critical combinations in the single, double, and multiple substations for further investigation.

3.6.3 Case Study III: Highest Impact Factor Scenarios

Three more cases are selected for this study. In Table 3.5, each study case contains 17 substations. Single-, double- and multiple-substation contingencies are selected from among the 17 candidate substations. The purpose here is to illustrate how the proposed method can be used to identify the worst cases where cyber vulnerability improvements are desirable. Based on the proposed algorithm, the worst cases for Study Case A, B, C are listed in Table 3.4. For Case A, it is seen that the top 5 scenarios involve the same substation, 56. The impact factor is 1, the highest level. This is an indication that substation 56 is critical for anomaly detection and monitoring. It is also seen that Study Case B identifies substation 77 and the results of Study Case C points to substation 15 as the critical substation.

Table 3.5: Case setup for simulation

Study Case	Scenario
A	77, 67, 85, 42, 58, 62, 49, 15, 87, 56, 9, 104, 74, 1730, 97, 27, 35
B	15, 16, 23, 29, 32, 34, 41, 42, 47, 53, 57, 77, 91, 92, 96, 109, 110
C	45, 18, 90, 5899, 85, 99, 95, 47, 24, 77, 58, 62, 49, 15, 87, 56, 9

The computational performance of the proposed cyber vulnerability assessment algorithm is shown in Table 3.7. It is seen that for Study Case A, it takes 321.82 sec to complete the computation of 153 + 94 = 247 scenarios of single-, double-, and multiple-substation contingencies.

Table 3.6: Critical 5 scenarios for each study case in Table 3.5

	Scenario Number	Max ϱ	Impact Factor	Vulnerability Index
Case A				
1	49, 56	56	1	0.84979
2	56, 9	56	1	0.84979
3	56, 104, 27	56	1	0.84979
4	56, 74, 27	56	1	0.84979
5	77, 58, 56, 1730	56	1	0.84979
Case B				
1	77, 96	77	1	0.84710
2	77, 109	77	1	0.84710
3	23, 42, 77	77	1	0.84710
4	15, 16, 23, 77	77	1	0.84710
5	15, 42, 47, 53, 77	77	1	0.84710
Case C				
1	5899, 15	15	1	0.84797
2	15, 9	15	1	0.84797
3	45, 85, 49, 15	15	1	0.84797
4	77, 49, 15	15	1	0.84797
5	99, 47, 49, 15, 87	15	1	0.84797

Table 3.7: Number of scenarios and calculation time in Table 3.5

	Number of Single and Double Substations Scenarios	Number of Multiple Substations Scenarios	Total Calculation Time for all Scenarios (sec)
Case A	153	94	321.82
Case B	153	69	352.57
Case C	153	85	291.51

The proposed anomaly detection algorithm can also be applied to a physical intrusion by manipulating the microprocessor-based devices in substations. The malicious behaviors can be captured, *e.g.*, execute disruptive switching actions by attempting to logon to IED directly or trying to change IED settings manually. All these "footprints" will be logged in the device and captured through the 2 remaining parameters that are (3) change of IED critical setting and (4) change of status on switching. The method will assign the value 1 for (3) and (4), respectively.

As described in Section 3.4, once vulnerability indices are computed, cyber security can be enhanced by different measures. Dispatcher and security analysts are able to temporarily disable the communication functions to disconnect remote connections to malicious users. This requires integration of boundary protection with the proposed anomaly detection framework. IEC 1686 standard recommends that an enhancement is needed for cyber security in substation IEDs [68]. For an IED, a combination with numbers, characters, and capital and lower cases is needed for the password construction. Different privilege IDs and passwords are helpful to identify the administrator, system engineers, and switch operators. A password threshold and auto timed logout is useful for preventing cyber attacks. It is important to generate real-time audit logs including password, control, setting and measurements. The audit logs can be used for the proposed anomaly detection algorithm and analysis of intruder's attack patterns.

Chapter 4. Integrated Anomaly Detection for Cyber Security of the Substations

4.1 Introduction

A smart grid is an enhanced power grid that generates, transmits and uses electricity with the support of information and communications technology (ICT) for advanced remote control and automation [69, 70]. Smart grid has the potential to benefit power systems and customers, such as improved reliability, efficiency and reduced costs. For example, with advanced automation technology, a power grid can identify and isolate the faulted area(s) and restore unaffected areas by self-healing technologies [2]. Smart meters allow data acquisition from the customers to be conducted frequently and enable customer participation through various demand side response mechanisms [71]. Automation of the power grid includes substation and distribution automation. The subject of smart substations is a critical issue for the smart grid as it plays an important role in advanced monitoring and control of the power grids. The substation is installed with critical devices and communication networks such as IEDs, transformers, distribution feeders, circuit breakers, and communication systems. A smart substation enhances reliability and efficiency of operation, monitoring, control and protection [70].

Cyber security of substations has been recognized as a critical issue [72]. For example, well organized simultaneous cyber attacks to multiple substations can trigger a sequence of cascading events, leading to a system blackout [3, 4]. Therefore, an effective measure to address this issue is to prevent, detect, and mitigate malicious activities at the substations. Anomaly detection refers to

the task of finding abnormal behaviors in data networks; it is a concept widely adopted in computer networks [73]. The term, Intrusion Detection System (IDS), is also used for cyber security in a substation. The concept of Intrusion Detection System (IDS) was proposed by [74]. It monitors user access logs, file access logs, and system event logs to see if there is any anomaly in the host system. The work of [75] provides a model of an IDS that became a starting point of the recent IDSs. This model uses statistics for anomaly detection and an intrusion detection expert system (IDES). Typical approaches to intrusion detection are either network or host-based methods. A network-based IDS (NIDS) collects packets from a communication network and analyze network activities. References [76] and [77] propose network-based anomaly detection systems. A host-based IDS monitors a host system and generates alarms when anomalies and malicious activities are observed. The authors of [78] and [79] propose host-based anomaly detection. However, both network- and host-based intrusion detection methods have their own weaknesses. For example, host-based detection can fail to detect multiple hosts or applications. Network-based detection, on the other hand, can have a high rate of false alarms. In [80] and [81], the authors propose an integrated (or hybrid) anomaly detection system in order to compensate for the weakness of each system. The work of [11] proposes an intrusion detection system for IEC 61850 automated substations. A cyber-physical security vulnerability index has been proposed [44]. Temporal event construction based anomaly detection has been developed in the work of [13]. Reference [82] reports a framework for cyber-physical security. A system-level security design for power systems has been developed [83]. Cyber security technologies for anomaly detection at a substation are in an early stage of development. Technologies to detect anomalies for substation automation protocols are critically needed, such as GOOSE, SMV, and Manufacturing Message Specification (MMS).

Stuxnet, Duqu and Flame could be highly relevant cyber attacks (malwares) that are aimed at critical power infrastructure control systems [39]. Other cyber security concerns and potential threats to the power infrastructures have been reported by governments and other organizations, e.g., General Accounting Office (GAO), NIST or Interagency Reports National Institute of Standards and Technology Interagency Report (NISTIR) and Department of Energy (DOE) [84, 85, 86]. In addition, substation automation standards existed before cyber security became a major concern for power grid. As a result, full security measures have not been incorporated in the open standards [5]. Multicast distribution techniques that are used for GOOSE and SMV enable an efficient communication mechanism; however, it also causes cyber security issues and vulnerabilities, e.g., open group membership and open access [87]. Due to the fast transmission time requirement (within 4 ms), most encryption techniques or other security measures that increase transmission delays may not be practical for GOOSE and SMV. Although the work of [5] proposes an authentication method through a digital signature, the performance test is yet to be performed. Current intrusion detection or anomaly detection methods do not normally support substation automation protocols, e.g., GOOSE and SMV; they are more focused on cyber attacks through Distributed Denial of Service attack (DDoS), and website and operating system (OS). Cyber intrusions related to GOOSE and SMV can cause serious damages. Intruder(s) can modify GOOSE control messages and trip circuit breakers in a substation. They can also send fabricated (and improper) protection coordination messages to other substations. A SMV message attack can generate fabricated analog values to a control center, leading to undesirable operations.

This dissertation is concerned with anomaly detection at a substation. An integrated method for host-based and network-based anomaly detection schemes is proposed. The host-based anomaly detection uses a systematic extraction technique for intrusion footprints that can be used to identify

credible intrusion events within a substation, e.g., firewall, user-interface, IEDs, and circuit breakers. The network-based anomaly detection is focused on multicast messages in a substation network; it also detects, in a real-time environment, anomalies that demonstrate abnormal behaviors. The main contribution of this dissertation is a new method for (1) an integrated anomaly detection system for protection of IEC 61850 based substation automation system, e.g., IEDs, user-interface and firewall, and (2) a network-based anomaly detection algorithm that can be used to detect malicious activities of IEC 61850 based multicast protocols, e.g., GOOSE and SMV, across the substation network. Anomaly detection for multicast messages in a substation automation network is a new field of research for the power grids. In this research, a cyber security testbed has been developed and used to validate the proposed anomaly detection algorithms. Cyber intrusions are simulated using the testbed including protective IEDs. The test results demonstrate that proposed anomaly detection algorithms are effective for the detection of simulated attacks.

In the remaining of this dissertation, Section 4.2 describes cyber security vulnerabilities in a substation network. Section 4.3 includes algorithms for host- and network-based anomaly detection schemes. In Section 4.4, the network-based substation multicast messages are analyzed for anomaly detection. Section 4.5 provides the test results of the proposed anomaly detection system and the simultaneous intrusion detection at multiple substations. The conclusions and recommendations for future work are given in Section 4.6.

4.2 Cyber Security Vulnerability of a Substation

A power substation may consist of various types of equipment such as network devices, user-interface, server, global positioning system (GPS), firewall, IEDs, and remote access points. IEC 61850 based protocols are used by substation automation facilities, e.g., GOOSE, SMV and MMS.

GOOSE is used to send tripping signals from IEDs to circuit breakers. Sampled measured voltage and current values (SMV) are sent from a Merging Unit (MU) to an IED. Many devices are synchronized by GPS. MMS is used for monitoring, control and reporting between the user-interface and IEDs. Vulnerabilities of the substation network and mitigation of cyber attacks are critical subjects for anomaly detection. Remote access to a substation network from corporate offices or locations external to the substation is not uncommon for control and maintenance purposes. Dial-up, Virtual Private Network (VPN), and wireless are available mechanisms between remote access points and the substation Local Area Network (LAN) [10]. These access points are potential sources of cyber vulnerabilities. When remote access points have been compromised by an intruder, malicious attacks to operate circuit breakers and/or to access critical information, such as Substation Configuration Description (SCD), can be launched. IEDs may have a web server to allow remote configuration change and control. This dissertation assumes that the remote access point is the main intrusion point to the substations. An intruder may be able to access the substation network after the firewall is compromised. (S)he may capture, modify, and retransfer GOOSE packets and operate circuit breakers in a substation. The attacker may also send fabricated (and improper) GOOSE to other substations, causing unauthorized breaker operations. The consequence of a fabricated SV message attack can generate high current values to a control center and it may lead to an undesirable operation. After malicious activities or anomalies are detected in a substation network using the proposed integrated anomaly detection system, an intruder can be disconnected by collaboration between the IDS and firewall in the substation network. For a firewall, this can be achieved by dynamic rejection rules or disconnecting open ports. The proposed IADS uses anomaly and specification-based detection algorithms. Therefore, it is not able to detect unknown or intelligent attacks that are not defined in the algorithm. Periodic updates of the attack models will be important.

As illustrated in Fig. 4.1, possible intrusions to the substation communication network can originate from outside or inside a substation network, e.g.,

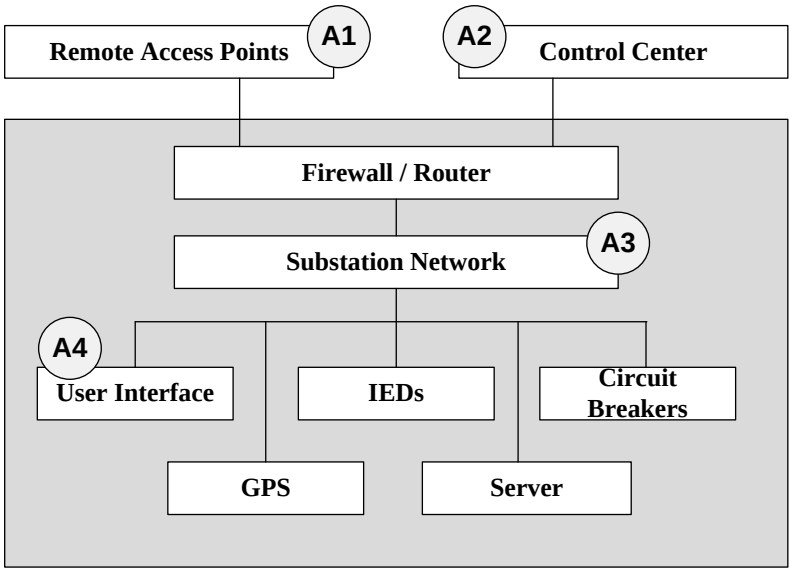

Fig. 4.1 Intrusion points in a substation automation system

- From outside a substation network: Intrusions can originate from remote access points (A1) or a control center (A2) to the firewall and the substation communication network (A3). Once an intruder can access the substation communication network, (s)he can access other facilities in the substation.

- From inside a substation network: Intrusions can enter from the substation communication network (A3) or user-interface (A4) and then gain access to other facilities in the substation.

Here are examples on how an intrusion from inside and outside of a substation can be launched on a substation network:

Inside attack: if a USB is already infected by an attacker, it may be used to install malware on the substation user-interface. Then it may be used to open a predefined communication port or execute hacking tools.

Outside attack: Remote access points may be used for maintenance, control or operation. Once an intruder compromises the access points, the attack may be able to pass the firewall and gain access to the substation ICT network.

Both inside and outside intrusions can be host-based or network-based attacks. A critical host-based attack is to compromise the user-interface machine. The user-interface system has the Human Machine Interface (HMI) and engineering tools that allow an operator or engineer to control, monitor or modify settings of the IEDs. A compromised user-interface can lead to undesirable operations of circuit breakers and settings for IEDs and transformer taps. Network-based intrusions can be conducted through packet monitoring, modification and replay attacks. Intruders can open circuit breakers by modifying GOOSE, SMV and MMS messages in a substation network. Modification of Simple Network Time Protocol (SNTP) messages can disrupt time synchronization. Each of attacks may cause severe damages.

4.3 Anomaly Detection

Anomaly detection refers to finding patterns that indicate abnormal or unusual behaviors. It is a method for detection of cyber security intrusions [73] that requires data analysis and correlation of events.

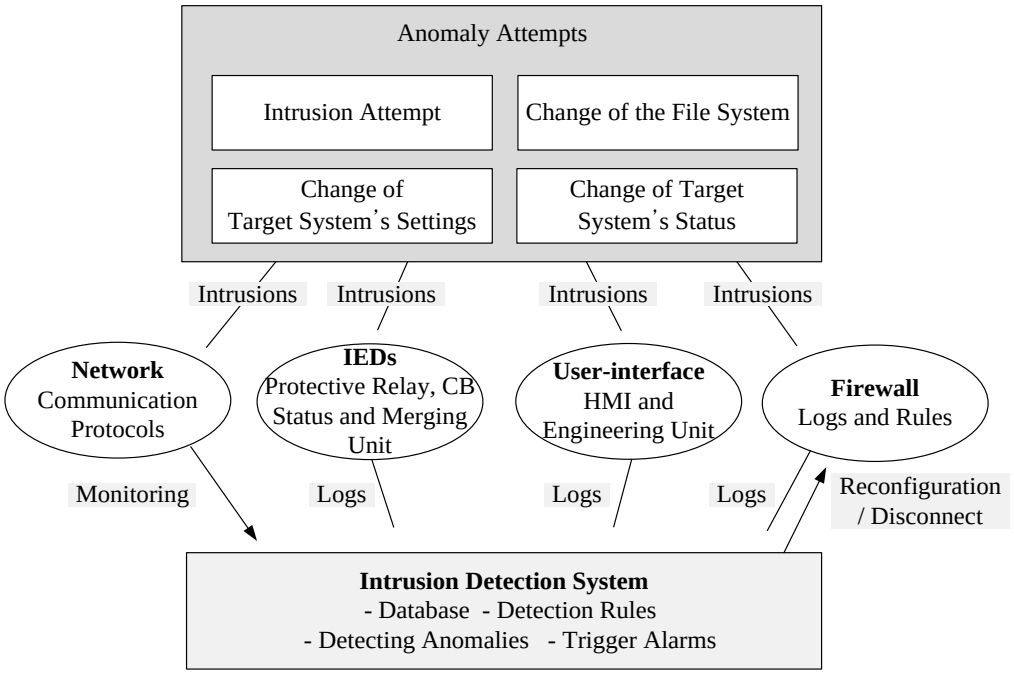

Fig. 4.2 Intrusion detection in a substation

As depicted in Fig. 4.2, intruders' behaviors generate logs across the substation-level networks, e.g., IEDs, firewall, user interface, and communication networks. For instance, the Stuxnet attack is based on: (1) intrusions, (2) changing the file system, (3) modifications of target system settings, and (4) altering the target system status [39]. If intruders try to compromise the substation targets, e.g., IEDs, networks, user interface and firewall, their behaviors will leave footprints in substation networks. Anomaly detection is performed based on logs of intruders' footprints.

4.3.1 Host-based Anomaly Detection

This section proposes a temporal anomaly detection method for host-based anomaly detection which is a generalization of the authors' previous work [13]. Generalizations from the authors' prior work are: (1) this dissertation proposes an integrated anomaly detection system, whereas [13] is concerned only with host-based anomaly detection in a substation, (2) this dissertation proposes a more efficient algorithm for attack similarity compared to the previous results, and (3) the generalized method incorporates a comprehensive set of substation logs and messages and extends the capability to scenarios involving multiple substations. The main assumption of the temporal anomaly detection for host-based anomaly detection is that the engineering software and hardware are able to generate system and security logs. For instance, if an intruder makes a wrong password attempt to IED or the user-interface, this action will generate a wrong password attempt flag. Similarly, if an intruder tries to copy or change a file in the user-interface, it will generate an unauthorized file change flag. The generalized method incorporates a comprehensive set of substation logs and messages and extends the capability to scenarios involving multiple substations. Temporal anomaly is used for host-based anomaly detection and can be determined from discrepancies between event logs from different time periods. As shown in Fig. 4.2, data logs at substation networks are used for the host-based anomaly detection algorithm.

The anomaly between two different time snapshots can be determined by a metric. The proposed technique is explained through an example. In Table 4.1, the event log matrix Ω, with a dimension of 7 by 4, contains 7 rows of anomaly indicators at the same substation for 7 consecutive time instants. Each of the 4 columns represents a specific type of host-based anomaly indicator, i.e., ψ^a (intrusion attempt on user interface or IED), ψ^{cf} (change of the file system), ψ^{cs} (change of IED

critical settings), and ψ^o (change of status of breakers/switches or transformer taps), respectively. If a specific type of anomaly is detected at time t, the value of the corresponding element in matrix $\mathbf{\Omega}$ will be changed from 0 (no anomaly) to 1 (anomaly). Detection of temporal anomalies is performed by comparing consecutive row vectors representing a sequence of time instants. The host-based ADS module imports the system and security logs from the user-interface, IEDs and firewalls at a predefined time. In this dissertation, the predefined polling time of system and security logs data is 10 seconds. An example of $\mathbf{\Omega}$ matrix describes a temporal anomaly detection for a sequence of 7 time instances of a substation A and B, and the time difference from t_1 to t_2 in $\mathbf{\Omega}$ matrix is 10 second. After subscribing to the logs, a data convertor module will change all temporal logs to binary values as shown in Table 4.1 (substation A) and Table 4.2. For example, Table 4.1 (substation A) has the converted binary values from Table 4.2. A detailed explanation is given in the following:

- At 10:20:000, there is no anomaly so t_1 is [0 0 0 0].
- At 10:30:000, ADS detects a wrong password attempt to IED 1 so t_2 is [1 0 0 0].
- At 10:40:000, ADS detects an unauthorized file change to the user-interface so t_3 is [1 1 0 0].
- At 10:50:000, there is no change so t_4 is [1 1 0 0].
- At 11:00:000, there is no change so t_5 is [1 1 0 0].
- At 11:10:000, ADS detects two anomalies, unauthorized setting change to IED 2 and unauthorized tap change to transformer 1 so t_6 is [1 1 1 1].
- At 11:20:000, there is no change so t_7 is [1 1 1 1].

An example of $\mathbf{\Omega}$ matrix describes a temporal anomaly detection for a sequence of 7 time instances of a substation A as shown in Table 4.1.

Table 4.1: An example of temporal anomaly detection in substations

Substation A	Substation B
$\boldsymbol{\Omega} = \begin{matrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$	$\boldsymbol{\Omega}' = \begin{matrix} t_1 \\ t_2 \\ t_3 \\ t_4 \\ t_5 \\ t_6 \\ t_7 \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}$

Table 4.2: System logs of a substation A

Substation A				
o.	**Date**	**Time**	**Contents**	**Issue**
45	15.09.2013	10:28:33,560	IED 1	Wrong password attempt
46	15.09.2013	10:35:43,159	User-interface	Unauthorized file change
47	15.09.2013	11:02:04,368	IED 2	Unauthorized setting change
48	15.09.2013	11:03:14,270	Transformer 1	Unauthorized tap change

An assumption of temporal anomaly detection for host-based anomaly detection is that the engineering software and hardware are able to generate system and security logs. For instance, if an intruder makes a wrong password attempt to IED or the user-interface, this action will generate a wrong password attempt flag. In the same manner, if an intruder tried to copy or change a file in the user-interface, it will generate an unauthorized file change flag. Some products have this log generating function but not all. If a specific type of anomaly is detected at time t, the value of the corresponding element in matrix Ω will be changed from 0 (no anomaly) to 1 (anomaly). The binary

number 1 (anomaly) will be kept until the operator resolves the issue and resets the integrated anomaly detection system. After resetting, all elements in matrix Ω will be set to zero (no anomaly). The main reason to use binary values for temporal anomaly detection is to minimize the calculation time for simultaneous anomaly detection at multiple substations.

If a discrepancy exists between two different periods (rows), the vulnerability index V_h^{Ω} is a number between 0 and 1. A value of 0 implies no discrepancy whereas 1 indicates the maximal discrepancy. A scalar index for temporal anomaly at time t=ti is defined as

$$V_{h(i)}^{\boldsymbol{\Omega}} = \frac{\sum_{j=1}^{n} |\boldsymbol{\Omega}_{(i,j)} - \boldsymbol{\Omega}_{(i+1,j)}|}{n}, \text{ i=1,..., 6,} \tag{4-1}$$

where n is the total number of anomaly indicators (n=4 for this example). Based on Eq. (4-1) one can obtain a vector for temporal anomaly that provides irregularities of events during the selected time period, $\mathrm{t} = \mathrm{t}_1, \ldots, \mathrm{t}_7$, from Ω matrix, i.e.,

$$V_h^{\boldsymbol{\Omega}} = (0.25,\ 0.25,\ 0,\ 0,\ 0.5,\ 0). \tag{4-2}$$

The first element of Eq. (4-2) is the value from the calculation based on first and second row of Ω in Table 4.1, similarly for other elements. The anomaly of this substation is determined by the vector V_h^{Ω}. If V_h^{Ω} is a zero vector, then there is no anomaly event on this substation. Otherwise, the substation will be included in the credible list to be evaluated further.

The proposed temporal anomaly detection is extended to detect simultaneous anomaly detection among multiple substations. The simultaneous anomaly detection is achieved in 3 steps, i.e., 1) Find the total number of types of attacks, 2) Find the same attack groups, and 3) Calculate the similarity between attacks in the same group. The total number of types of attack can be calculated by

$$\text{Total number of types of attack} = \sum_{k=1}^{n} \frac{n!}{k!(n-k)!} + 1, \tag{4-3}$$

where n is the total number of anomaly indicators. Eq. (4-3) is based on binomial coefficients. The total number of types of attacks for the specific example above is 15 since it has 4 anomaly indicators (number of columns). Let the event log matrix Ω' be an indicator for a different substation, as shown in Table 4.1. In comparison with Ω, it is assumed that the Ω' matrix has identical values except for the 5th row which is [1, 1, 0, 1]. Then the attack patterns of Ω and Ω' are considered to be the same since they eventually have the same values in the last row, i.e., [1, 1, 1, 1]. It indicates that substations Ω and Ω' are under a simultaneous attack but the attack sequences are different. Once the same type of attack groups is found as described above, the similarity between attacks can be calculated by

$$\text{Attack Similarity} = 1 - \frac{\sum_{i=1}^{x} \sum_{j=1}^{y} \left| \boldsymbol{\Omega}_{(i,j)} - \boldsymbol{\Omega}'_{(i,j)} \right|}{x \cdot y}, \tag{4-4}$$

where x and y are total number of rows and columns of matrix, respectively (x=7 and y=4 for this example). Attack similarity value of 0 indicates no overlap and a value 1 indicates a complete overlap. Therefore, by Eq. (4-4), the similarity index between substation Ω and Ω' is 0.9643.

4.3.2 Network-based Anomaly Detection

The proposed method also provides a network-based anomaly detection algorithm for multicast messages in the substation automation network. The multicast messages are based on IEC 61850 standard, e.g., GOOSE and SMV. The proposed Substation Multicast Message Anomaly Detection (SMMAD) model in Fig. 4.3 is divided into 3 process modules, i.e., packet filtering, anomaly detection, and evaluation. The packet filtering module consists of functions to identify GOOSE and SMV messages. The filter will only allow passing for GOOSE and SMV messages so the burden of processing can be reduced and the system performance will increase. The anomaly detection module is used to find violations based on predefined rules. The evaluation module will decide if the detected anomaly status is "abnormal" or "attack." Details will be explained in the next section.

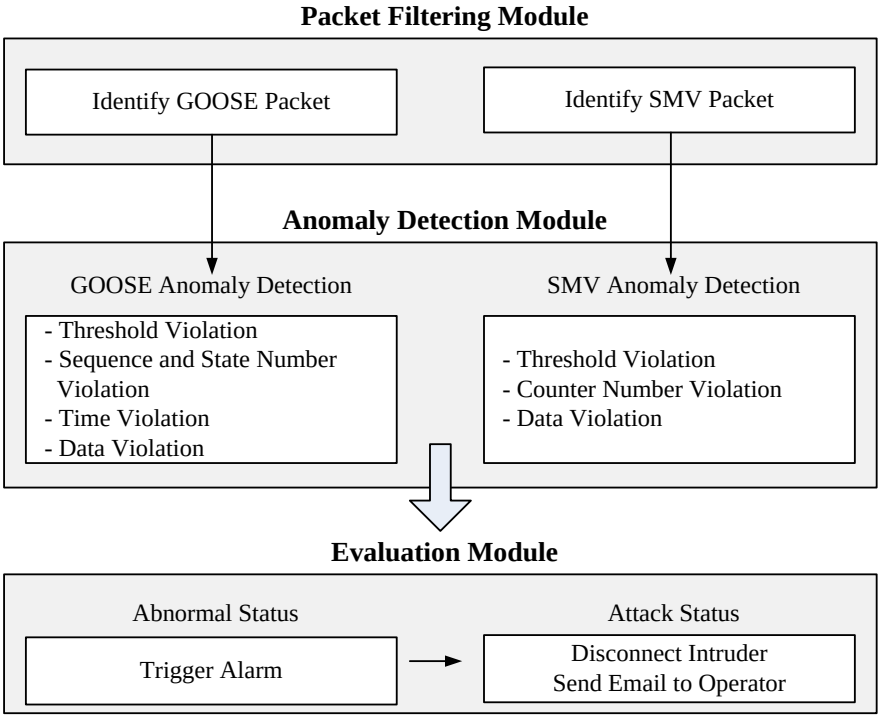

Fig. 4.3 SMMAD modeling for ADS

4.4 Substation Multicast Message Anomaly Detection

4.4.1 Multicast Messages in IEC 61850

Multicast messages in IEC 61850, e.g., GOOSE and SMV, are different from other protocols used in substation automation because they use three layers in Open Systems Interconnection (OSI) model stack, i.e., physical, data link, and application layer, in a real-time requirement. The multicast scheme uses the Media Access Control (MAC) address [88]. The GOOSE service uses a re-transmission scheme to enhance the communication speed and reliability, i.e., the same GOOSE message is re-transmitted at different time intervals but no response is sent from the receiver. The sequence number of GOOSE messages will be increased for each transmission and the state number will increase when the data status is changed. The sequence number will be set to 0 when the state number is changed. However, the specific time of re-transmission (interval) is not defined in the IEC 61850 standard so different vendors' GOOSE re-transmission times may vary [35].

SMV of voltage and current messages are published from the Merging Unit and subscribed by IEDs. The resolution amplitude of the Merging Unit in this project is 16 bits so it will send 960 SMV voltages and currents to IEDs in a second [34]. The message counter is incremented each time when a new sampled packet is published.

4.4.2 Detection Method

Unwanted multicast message packets can be identified by rules that match known signatures. Therefore, anomalies which match the predefined rules can be detected by the ADS. Each rule has been defined based on the IEC 61850 standard. First, an "anomaly" state has been identified as a result of violation of predefined rules. Second, an "attack" state is identified if the detected anomaly will adversely affect proper functioning of the substation control and measurement, e.g., open

circuit breaker and change of voltage and current values. A binary status is used as indicators of the status, i.e., “0” means no anomaly and “1” indicates that an anomaly is detected.

Port mirroring is a function to copy all packets from port(s) to the specific port in order to monitor and analyze packets. General network-based ADS will need port mirroring to capture all communication packets in the network [11]. Note that the proposed ADS is able to capture the GOOSE and SMV without the port mirroring function as it is focused on multicast messages and not other packets.

4.4.3 Main Framework

After calculating the violation detection indicators in GOOSE and SMV anomaly detection modules, the anomaly detection module will determine if there is an anomaly using the rules in Appendix I. As shown in Appendix 1, Line 7 is used for GOOSE anomaly detection, i.e., any detected anomaly in threshold violation $\alpha_{\mathrm{Th}}^{\mathrm{G}}$, sequence and state number violation $\beta_{\mathrm{S}}^{\mathrm{G}}$, GOOSE time violation $\gamma_{\mathrm{Ti}}^{\mathrm{G}}$, and GOOSE data violation $\delta_{\mathrm{d}}^{\mathrm{G}}$ will change GOOSE network-based anomaly indicator ψ^{G} from false to true. On the other hand, Line 12 is developed for SMV anomaly detection, i.e., any detected anomaly in SMV threshold violation $\varepsilon_{\mathrm{Th}}^{\mathrm{SV}}$, counter number violation $\theta_{\mathrm{cn}}^{\mathrm{SV}}$, and SMV data violation $\mu_{\mathrm{d}}^{\mathrm{SV}}$ will set the SMV network-based anomaly indicator ψ^{SV} from false to true.

After the anomaly detection task is completed, a network-based substation vulnerability index $\mathrm{V}_{\mathrm{n}}^{\mathrm{GS}}$ is defined as follows:

$$V_n^{GS} = \begin{cases} 1, \text{if } \psi^G = true \\ 1, \text{if } \psi^{SV} = true \\ 0, \text{ otherwise,} \end{cases} \quad (19)$$

where ψ^G is the GOOSE network-based anomaly indicator and ψ^{SV} is the SMV network-based anomaly indicator. A result of $V_n^{GS} = 1$ indicates the existence of an intrusion based on GOOSE and SMV messages whereas $V_n^{GS} = 0$ indicates that there is no evidence of a multicast message based cyber intrusion.

The proposed SMMAD examines all GOOSE and SMV packets in the substation network, and then checks if there is a security violation, as shown in Appendix I. SMMAD has two phases: initialization and detection. Line 1 represents the initialization of the examination process. Line 2 captures all packets in a substation network. Lines 3 and 10 are to check whether this is a GOOSE or SMV message. Line 4 and 11 are used to analyze the captured packets. Lines 5 and 6 create anomaly detection threads if there is more than one type of GOOSE messages. Lines 7 and 12 are used to check if there is a security violation. Finally, Lines 8, 9 and 13, 14 show whether there is an intrusion.

4.4.4 GOOSE Anomaly Detection

The threshold of GOOSE packets G_{th} can be calculated by the pre-defined re-transmission rule. The proposed ADS can filter the GOOSE packets by checking recommended MAC address from 01-0C-CD-01-00-00. Then the count of GOOSE packet is maintained, and details of this packet are saved. When the captured number of GOOSE packets G_{cnp} within predefined time G_{th}^T is greater than the predefined threshold for GOOSE packets G_{th} within G_{th}^T or there is no captured GOOSE packet

within G_{th}^{T}, an anomaly is deemed to be occurring and details are written to the log file. This process can also detect a GOOSE based denial-of-service (DoS) attack. Hence, the GOOSE violation indicator (GVI) α_{Th}^{G} is changed from 0 to 1. Line 1 in Appendix II is used for the detection of threshold violation α_{Th}^{G}.

The state number of GOOSE messages G_{st} will change and the sequence number of GOOSE G_{sq} will be set to 0 when the GOOSE state is changed. The sequence number of GOOSE will increase when GOOSE is published. Hence, if a captured GOOSE message's sequence number is not set to zero after the state is changed or sequence number is not matched as a sequence, it will detect the anomalies that are suspicious as attacker(s)'s packet modification or injection to the substation network. The GVI β_{S}^{G} will be changed from 0 to 1. Line 2 in Appendix II is for the GOOSE sequence and state number violation β_{S}^{G} detection.

In general, the GOOSE clients and servers are synchronized within a few microseconds for the critical protection and control functions. The time stamp will be implemented in the GOOSE packet by the sender. So anomaly will be detected when the generated time stamp G_{ge}^{T} is greater than the receiver's time G_{re}^{T}. The recommended GOOSE transfer time G_{tr}^{T} is defined in the IEC 62351-1 standard, which is 4 ms. If the difference between the generated time and received time is greater than the transfer time, it will be considered an anomaly. The GVI γ_{Ti}^{G} will be changed from 0 to 1. Line 3 in Appendix II is for the GOOSE time violation γ_{Ti}^{G} detection.

When the GOOSE indicator that contains the binary control value is changed from false to true or vice versa, the state number of GOOSE will be changed to the next number and the sequence

number will be set to 0. Therefore, if there is any violation of this rule, the GVI δ_d^G will be changed from 0 to 1. Line 4 in Appendix II is to perform the detection of GOOSE data violation δ_d^G.

4.4.5 Sampled Measured Values Message Anomaly Detection

The threshold for SMV packets S_{th} depends on the sampling rate. The proposed ADS will capture the SMV message by checking MAC address which starts from 01-0C-CD-04-00-00. Then it will count the number of SMV every second, and save the detailed information. When the captured number of SMV packets S_{cnp} within predefined time S_{th}^T is greater than the predefined threshold for SMV packets S_{th} within S_{th}^T or there is no captured SMV packet within S_{th}^T, an anomaly is deemed to be occurring and details are written to the log file. This can also detect a SMV based denial-of-service attack. A SMV violation indicator (SVI) ε_{Th}^{SV} will be changed from 0 to 1. Line 5 in Appendix II is used to perform the detection of SMV Threshold violation ε_{Th}^{SV}.

For the counter number violation detection, "SmpCnt" is a SMV protocol attribute and its attribute type is INT16U. This value will be incremented each time SMV is published. The count will be set to zero when sampling is synchronized by a clock signal [88]. The SMV message counter S_{mc} corresponds to SmpCnt so it will also increase after each transmission. If the SMV message counter is not increased or equal to the previous count when sampling is not synchronized, the SVI θ_{cn}^{SV} will be changed from 0 to 1. Line 6 in Appendix II is used to carry out the counter number violation θ_{cn}^{SV} detection.

Each group of SMV message has its own identification S_{id} and name of dataset S_{ds} [88]. They will not change unless the configuration of the Merging Unit is changed. Therefore the proposed

algorithm will detect the anomalies when there is a modification of the name of identification and dataset, and they still contain the same source and destination MAC address. Then, the SVI μ_d^{SV} will be changed from 0 to 1. Line 7 in Appendix II is to detect the SMV data violation μ_d^{SV}.

4.5 Simulation Results

A testbed is developed at WSU to perform different types of cyber intrusions and analyze the effectiveness of the proposed detection and mitigation techniques in a realistic substation environment. Government agencies and other organizations have been using various testbeds for cyber security testing [89, 90, 91]. In this dissertation, several types of cyber attacks have been generated for validation of the proposed anomaly detection algorithms, e.g., replay, packet modification, injection, generation and DoS using the testbed. The results of the simulated attacks are shown in Tables 4.3 and 4.4 and Fig. 4.5.

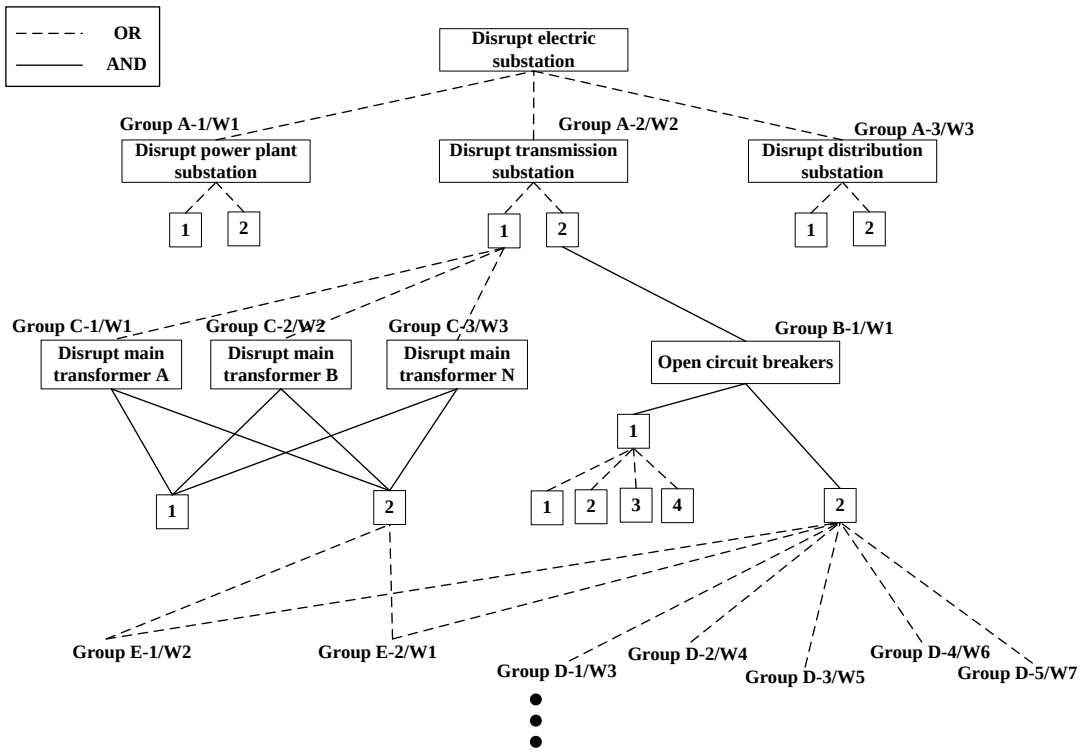

Fig. 4.4 Attack tree for the substations

Fig. 4.4 shows a portion of the attack tree for the substations that have been used in Case study I, II and III. For instance, the goal of group B-1/W1 is to open a circuit breaker. The preconditions of this attack are: an attacker can open a circuit breaker via IED, the control center user interface, and substation user interface. The goal of this attack is achieved with two AND conditions, i.e., 1) find target CB, and 2) open target CB, as shown in Fig. 4.4. Find target has four OR conditions: 1) send a GOOSE message to CB using IED, 2) use control center user interface, 3) use substation user interface, and 4) modify the protection setting (to low value) of IED. The post condition of this attack is that the attacker will open target CB. It is shown that some intrusions are able to execute a switching action on the circuit breaker. The C language based source code library was used for the proposed integrated ADS. The proposed anomaly detection algorithms are implemented in the C

language. C++ has been used for ADS HMI in order to test the real-time anomaly detection and alarms to the substation operator. The circuit breaker is designed to subscribe GOOSE messages generated from the IEDs. IED A is designed to subscribe to SMV messages that are from the Merging Unit. Free available software tools are used for all intrusion processes, e.g., Wireshark, Colasoft Packet Builder, Nmap, etc.

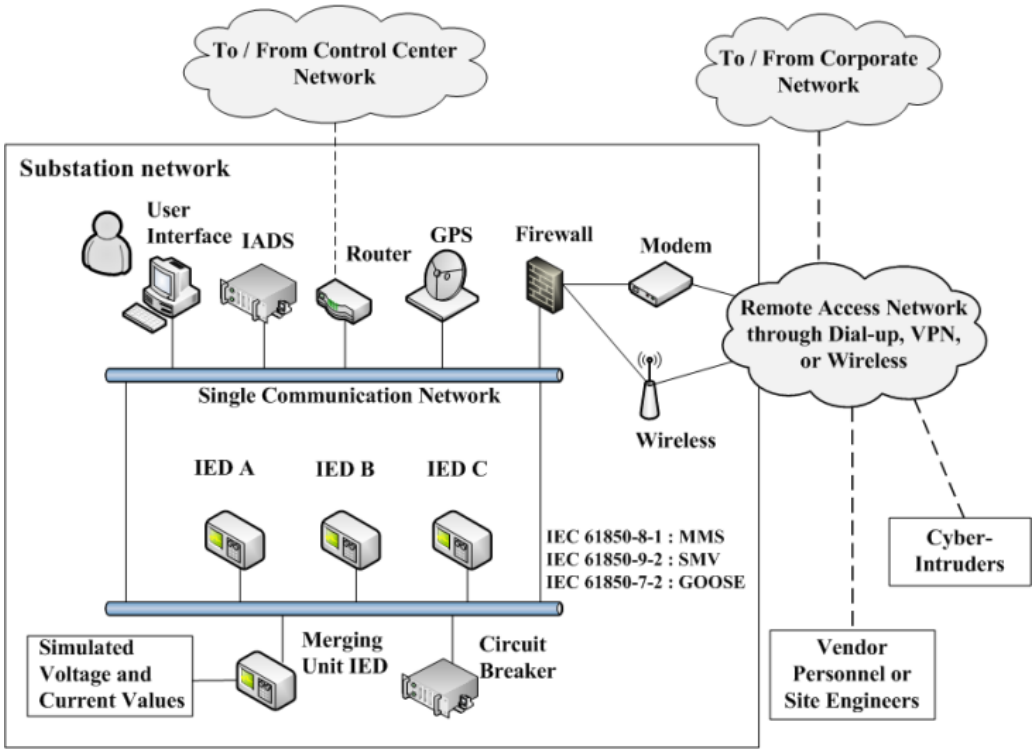

Fig. 4.5 WSU cyber security testbed for the substation

Table 4.3: Consequence of GOOSE based malicious behaviours without anomaly detection system

Action	Result
Disconnect Ethernet cable from IED	Lost availability of IED
Send normal control	Open CB
Replay attack	Open CB
Modify sequence & state number	Warning occurred at CB
Modify transferred time	Warning occurred at CB
Modify GOOSE control data	Open CB
Denial of Service attack	Lost availability of CB
Generate GOOSE control data	Open CB

Table 4.4: Consequence of SMV based malicious behaviours without anomaly detection system

Action	Result
Disconnect Ethernet cable from MU	Lost availability of MU
Increase measured values	Open CB
Replay attack	Open CB
Modify counter number	Warning occurred at IED
Modify SMV dataset	Warning occurred at IED
Denial of Service attack	Lost availability of IED
Generate SMV data	Open CB

The simulation results include 3 Study Cases. Case I shows the GOOSE cyber intrusions and detection on the substation communication network. Both single and simultaneous attacks are considered. The results demonstrate that the proposed method detects all intrusions and triggers the appropriate alarms. Case II is a simulation of SMV intrusions and detection. SMV packets are captured and retransferred to the substation network after they are falsified to include high current

and voltage values. The results showed that the proposed anomaly detection can detect simulated intrusions and then trigger the alarms. Case III is concerned with simultaneous anomaly detection at multiple substations. Cyber intrusions are generated by attacker(s) and detected by the ADS up to 2000 substations. The results show that proposed algorithm is faster than others.

4.5.1 Case Study I: GOOSE Anomaly Detection

As shown in Table 4.5, the threshold of GOOSE messages G_{th} has been set to 12, including a margin of error of 20%, since the peak number of GOOSE messages when normal control was issued is 10. The GOOSE anomaly detection results are given in Table 4.5.

Table 4.5: GOOSE anomaly detection test results

Test case	Set packet threshold G_{th} (per 1 sec)	Normal control issued	Disconnect Ethernet cable from IED	Detected anomalies	Alert issued
T1	12	No	No	-	No
T2	12	Yes	No	-	No
T3	12	No	No	$\alpha_{Th}^{G}, \beta_{S}^{G}, \gamma_{Ti}^{G}$	Yes
T4	12	No	No	β_{S}^{G}	Yes
T5	12	No	No	γ_{Ti}^{G}	Yes
T6	12	No	No	δ_{d}^{G}	Yes
T7	12	No	No	$\alpha_{Th}^{G}, \beta_{S}^{G}, \gamma_{Ti}^{G}$	Yes
T8	12	No	No	δ_{d}^{G}	Yes
T9	12	No	Yes	α_{Th}^{G}	Yes
T10	12	No	No	$\alpha_{Th}^{G}, \beta_{S}^{G}, \gamma_{Ti}^{G}, \delta_{d}^{G}$	Yes
- The peak number of normal GOOSE message when control was issued: 10 (per second) - Number of normal GOOSE message: 1 (per second)					

- T1, normal status: There was no alarm under a normal operating condition.

- T2, normal control issued: There was no alarm when normal control was issued to IED.

- T3, replay attack (20 packets/sec): The normal control GOOSE packet was captured from T2 and retransferred to the substation network by the attacker without any modification.

- T4, sequence and state number modification attack (5 packets/sec): Change sequence and state number of GOOSE packets and then transfer to substation network by the attacker.

- T5, transferred time modification attack (5 packets/sec): Change time stamp of GOOSE packets and then transfer to substation network by the attacker.

- T6, GOOSE control data modification attack (5 packets/sec): Change control data of GOOSE packets and then transfer to the substation network by the attacker.

- T7, Denial of Service attack (2000 packets/sec): Execute GOOSE based DoS attack by the attacker.

- T8, generating GOOSE control data attack (5 packets/sec): Generate GOOSE control messages and publish to the substation network by the attacker.

- T9, disconnect Ethernet cable: Disconnect Ethernet cable from IED by the attacker so there was no GOOSE message in the substation network.

- T10, simultaneous attack: Change sequence, state number, time stamp and control data of GOOSE packets and then transfer to substation network by the attacker.

4.5.2 Case Study II: SMV Anomaly Detection

Table 4.6 shows that the threshold of SMV messages S_{th} has been set to 1178, including a margin of error of 20%, since the peak number of SMV messages when normal control was issued is 982.

Table 4.6: SMV anomaly detection test results

Test case	Set packet threshold S_{th} (per 1 sec)	Disconnect Ethernet cable from MU	Detected anomalies	Alert issued
T11	1178	No	-	No
T12	1178	No	ε_{Th}^{SV}, θ_{cn}^{SV}	Yes
T13	1178	No	θ_{cn}^{SV}	Yes
T14	1178	No	μ_{d}^{SV}	Yes
T15	1178	No	ε_{Th}^{SV}, θ_{cn}^{SV}	Yes
T16	1178	No	ε_{Th}^{SV}, θ_{cn}^{SV}	Yes
T17	1178	Yes	ε_{Th}^{SV}	Yes
T18	1178	No	ε_{Th}^{SV}, θ_{cn}^{SV}, μ_{d}^{SV}	Yes
- The peak number of SMV message: 982 (per second)				

- T11, normal status: There was no alarm under a normal operating condition.

- T12, replay attack (200 packets/sec): The normal SMV packet was captured and retransferred to the substation network without modification by the attacker.

- T13, counter number modification attack (20 packets/sec): Change the counter number of SMV packets and then transfer to substation network by the attacker.

- T14, SMV dataset modification attack (20 packets/sec): Change the dataset of SMV packets and then transfer to the substation network by the attacker.

- T15, Denial of Service attack (2000 packets/sec): Execute SMV based DoS attack by the attacker.

- T16, generating SMV data attack (100 packets/sec): Generate SMV messages that contain high current and voltage values, and publish to the substation network by the attacker.

- T17, disconnect Ethernet cable: Disconnect Ethernet cable from MU by the attacker so there was no SMV message in the substation network.

- T18, simultaneous attack: Change the counter number and dataset of SMV packets and then transfer to the substation network by the attacker.

Once ADS detects an anomaly in a substation network, it will trigger an alarm and send a message to operators. Also ADS will send a disconnect control command to the firewall and block the intruder's connection as a mitigation action.

4.5.3 Case Study III: Multiple Substation

An anomaly detection system is intended to find malicious behaviors quickly so that system operators can disconnect the intruder(s) from the network and take other mitigation actions. If there are simultaneous intrusions from multiple attackers, however, it is difficult to mitigate the situation since different types of intrusions will require corresponding countermeasures. The ability to find the same type of attacks and their locations will reduce the mitigation time and effort. The total number of types of attack is 57 since the proposed ADS has 6 anomaly indicators (4 of host-based anomaly indicators from Section III-A and 2 of network-based anomaly indicators from Eq. (4-5)) as shown in Tables 4.7 and 4.8. Tables 4.7 and 4.8 also report sample ADS logs of substations 1 and 2, respectively, where 0 indicates no anomaly and 1 indicates a detected anomaly. Table 4.8 includes logs indicating an intrusion into substation 1, leading to a change of settings and GOOSE attack. This attack is shown to start from intrusion attempts ψ^{a} at t_2. Then logs indicate an unauthorized change of settings ψ^{cs} for a protective device at t_3. This type of attacks may happen when attackers know the password for the IED configuration tool. The intruder also attempts the GOOSE based attack at t_4. Table 4.8 provides logs from the ADS in substation 2. It shows the same attack as the one at substation 1 since the attack pattern of substation 1 is [1, 0, 1, 0, 1, 0] and

substation 2 also has [1, 0, 1, 0, 1, 0] at t_7 but the attack time is different. Therefore, by Eq. (4-4), the attack similarity index between substations 1 and 2 is 0.9048.

Table 4.7: Detected anomaly log substation 1

Time	Host-based				Network-based	
	ψ^a	ψ^{fs}	ψ^{cs}	ψ^o	ψ^G	ψ^{SV}
t_1	0	0	0	0	0	0
t_2	1	0	0	0	0	0
t_3	1	0	1	0	0	0
t_4	1	0	1	0	1	0
t_5	1	0	1	0	1	0
t_6	1	0	1	0	1	0
t_7	1	0	1	0	1	0

Table 4.8: Detected anomaly log substation 2

Time	Host-based				Network-based	
	ψ^a	ψ^{fs}	ψ^{cs}	ψ^o	ψ^G	ψ^{SV}
t_1	0	0	0	0	0	0
t_2	0	0	0	0	0	0
t_3	0	0	0	0	0	0
t_4	1	0	0	0	1	0
t_5	1	0	1	0	1	0
t_6	1	0	1	0	1	0
t_7	1	0	1	0	1	0

The simulation steps are explained as follows. First, different types of attacks are randomly generated from multiple attackers. Second, all anomalies are captured and detected by the proposed ADS, and then the ADS generates logs at each substation. Third, simultaneous intrusion detection has been performed using generated logs. The proposed methodology for simultaneous anomaly detection at multiple substations is validated using the simulated data shown in Fig. 4.6. The proposed simultaneous anomaly detection method is able to identify the same type of attacks and its

similarity within 0.18 seconds among 2000 substations. It also shows that the computational performance of the proposed host-based anomaly detection algorithm is faster than the previous algorithm developed by the authors that uses Pearson's Similarity and the other similarity coefficient algorithms [92].

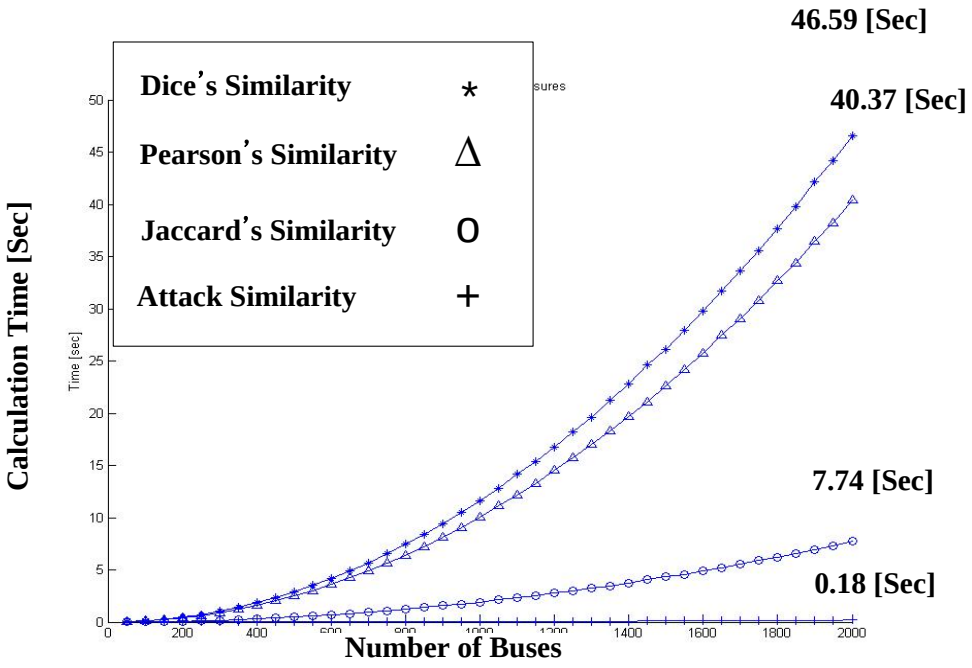

Fig. 4.6 Comparison of similarity coefficient algorithms

4.5.4 ADS Evaluation

The false positive ratio (FPR) is defined as the number of misclassified normal packets divided by the total number of normal packets. The false negative ratio (FNR) is defined as the number of misclassified abnormal packets divided by the total number of abnormal packets. The FPR and FNR of the proposed host-based anomaly detection system depend on the accuracy of the event log matrix Ω generated from the substation logs. They are 0.00013 and 0.0002, respectively. FPR and

FNR of the proposed network-based anomaly detection system depend on the number of packets per second. This is due to the fact that ADS may lose packets when the number of packets exceeds 2000 per second. FPR and FNR are 0.00013 and 0.00016 for the case of 2100 packets, respectively. In order to compare the performance of the proposed network-based anomaly detection, a rule-based detection system using Tshark is used [93]. TShark is a network protocol analyzer. It allows users to capture packet data from a live network, or read packets from a previously saved capture file, either printing in a decoded form to the standard output or writing the packets to a file [94]. The resulting FPR and FNR of the rule-based detection system is 0.00142 and 0.0019, respectively. Therefore the proposed network-based anomaly detection shows a higher performance.

4.6 Appendix I

The GOOSE and SMV messages have its own recommended MAC address as defined in IEC 61850-8-1 standard. The first three octets are assigned by IEEE with 01-0C-CD. Then fourth octet shall be 01 for GOOSE and 04 for multicast sampled values. The last two octets shall be used as individual addresses assigned by the range defined in Table 4.9.

Therefore, the proposed ADS filters the GOOSE and SMV packets by checking the recommended MAC addresses, 01-0C-CD-01-00-00 and 01-0C-CD-04-00-00, respectively. The ADS can create anomaly detection threads if there is more than one type of GOOSE messages by checking the MAC address. For instance, the first GOOSE MAC address is 01-0C-CD-01-00-01 and, if there is another GOOSE packet that has a MAC address 01-0C-CD-01-00-02, ADS will create a new anomaly detection thread. The proposed ADS can handle up to two different types of GOOSE messages. If a captured packet is a GOOSE, ADS will analyze the captured packets. Then ADS detects malicious activities and abnormal behaviors that match predefined security rules described

in Section IV-D. Finally ADS shows to the operator whether there is a GOOSE related intrusion or an anomaly. The creation of the new detecting thread is not applicable for SMV detection at this moment since ADS cannot handle too much data. The resolution (bits) amplitude of SV for protection and control is defined in IEC 61850-5, e.g., 8 bits (P1 class), 16 bits (P2 class) and 32 bits (P3 class). For example, SMV used in this research publishes approximately 960 packets in a second (using 16 bits). In the same manner, if a captured packet is a SMV, ADS will analyze the captured packets. Then ADS will detect malicious activities and abnormal behaviors that match predefined security rules described in Section 4.4. Finally, ADS provides an indication to the operator whether there is a SMV related intrusion or an anomaly.

SMMAD Algorithm

1. $\psi^G,\ \psi^{SV},\ V_n^{GS} = 0$; // Initialize
2. capture C_{pkt}; // Capture all packets in the substation network
3. if (C_{pkt} is IEC GOOSE);
4. $G_{cp} = [G_{st}, G_{sq}, G_{ge}^T, G_{re}^T]$; // Parse packet
5. if ($G_{cp}[G_{sm},\ G_{dm}] \neq G_{cp-1}[G_{sm},\ G_{dm}]$); // Find different GOOSE
6. make $G_{at\ new}$; // Create new anomaly detection thread
7. $\alpha_{Th}^G \vee \beta_S^G \vee \gamma_{Ti}^G \vee \delta_d^G \rightarrow \psi^G$; // Calculate GOOSE intrusion
8. if ($\psi^G = true$), set $V_n^{GS} = 1$; // Detect GOOSE intrusion
9. else set $V_n^{GS} = 0$; // No intrusion
10. elseif (C_{pkt} is IEC SMV);
11. $S_{cp} = [S_{mc}, S_{ds}, S_{id}, S_{sm}, S_{dm}]$; // Parse packet
12. $\varepsilon_{Th}^{SV} \vee \theta_{cn}^{SV} \vee \mu_d^{SV} \rightarrow \psi^{SV}$ // Calculate SMV intrusion
13. if ($\psi^{SV} = true$), set $V_n^{GS} = 1$; // Detect SMV intrusion
14. else set $V_n^{GS} = 0$; // No intrusion
15. return V_n^{GS};

Table 4.9: Recommended address range assignments

Service	Starting address (hexadecimal)	Ending address (hexadecimal)
GOOSE	01-0C-CD-01-00-00	01-0C-CD-01-01-FF
SMV	01-0C-CD-04-00-00	01-0C-CD-04-01-FF

4.7 Appendix II

Examples are provided on how the proposed host- and network-based anomaly detection system can find the GOOSE and SMV related anomalies and intrusions.

Example I: An intruder gains access to the substation network via VPN. (S)he scans all IP address and opens ports using a scanning tool. After the information of protection IED is found, (s)he captures GOOSE packets of the target IED. Then the intruder modifies control data of GOOSE messages and retransfers to the substation network. Now ADS will detect the modified GOOSE message since the intruder fails to synchronize the sequence number, state number, and time stamp of GOOSE.

Example II: An intruder gains access to the substation network via a dial-up connection. (S)he has a communication topology diagram and information. Intruder checks whether MU is live. After the information of the merging unit is found, (s)he captures SMV packets of the target merging unit. Then the intruder modifies the measured current values of the SMV message and retransfers to the substation network. Now ADS will detect the modified SMV messages since the counter number of injected SMV messages is not synchronized with the original SMV messages.

GOOSE and SMV Violation Indicators

1. α_{Th}^{G}: $[(G_{max}$ within $G_{th}^{T} > G_{cnp}$ within $G_{th}^{T})\ \vee\ (G_{cnp} = 0$ within $G_{th}^{T})]$.
2. β_{S}^{G}: $[(G_{sq} \geq G_{sq-1}) \wedge (G_{st} \leq G_{st-1})] \vee \left[(G_{sq} \leq G_{sq-1}) \wedge (G_{st} \geq G_{st-1})\right]$.
3. γ_{Ti}^{G}: $(G_{ge}^{T} \geq G_{re}^{T}) \vee [(G_{re}^{T} - G_{ge}^{T}) > G_{tr}^{T}\ ]$.
4. δ_{d}^{G}: $(G_{cp} \neq G_{cp-1}) \wedge \left[(G_{sq} \leq G_{sq-1}) \wedge (G_{st} \leq G_{st-1})\right]$.
5. ε_{Th}^{SV}: $(S_{max}$ within $S_{th}^{T} > S_{cnp}$ within $S_{th}^{T}) \vee (S_{cnp} = 0$ within $S_{th}^{T})$.
6. θ_{cn}^{SV}: $(S_{mc} \leq S_{mc-1})$ when $(S_{si} = S_{si-1} = False)$.
7. μ_{d}^{SV}: $[(S_{sm} = S_{sm-1}) \vee (S_{dm} = S_{dm-1})] \wedge [(S_{ds} \neq S_{ds-1}) \vee (S_{id} \neq S_{id-1})]$.

Table 4.10: An example of normal GOOSE operation and anomaly in a substation

	Normal operation			Anomaly		
Time	State number	Sequence number	Data	State number	Sequence number	Data
1	3	145	False	3	145	False
2	3	146	False	3	146	False
3	4	0	True	3	146	True
4	4	1	True	3	146	True
5	4	2	True	3	146	True

Example III: The left column of Table 4.10 shows a normal operation whereas the right column shows a GOOSE modification attack. When there is an open circuit breaker control event between time 2 and time 3, the state number is changed from 3 to 4 and the sequence number is set to 0. Then the sequence number is increased from 0 to 1, 1 to 2, etc. However, if an intruder captures, modifies data and retransfers GOOSE messages to the substation network, the state number and sequence number are not changed even though GOOSE data have changed. ”

Example IV: Suppose that there is a SMV packet insertion to the substation network using captured SMV packets. This action will trigger the SMV threshold violation $\varepsilon_{\mathrm{Th}}^{\mathrm{SV}}$ if the total numbers of SMV

packets (inserted packets + normal SMV packet) are higher than the SMV threshold. This will trigger the counter number violation θ_{cn}^{SV} since the inserted SMV packets will violate "SmpCnt" as explained in Section IV-E. This may also trigger the data violation μ_{d}^{SV} if the intruder inserts packets after modification of the SMV messages. It will show an alarm to the operator, who can find more details from the alarm logs and event logs.

4.8 Appendix III (Nomenclature)

α_{Th}^{G}	GOOSE threshold violation indicator
β_{S}^{G}	GOOSE sequence and state number violation indicator
γ_{Ti}^{G}	GOOSE time violation indicator
δ_{d}^{G}	GOOSE data violation indicator
ε_{Th}^{SV}	SMV threshold violation indicator
θ_{cn}^{SV}	SMV counter number violation indicator
μ_{d}^{SV}	SMV data violation indicator
ψ^{a}	Intrusion attempts upon user-interface or IEDs host-based anomaly indicator (HAI)
ψ^{cf}	Change of the file system HAI
ψ^{cs}	Change of IED critical settings HAI
ψ^{o}	Change of status on switches or transformer taps HAI
ψ^{G}	GOOSE network-based anomaly indicator (NAI)
ψ^{SV}	SMV network-based anomaly indicator
T	Predefined time for each anomaly detection indicator
C_{pkt}	Captured packets in a substation network
$V_{h}^{\mathbf{\Omega}}$	Substation vulnerability index for host-based anomaly
V_{n}^{GS}	Substation vulnerability index for network-based anomaly
G_{sm}	GOOSE source MAC address
G_{dm}	GOOSE destination MAC address
G_{at}	Anomaly detection thread for GOOSE
G_{cnp}	Captured number of GOOSE packets
G_{st}	State number of GOOSE packets
G_{sq}	Sequence number of GOOSE packets
G_{th}	Predefined threshold for GOOSE packets (depending on the re-transmission time)
G_{th}^{T}	Predefined time for GOOSE threshold violation detection

G_{ge}^{T}	GOOSE packet, time at which it is generated
G_{re}^{T}	GOOSE packet, time at which it is received
G_{tr}^{T}	GOOSE transfer time (4 ms, defined in IEC 62351-1 [5])
G_{cp}	Data of captured GOOSE packet
S_{th}	Predefined threshold for Sampled Values packets (depending on the sampling rate)
S_{cnp}	Captured number of Sampled Values packets
S_{cp}	Captured SMV packet
S_{mc}	SMV message counter
S_{ds}	Object reference of the data set (datSet)
S_{id}	Value of attributes MsvID of the MSVCB (smvID) [88]
S_{sm}	SMV source MAC address
S_{dm}	SMV destination MAC address
S_{si}	SMV synchronization indicator (*true* = synchronized by a clock signal, *false* = not synchronized)
S_{th}^{T}	Predefined time for SMV threshold violation detection

Chapter 5. Conclusions and Future Work

5.1 Conclusions

The proposed cyber-physical security framework is intended to improve the cyber security of existing substation computer networks. The equipment and software deployed at the substations have been equipped with communication technologies. Therefore, the requirements for identifying relevant properties of cyber security and performance are crucial. The contribution of this dissertation is a new substation anomaly detection algorithm that can be used to systematically extract malicious "footprints" of intrusion-based steps across substation networks. The proposed integrated anomaly detection system contains host- and network-based anomaly detection for a single substation, and simultaneous anomaly detection for multiple substations. The host-based ADS uses logs that are extracted from malicious footprints of intrusion-based steps across substation facilities. The network-based ADS can detect malicious behaviors that are related to multicast messages in the substation network. The proposed simultaneous intrusion detection method is able to find the same type of attacks on multiple substations and their locations, whereas the impact factor is used to evaluate how substation outages impact the entire system. The methods have been validated by testing with realistic intrusion scenarios using the testbed, e.g., replay, modification, man-in-the-middle, generation, and DoS.

5.2 Future Work

In order to increase the resiliency of power grids against cyber attacks, the following aspects should be investigated further:

1. In order to enhance the detection rate, substation systems need to generate more system and security logs since the proposed host-based anomaly detection depends on the generated logs. The network-based anomaly detection algorithm should be updated periodically since it is not able to detect unknown attacks that are not defined in the algorithm. In the future work, it will be useful to include other substation automation communication protocols, e.g., MMS, SNTP, DNP, Modbus, and IEC 60870-5 based anomalies.

2. Cyber-physical vulnerability assessment analysis that includes all substations should be proposed. A cyber-physical vulnerability index of each substation should be different since each substation has a different type of ICT devices, security feature, and impact factor on the power grid (i.e., a high voltage substation is normally more important than a low voltage substation). After calculating the cyber-physical vulnerability index for all substations, a power system will be able to identify the substations where cyber security needs to be enhanced first.

3. A coordinated simultaneous cyber attack detection algorithm using both ADS data and power system measurements need to be developed. In this research, two applications (e.g., impact evaluation and attack similarity), which use the ADS data, are proposed. However, the problem of these applications is that the accuracy of these applications is highly dependent on the false ratio of ADS data. In the same way, power system measurements highly rely on the ICT network. In order to make up for these weaknesses, a collaborative anomaly detection algorithm that uses both the physical system (power system measurements) and cyber system (ADS) data has to be developed.

Bibliography

[1] CRO Forum, Power Blackout Risks: Risk Management Options, Emerging Risk Initiative - Position Paper, Nov. 2011 [Online]. Available: https://www.allianz.com/v_1339677769000/media/responsibility/documents/position_paper_power_blackout_risks.pdf

[2] M. Kezunovic, "Smart fault location for smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 1, pp. 11-22, Mar. 2011.

[3] J.-W. Wang and L.-L. Rong, "Cascade-based attack vulnerability on the US power grid," *Safety Science*, vol. 47, no. 10, pp. 1332-1336, Dec. 2009.

[4] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-physical system security for the electric power grid," *Proc. IEEE*, vol. 100, no. 1, pp. 210-224, Jan. 2012.

[5] Power Systems Management and Associated Information Exchange - Data and Communications Security, IEC TS 62351-1 Standard: Part 1: Communication Network and System Security - Introduction to Security Issues, May 2007, 1st Edition.

[6] J. McGhee, and M. Goraj, "Smart high voltage substation based on IEC 61850 process bus and IEEE 1588 time synchronization," *IEEE Smart Grid Communications (SmartGridComm)*, pp. 489-494, Oct. 2010.

[7] Z. Vale and A. Machado e Moura, “An expert system with temporal reasoning for alarm processing in power system control centers,” *IEEE Trans. Power Syst.*, vol. 8, no. 3, pp. 1307-1314, Aug. 1993.

[8] D. Kirschen and B. Wollenberg, “Intelligent alarm processing in power systems,” *Proc. IEEE*, vol. 80, no. 5, pp. 663-672, May 1992.

[9] N. Liu, J. Zhang, H. Zhang, and W. Liu, “Security assessment for communication networks of power control systems using attack graph and MCDM,” *IEEE Trans. Power Del.*, vol. 25, no. 3, pp. 1492-1500, Jun. 2010.

[10] C.-C. Liu, A. Stefanov, J. Hong, and P. Panciatici, “Intruders in the grid,” *IEEE Power and Energy Magazine*, vol. 10, no. 1, pp. 58-66, Feb. 2012.

[11] U.-K. Premaratne, J. Samarabandu, T.-S. Sidhu, R. Beresh, and J.-C. Tan, “An intrusion detection system for IEC 61850 automated substations,” *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2376-2383, Oct. 2010.

[12] T. Morris and K. Pavurapu, “A retrofit network transaction data logger and intrusion detection system for transmission and distribution substations,” *IEEE International Conference on Power and Energy (PECon)*, pp. 958-963, Nov. 2010.

[13] C.-W. Ten, J. Hong, and C.-C. Liu, “Anomaly detection for cybersecurity of the substations,” *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 865-873, Dec. 2011.

[14] North American Electric Reliability Corporation (NERC) Standards, Cyber Security – Critical Cyber Asset Identification, Critical Infrastructure Protection (CIP) 002 – 009, Dec. 2009. [Online]. Available: http://www.nerc.com/pa/Stand/Pages/CIPStandards.aspx

[15] Logical Security Architecture Key Concepts and Assumptions on Intrusion Detection for Power Equipment – The Smart Grid Interoperability Panel – Cybersecurity Working Group, *Guidelines for Smart Grid Cybersecurity: Vol. 1, Smart Grid Cyber Security Strategy, Architecture, and High-Level Requirements*. NIST Interagency Report 7628, National Institute of Standards and Technology, US Department of Commerce, Aug. 2010.

[16] The Smart Grid Interoperability Panel - Cybersecurity Working Group, *Guidelines for Smart Grid Cybersecurity: Vol. 3, Supportive Analyses and References*. NIST Interagency Report 7628, National Institute of Standards and Technology, US Department of Commerce, Aug. 2010.

[17] "National SCADA test bed: Fact sheet," Idaho National Laboratory (INL), 2007.

[18] Common Cyber Security Vulnerabilities Observed in Control System Assessments by the INL NSTB Program, Idaho National Laboratory (INL), Nov. 2008.

[19] M.-R.-P. Rohde, "Cyber assessment methods for SCADA security," *Instrumentation, Systems and Automation Society (ISA)*, Tech. Rep., 2005.

[20] M.-J. McDonald, G.-N. Conrad, T.-C. Service, and R. H. Cassidy, "Cyber effects analysis using VCSE," Promoting Control System Reliability, Sandia National Laboratories, SAND2008-5954, Sep. 2008.

[21] M.-J. McDonald, "Modeling and simulation for cyber-physical system security research," Development and Applications, Sandia National Laboratories, SAND2010-0568, Feb. 2010.

[22] A. Hahn, A. Ashok, S. Sridhar, and M. Govindarasu, "Cyber-physical security testbeds: architecture, application, and evaluation for smart grid," *IEEE Trans. Smart Grid*, vol. 4, no. 2, pp. 847-855, June 2013.

[23] D. C. Bergman, D. Jin, D. M. Nicol, and T. Yardley, "The virtual power system testbed and inter-testbed integration," *Proc. 2nd Workshop Cyber Security Exp. Test*, Aug. 2009.

[24] M. Mallouhi, Y. Al-Nashif, D. Cox, T. Chadaga, and S. Hariri, "A testbed for analyzing security of SCADA control systems (TASSCS)," *Proc. IEEE PES Innov. SmartGrid Technol. (ISGT)*, Jan. 2011.

[25] G. Dondossola, G. Garrone, J. Szanto, G. Deconinck, T. Loix, and H. Beitollahi, "ICT resilience of power control systems: Experimental results from the CRUTIAL testbeds," *Proc. IEEE/IFIP Int. Conf. Dependable Syst. Netw. (DSN)*, pp. 554-559, Jul. 2009.

[26] G. Dondossola, G. Deconinck, F. Garrone, and H. Beitollahi, "Testbeds for acsessing critical scenarios in power control systems," Berlin, Germany: Springer-Verlag, pp. 223-234, 2009.

[27] J. Hong, S.-S. Wu, A. Stefano, A. Fshosha,C.-C.Liu, P. Gladyshev, and M. Govindarasu, "An intrusion and defense testbed in a cyber-power system environment," *IEEE Power and Energy Society General Meeting*, Jul. 2011.

[28] C. Queiroz, A. Mahmood, and Z. Tari, "SCADASim: A framework for building SCADA simulations," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 589-597, Dec. 2011.

[29] V. Vyatkin, G. Zhabelova, N. Higgins, K. Schwarz, and N.C. Nair, "Towards intelligent smart grid devices with IEC 61850 interoperability and IEC 61499 open control architecture," *IEEE PES Transmission and Distribution Conference*, April 2010.

[30] R.E. Mackiewicz, "Overview of IEC 61850 and benefits," *IEEE PES Transmission and Distribution Conference*, pp. 376-383, May 2006.

[31] G. Clarke, D. Reynders, and E. Wright, Practical Modern SCADA Protocols, IDC technologies, 2004.

[32] Communication Networks and Systems for Power Utility Automation, IEC 61850-90-1 Standard: Use of IEC 61850 for the Communication between Substations, Mar. 2010, 1st Edition.

[33] Electrical Single Line Diagram - Part Two, Electrical Knowhow [Online]. Available: http://www.electrical-knowhow.com/2012/12/electrical-single-line-diagram-part-two.html

[34] Communication Networks and Systems in Substations, IEC 61850-5 Standard: Communication Requirements for Functions and Device Models, July 2003, 1st Edition.

[35] Specific Communication Service Mapping (SCSM), IEC 61850 8-1 Standard: Mapping to MMS (ISO/IEC9506-1 and ISO/IEC 9506-2), May 2004, 1st edition.

[36] J. Hong, C.-C. Liu, and M. Govindarasu, "Detection of cyber intrusions using network-based multicast messages for substation automation," *Proc. IEEE PES Innov. SmartGrid Technol. (ISGT)*, 2014.

[37] T. Pender, When Power Goes Out, a Squirrel is Likely to Blame, The Record, Oct. 2013 [Online]. Available: http://www.therecord.com/news-story/4164925-when-power-goes-out-a-squirrel-is-likely-to-blame/

[38] R.-J. Campbell, Weather-Related Power Outages and Electric System Resiliency, Congress Research Service 7-5700 [Online]. Available: http://www.fas.org/sgp/crs/misc/R42696.pdf

[39] D. Kushner, “The real story of Stuxnet,” IEEE Spectrum, vol. 50, no. 3, pp. 48-53, Mar. 2013.

[40] G.-L. Orgill, G.-W. Romney, M.-G. Bailey, and P.-M. Orgill, “The urgency for effective user privacy-education to counter social engineering attacks on secure computer systems,” *Proc. 5th conference on Information technology education (CITC5)*, pp. 177-181, 2004.

[41] B. Schneier, “Attack trees: Modeling security threats,” Dr. Dobb’s Journal, December 1999.

[42] J. Dawkins and J. Hale, “A systematic approach to multi-stage network attack analysis,” *Second IEEE International Information Assurance Workshop*, pp. 48-56, April 2004.

[43] A.-P. Moore, R.-J. Ellison, and R.-C. Linger, “Attack modeling for information security and survivability,” Survivable Systems, Technical Note CMU/SEI-2001-TN-001, Mar. 2001.

[44] C.-W. Ten, C.-C. Liu, and G. Manimaran, “Vulnerability assessment of cybersecurity for SCADA systems,” *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1836-1846, Nov. 2008.

[45] North American Electric Reliability Corporation, Cyber Attack Task Force, Final Report, May 2012 [Online]. Available: http://www.nerc.com/docs/cip/catf/12-CATF_Final_Report_BOT_clean_Mar_26_2012-Board%20Accepted%200521.pdf

[46] B. Kordy, L. Pietre-Cambacedes, and P. Schweitzer. “DAG-based attack and defense modeling: Don’t miss the forest for the attack trees,” arXiv preprint arXiv:1303.7397 (2013).

[47] R. Baldick, B. Chowdhury, I. Dobson, Z. Dong, B. Gou, D. Hawkins, Z, Huang, M. Joung, J. Kim, D. Kirschen, S. Lee, F. Li, J. Li, Z. Li, C. C. Liu, X. Luo, L. Mili, S. Miller, M. Nakayama, M. Papic, R. Podmore, J. Rossmaier, K. Schneider, H. Sun, K. Sun, D. Wang, Z. Wu, L. Yao, P. Zhang, W. Zhang, and X. Zhang, “Vulnerability assessment for cascading failures in electric power systems,” *Proc. IEEE PES PSEC*, pp. 1-9, Mar. 2009.

[48] United States Computer Emergency Readiness Team (US-CERT). Quarterly trends and analysis report, June, 2009. [Online]. Available: http://www.hsdl.org/?abstract&did=28501

[49] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proc. IEEE*, vol. 93, no. 11, pp. 1890-1908, Nov. 2005.

[50] T. Dy-Liacco, "Modern control centers and computer networking," *IEEE Comput. Appl. Power*, vol. 7, no. 4, pp. 17-22, Oct. 1994.

[51] D. Denning, "An intrusion-detection model," *IEEE Trans. Softw. Eng.*, vol. SE-13, no. 2, pp. 222-232, Feb. 1987.

[52] G. Coates, K. Hopkinson, S. Graham, and S. Kurkowski, "Collaborative, trust-based security mechanisms for a regional utility intranet," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 831-844, Aug. 2008.

[53] N. Liu, J. Zhang, and W. Liu, "A security mechanism of web services based communication for wind power plants," *IEEE Trans. Power Del.*, vol. 23, no. 4, pp. 1930-1938, Oct. 2008.

[54] C.-W. Ten, G. Manimaran, and C.-C. Liu, "Cybersecurity for critical infrastructures: attack and defense modeling," *IEEE Trans. Syst., Man, Cybern.*, vol. 40, no. 4, pp. 853-865, Jul. 2010.

[55] X. Guan, W. Wang, and X. Zhang, "Fast intrusion detection based on a non-negative matrix factorization model," *Journal of Network and Computer Applications*, vol. 32, no. 1, pp. 31-44, Jan. 2009.

[56] W. Wang, X. Guan, and X. Zhang, "Processing of massive audit data streams for real-time anomaly intrusion detection," *Computer Communications*, vol. 31, no. 1, pp. 58–72, Jan. 2008.

[57] S. Singh, H. Tu, W. Donat, K. Pattipati, and P. Willett, "Anomaly detection via feature-aided tracking and hidden markov models," *IEEE Trans. Syst., Man, Cybern.*, vol. 39, no. 1, pp. 144-159, Jan. 2009.

[58] C.-W. Ten, C.-C. Liu, and G. Manimaran, "Vulnerability assessment of cyberesecurity for SCADA systems," *IEEE Trans. Power Syst.*, vol. 23, no. 4, pp. 1836-1846, Nov. 2008.

[59] D. K. Holstein, "Wi-fi protected access for protection and automation," *Proc. IEEE Power Systems Conference and Exposition (PSCE)*, 2006, pp. 2004-2011.

[60] IEC 61850-6 standard, Configuration description language for communication in electrical substations related to IEDs, 1st ed. International Electrotechnical Commision, Mar. 2004.

[61] L. Hossenlopp, "Engineering perspectives on IEC 61850," *IEEE Power and Energy Magazine*, vol. 5, no. 3, pp. 45-50, May 2007.

[62] "Data and communication security - network and system management (NSM) data object models," in IEC/TS 62351-7 Ed. 1.0 (draft).

[63] D. Dzung, M. Naedele, T. P. V. Hoff, and M. Crevatin, "Security for industrial communication systems," *Proc. IEEE*, vol. 93, no. 6, pp. 1152-1177, Jun. 2005.

[64] C. Ozansoy, A. Zayegh, and A. Kalam, "The real-time publisher/subscriber communication model for distributed substation systems," *IEEE Trans. Power Del.*, vol. 22, no. 3, pp. 1411-1423, Jul. 2007.

[65] McAfee Foundstone Professional Services and McAfee Labs™, Global Energy Cyberattacks: "Night dragon," MacAfee, Feb. 2011. [Online]. Available:

http://www.mcafee.com/us/resources/white-papers/wp-global-energy-cyberattacks-night-dragon.pdf

[66] A. Matrosov, E. Rodionov, D. Harley, and J. Malcho, "Stuxnet under the microscope," ESET, North America, White Paper, Oct. 2010. [Online]. Available: http://www.eset.com/us/resources/white-papers/Stuxnet_Under_the_Microscope.pdf

[67] M. W. Berry, Z. Drmac, and E. R. Jessup, "Matrices, vector spaces, and information retrieval," *Society for Industrial and Applied Mathematics*, SIAM Review, vol. 41, no. 2, pp. 335-362, 1999.

[68] IEC 1686 Standard, IEEE Standard for substation intelligent electronic devices (IEDs) cybersecurity capabilities. IEEE, Feb. 2008.

[69] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Trans. Smart Grid*, vol. 1, no. 1, pp. 99-107, June 2010.

[70] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart transmission grid: Vision and framework," *IEEE Trans. Smart Grid*, vol. 1, no. 2, pp. 168-177, Sept. 2010.

[71] NIST 1108R2: NIST Framework and Roadmap for Smart Grid Interoperability Standards, National Institute for Standards and Technology, Feb. 2012 [Online]. Available: http://www.nist.gov/smartgrid/upload/NIST_Framework_Release_2-0_corr.pdf

[72] J. D. McDonald, *Electric Power Substations Engineering*. CRC press, 2012.

[73] V. Chandola, B. Arindam, and K. Vipin, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 15-58, July 2009.

[74] J. P. Anderson, “Computer security threat monitoring and surveillance,” Washing, PA, James P. Anderson Co., 1980.

[75] D. E. Denning, “An intrusion detection model,” *Proc. Seventh IEEE Symposium on Security and Privacy*, pp. 119-131, May 1986.

[76] J. Hall, M. Barbeau, and E. Kranakis, “Anomaly-based intrusion detection using mobility profiles of public transportation users,” *Proc. IEEE International Conference on Wireless And Mobile Computing Networking And Communications (WiMob’2005)*, vol. 2, pp. 17-24, Aug. 2005.

[77] B. Sun, F. Yu, K. Wu, Y. Xiao, and V.C.M. Leung, “Enhancing security using mobility-based anomaly detection in cellular mobile networks,” *IEEE Trans. Vehicular Technology*, vol. 55, no. 4, pp. 1385-1396, July 2006.

[78] N. Ye, S.M. Emran, Q. Chen, and S. Vilbert, “Multivariate statistical analysis of audit trails for host-based intrusion detection,” *IEEE Trans. Computers*, vol. 51, no. 7, pp. 810-820, Jul. 2002.

[79] J. Hu, X. Yu, D. Qiu, and H.-H. Chen, “A simple and efficient hidden markov model scheme for host-based anomaly intrusion detection,” *IEEE Network*, vol. 23, no. 1, pp. 42-47, Jan. 2009.

[80] M.J. Shevenell and R.F. Erbacher, “Design and implementation of an open network and host-based intrusion detection testbed with an emphasis on accuracy and repeatability,” *Proc. Ninth International Conference on Information Technology: New Generations (ITNG)*, pp. 409-416, April 2012.

[81] S. Bijan and A.M. Kazemitabar, "HIDMN: A host and network-based intrusion detection for mobile networks," *Proc. International Conference on Computer and Electrical Engineering (ICCEE)*, pp. 204-208, Dec. 2008.

[82] S.-M. Amin "Toward more secure, stronger and smarter electric power grids," *IEEE Power and Energy Society General Meeting*, July. 2011.

[83] G. Dan, H. Sandberg, M. Ekstedt, and G. Bjorkman, "Challenges in power system information security," *IEEE Security & Privacy*, vol. 10, no. 4, pp. 62-70, July 2012.

[84] M. Govindarasu, A. Hann, and P. Sauer, "Cyber-physical systems security for smart grid," Future Grid Initiative White Paper, PSERC, Feb. 2012 [Online]. Available: http://www.pserc.wisc.edu/documents/publications/papers/fgwhitepapers/Govindarasu_Future_Grid_White_Paper_CPS_May_2012.pdf

[85] GAO-11-117, Electricity Grid Modernization: Progress Being Made on Cyber Security Guidelines, but Key Challenges Remain to be Addressed. Government Accountability Office (GAO). Jan. 2011 [Online]. Available: http://www.gao.gov/new.items/d11117.pdf

[86] Guidelines for Smart Grid Cyber Security, National Institute for Standards and Technology, Aug. 2010 [Online]. Available: http://csrc.nist.gov/publications/nistir/ir7628/nistir-7628_vol2.pdf

[87] P. Judge and M. Ammar, "Security issues and solutions in multicast content distribution: A Survey," *IEEE Network*, vol. 17, no. 1, pp. 30-36. Jan. 2003.

[88] Specific Communication Service Mapping (SCSM), IEC 61850 9-2 Standard: Sampled Values over ISO/IEC 8802-3, Apr. 2004, 1st Edition.

[89] U.S. Department of Energy, Office of Electricity Delivery and Energy Reliability, National SCADA Test Bed Program, Jan. 2008 [Online]. Available: http://energy.gov/sites/prod/files/oeprod/DocumentsandMedia/DOE_OE_NSTB_Multi-Year_Plan.pdf

[90] A. Hahn and M. Govindarasu, "Cyber attack exposure evaluation framework for the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 835-843, June 2011.

[91] G. Dondossola, F. Garrone, G. Proserpio, and C. Tornelli, "Impact of DER integration on the cyber security of SCADA systems - The medium voltage regulation case study," *Integration of Renewables into the Distribution Grid, CIRED Workshop*, pp. 1-4, May 2012.

[92] M. Hillenmeyer, Machine Learning. Stanford University, July 2005. [Online]. Available: http://www.stanford.edu/~maureenh/quals/pdf/ml.pdf

[93] Aidan Harvey, "Cybersecurity enhancement in a power substation," M.S. Thesis, School of Electrical, Electronic and Mechanical Engineering, University College Dublin, 2011.

[94] Wireshark [Online]. Available: http://www.wireshark.org/docs/man-pages/tshark.html