Anomaly Detection for Cybersecurity of the Substations

Chee-Wooi Ten, *Member, IEEE*, Junho Hong, *Student Member, IEEE*, and Chen-Ching Liu, *Fellow, IEEE*

Abstract—**Cybersecurity of the substations in a power system is a major issue as the substations become increasingly dependent on computer and communication networks. This paper is concerned with anomaly detection in the computer network environment of a substation. An anomaly inference algorithm is proposed for early detection of cyber-intrusions at the substations. The potential scenario of simultaneous intrusions launched over multiple substations is considered. The proposed detection method considers temporal anomalies. Potential intrusion events are ranked based on the credibility impact on the power system. Snapshots of anomaly entities at substations are described. Simulation results using the modified IEEE 118-bus system have shown the effectiveness of the proposed method for systematic identification. The result of this research is a tool to detect cyber-intrusions that are likely to cause significant damages to the power grid.**

Index Terms—**Anomaly detection, cybersecurity of substations, defense system, network security.**

I. 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 [1]. Cybersecurity of intelligent electronic devices (IEDs) in the substations has been recognized as a critical issue for the smart grid [2]. 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 cybersecurity 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 [3]–[5]. A survey of the important issues related to cascading events has been reported [6].

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 [7]. This paper 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 [8]. NIST also identified key attributes of the logical design for intrusion-based attacks on power equipment that is critical to standardization and modeling [9], [10]. The hypothesized intrusion scenarios in this paper 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 [11], [12], 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 [13]. Trust-based security mechanisms have been designed to suppress cyber-attacks or other malicious events for event logging, analysis, or blocking power system operations [14], [15]. 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 [16]. 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 [17], [18].

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 [19]. Enhancement of the previous framework is required for two reasons [16], [20]: 1) cyberinfrastructure 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

Manuscript received November 16, 2010; revised April 04, 2011; accepted June 05, 2011. Date of publication August 30, 2011; date of current version November 23, 2011. This work is sponsored by Science Foundation Ireland (SFI) at University College Dublin (UCD), Ireland. Paper no. TSG-00214-2010.

C.-W. Ten is with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton, MI 49931 USA (e-mail: ten@mtu.edu).

J. Hong and C.-C. Liu are with the School of Electrical, Electronic and Mechanical Engineering, University College Dublin, Belfield, Dublin 4, Ireland (e-mail: junho.hong@ucdconnect.ie; liu@ucd.ie).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSG.2011.2159406

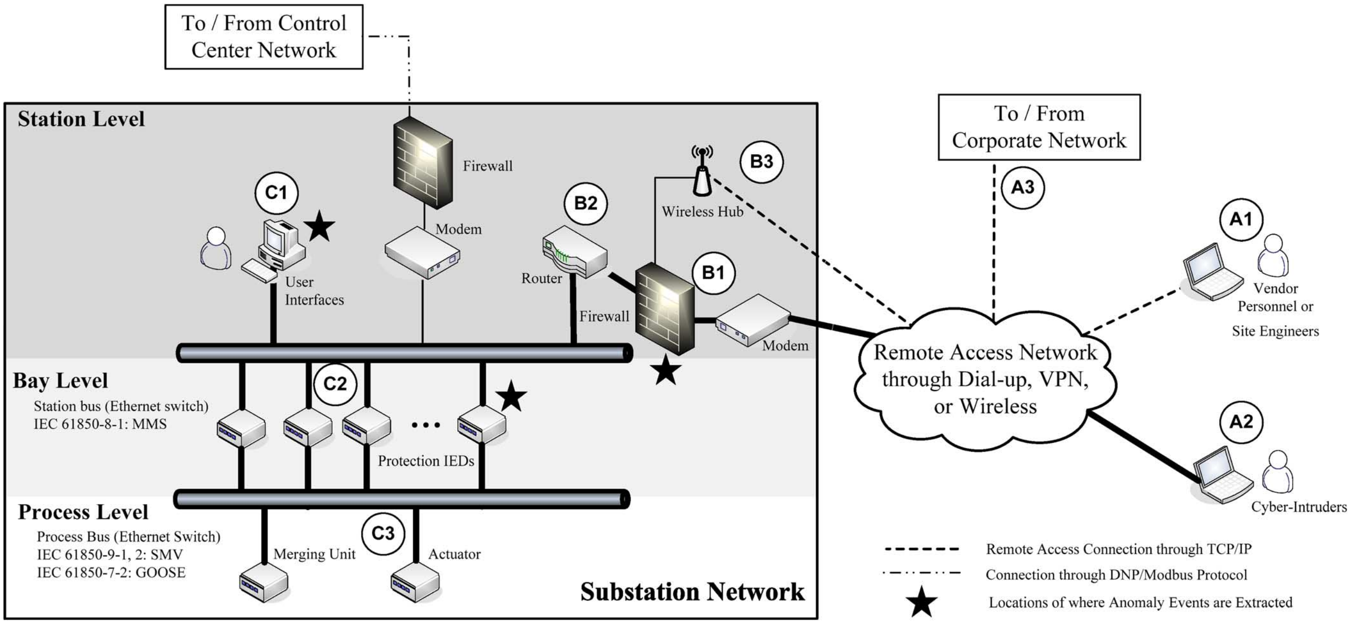
To / From Control Center Network
Station Level
Firewall
Modem
C1
User Interfaces
B2
Router
Firewall
Wireless Hub
B3
B1
Modem
To / From Corporate Network
A3
A1
Vendor Personnel or Site Engineers
Remote Access Network through Dial-up, VPN, or Wireless
A2
Cyber-Intruders
Bay Level
Station bus (Ethernet switch)
IEC 61850-8-1: MMS
C2
Protection IEDs
Process Level
Process Bus (Ethernet Switch)
IEC 61850-9-1, 2: SMV
IEC 61850-7-2: GOOSE
Merging Unit
C3
Actuator
Substation Network
Remote Access Connection through TCP/IP
Connection through DNP/Modbus Protocol
Locations of where Anomaly Events are Extracted

Fig. 1. Path combinations of intrusion scenarios to substation level networks (bold lines).

upon substations and the resulting impact is not captured or observed.

In this paper, 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 paper 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. In the authors' previous work, the conceptual design of RAIM was reported [16]. The focus of [20] 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 [20]. This paper contributes to the state of the art of cybersecurity 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 cybersecurity of the computer network at multiple substations in order to enhance resilience of the power grid.

Section II provides a generalized intrusion scenario. Section III describes a prototype design for real-time monitoring, anomaly detection, impact analysis, and mitigation strategies (RAIM) [16]. Section IV is concerned with anomaly construction based on temporal events. Section V provides an attack analysis with the identification of classes of contingencies with different levels of complexity. Section VI is a case study of simultaneous impact evaluations based on an anomaly set. Section VII provides the conclusion and recommendations for future work.

II. Hypothesized Intrusion Scenarios

As depicted in Fig. 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:

- 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 [21]. Discussion on intrusion scenarios through local wireless connection is outside the scope of this paper. 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 [22], [23]. Once the required information is identified, actions to operate circuit breakers can be launched through direct IED connection.

III. 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 [24]. This

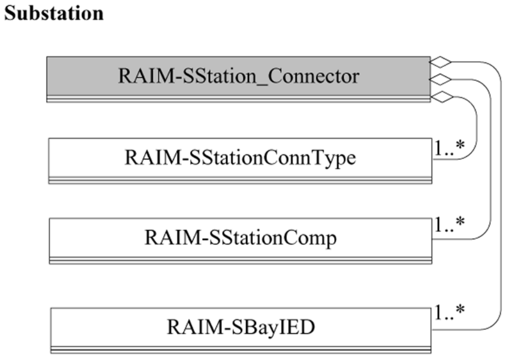

Fig. 2. The object modeling of RAIM for substation.

information can be used for event constructions to extract evidences for intrusions.

Fig. 2 describes the data object model abstraction for the RAIM framework developed in the previous work by the authors [16].

The current paper provides an anomaly detection algorithm for substation-based intrusions. The proposed substation RAIM model in Fig. 2 is divided into 3 data object models, i.e., *RAIM-SStationComp*, *RAIM-BayIED*, and *RAIM-SStation ConnType*.

A *RAIM-SStationComp* consists 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 [25]. 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) [26].

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 computers or IEDs.
2) The changes of file systems on local user interface.

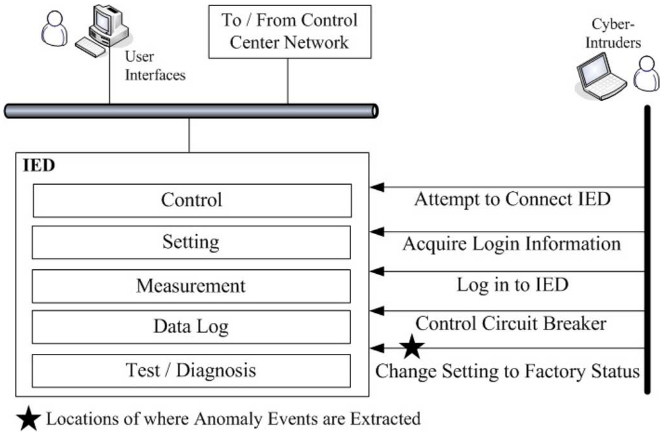

Fig. 3. Anomaly detection of a cyber-attack at the IED level.

3) The changes of IED critical settings that may misoperate the system operations.

These attributes are illustrated in Fig. 2. The intrusion attempts on each device or computer in substation level network are included in *RAIM-SStationSBayIED.attempts* and *RAIMSStationComp.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*.

IV. 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.

As shown in Fig. 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. 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* [27] and ii) *Stuxnet* [28]. 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 four 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 III, 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, $\pi^{\rm fs}$ represents a change of the file system, $\pi^{\rm cs}$ 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.,

$$\boldsymbol{\pi}_{(1\ \times k)} = \Big(1\ \ \alpha\pi^{a}_{(T\times L)}\ \beta\pi^{\rm fs}_{(T\times M)}\ \delta\pi^{\rm cs}_{(T\times N)}\ \epsilon\pi^{o}_{(T\times O)}\Big). \quad (1)$$

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 $\alpha < \beta < \delta < \epsilon$. The symbols, L, M, N, and O, denote the size of each element in (1). T is the number of records of anomaly for each period of time. An example is given in (2) as an example of the attributes described in (1) with $L = 1$, $M = 1$, $N = 1$, and $O = 1$ and weighting factor parameters, $\alpha = 1$, $\beta = 5$, $\delta = 10$, $\epsilon = 20$. 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,

$$\boldsymbol{\pi}_{(1\times 5)} = (1\ 1\ 0\ 10\ 0). \quad (2)$$

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

$$\boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 20 \\ 1 & 1 & 0 & 0 & 20 \\ 1 & 1 & 0 & 0 & 20 \\ 1 & 1 & 5 & 0 & 20 \\ 1 & 1 & 5 & 0 & 20 \\ 1 & 1 & 5 & 10 & 20 \\ 1 & 1 & 5 & 10 & 20 \end{bmatrix}. \quad (3)$$

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

$$\widehat{\boldsymbol{\pi}} = \frac{\boldsymbol{\pi}}{||\boldsymbol{\pi}||_2}. \quad (4)$$

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

$$\widehat{\boldsymbol{\Pi}} = \begin{bmatrix} 0.0499 & 0 & 0 & 0 & 0.9988 \\ 0.0499 & 0.0499 & 0 & 0 & 0.9975 \\ 0.0499 & 0.0499 & 0 & 0 & 0.9975 \\ 0.0484 & 0.0484 & 0.2420 & 0 & 0.9679 \\ 0.0484 & 0.0484 & 0.2420 & 0 & 0.9679 \\ 0.0436 & 0.0436 & 0.2178 & 0.4356 & 0.8712 \\ 0.0436 & 0.0436 & 0.2178 & 0.4356 & 0.8712 \end{bmatrix}. \quad (5)$$

Temporal anomaly is determined by the two vectors that occur at two different time instants [29]. 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

$$\boldsymbol{\Delta}_{\rm ta} = \mathbf{1} - \frac{\widehat{\boldsymbol{\pi}} \cdot \widehat{\boldsymbol{\pi}}_{-1}^{\top}}{||\widehat{\boldsymbol{\pi}}||_2 \cdot ||\widehat{\boldsymbol{\pi}}_{-1}||_2}. \quad (6)$$

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

$$\Delta_{\rm ta}^{\top} = (0\ \ 0.0012\ \ 0\ \ 0.0297\ \ 0\ \ 0.0999\ \ 0). \quad (7)$$

The first vector of $\Delta_{\rm ta}^{\top}$ is 0 because there is nothing to compare to for the first row of $\widehat{\boldsymbol{\Pi}}$ as this procedures starts. After the first element of (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 $\widehat{\boldsymbol{\Pi}}$ is formed by normalizing the matrix $\boldsymbol{\Pi}$ as illustrated in (6). The rank of $\widehat{\boldsymbol{\Pi}}$ for this substation is used to determine an index ζ, i.e.,

$$\zeta = {\rm rank}(\widehat{\boldsymbol{\Pi}}) - 1. \quad (8)$$

Based on (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 $\mathbf{p}$ is used to denote a column that is calculated from the anomaly corresponding to the weighted anomaly entities from $\widehat{\boldsymbol{\Pi}}$ and the temporal anomaly $\Delta_{\rm ta}$. B is used to represent an $n \times m$ matrix with all elements equal to 1. That is,

$$\mathbf{p} = \begin{cases} 0 & \text{if } \zeta = 0 \\ \widehat{\boldsymbol{\Pi}} \cdot \mathbf{B} \cdot \Delta_{\rm ta} & \text{Otherwise.} \end{cases} \quad (9)$$

The observation over a certain time period can be made throughout all substations using (9).

For the example in (5), $\zeta = 4 - 1 = 3$. Hence, $\mathbf{p}^{\top} = (0.1372, 0.1435, 0.1435, 0.1709, 0.1709, 0.2109, 0.2109)$. 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 $\mathbf{p}$, i.e.,

$$\varrho = \max \mathbf{p} - \overline{\mathbf{p}}. \tag{10}$$

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 ϱ^* 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.

V. 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:

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 nondivergent condition is an indication of an infeasible operating condition, e.g., voltage collapse.
2) *Single and Double Substations*: This category of substations 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 authors' previous work [20] 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

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

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

$$V_{\text{sub}}(S_1, \cdots, S_k) = \gamma_{1,\cdots,k} \cdot \max(\varrho, \cdots, \varrho_k). \tag{12}$$

where $1, \ldots k$ represent the hypothesized outage involving k substations and $\gamma_{1,\ldots,k}$ 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 $V_{\text{sub}}(S_1, \cdots, S_k)$.

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.

VI. 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. 4.

The IEEE 118-test system, as described in Fig. 4, consists of 109 substations with a total load of 4266 MW. The hypothesis here is that an intrusion into a substation will lead to operation of 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.

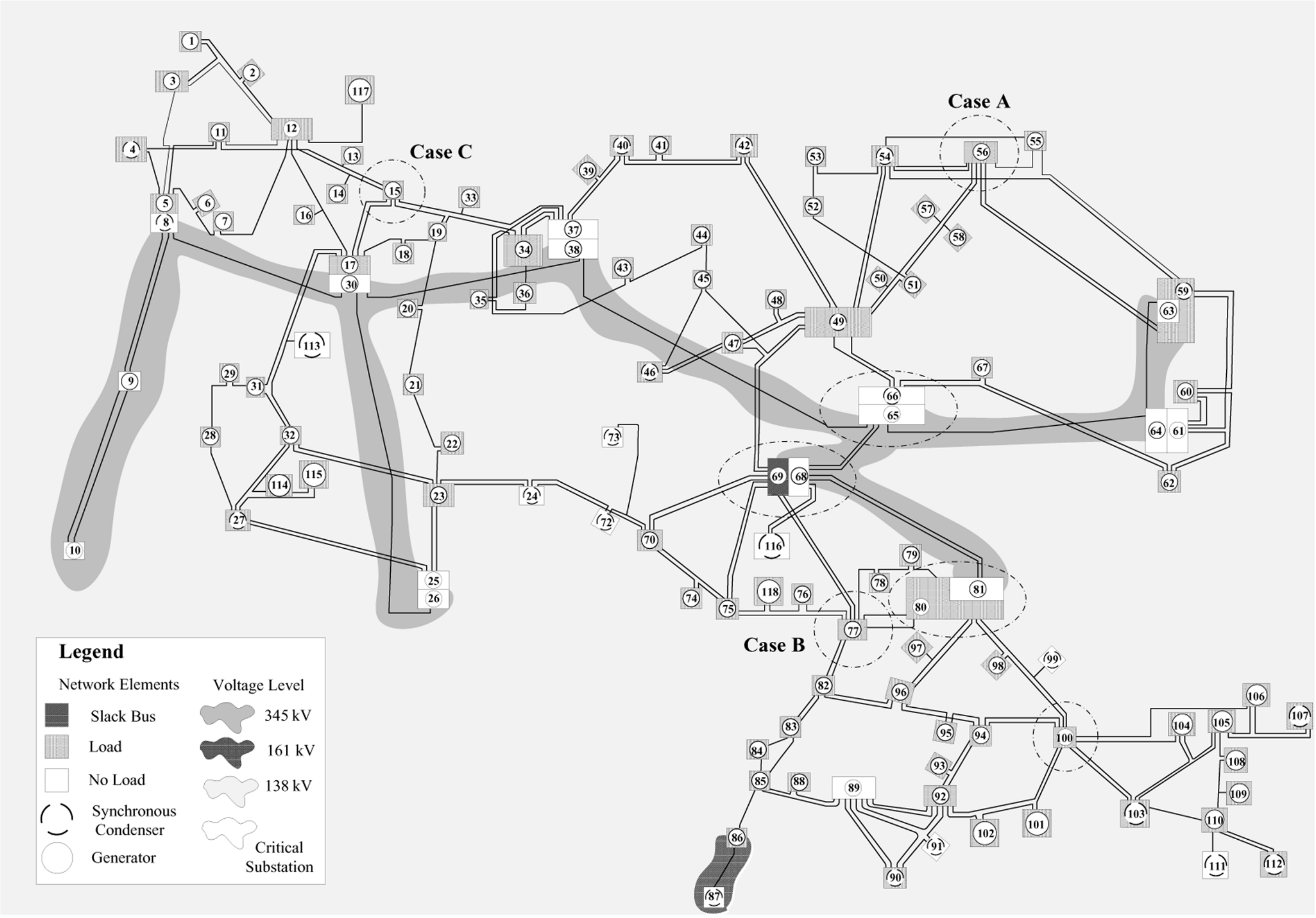

Fig. 4. IEEE 118-test system.

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 $\mathbf{\Pi}$ in (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.

A. Case Study I: Simultaneous Attack

Table I 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} as shown in the table.

The second case in Table I 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.

TABLE I
HYPOTHESIZED CYBER-ATTACK UPON SINGLE AND MULTIPLE SUBSTATION(S)

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

Tables II and III report sample IED logs of substations 49 and 2526, respectively. Table II 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 III 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.

Based on the logs in Tables II and III, matrix $\mathbf{\Pi}$ is constructed for each of the two substations, 49 and 2526. The results are

TABLE II
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 III
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

$$\mathbf{\Pi}_{\mathrm{sub49}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \\ 1 & 0 & 0 & 10 & 0 \end{bmatrix}. \tag{13}$$

$$\mathbf{\Pi}_{\mathrm{sub2526}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \\ 1 & 0 & 0 & 0 & 20 \end{bmatrix}. \tag{14}$$

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 I. 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 (1)–(10) based on the IED logs in (13), (14), i.e., Π_{sub49}, Π_{sub2526}, respectively. Therefore, by (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.

B. Case Study II: Most Critical Substations

The system given in Fig. 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. 5. Hence, these 4 cases are categorized as critical.

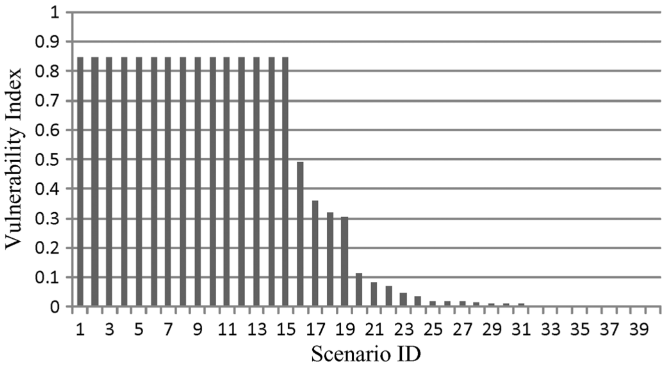

Fig. 5. Vulnerability ranking of enumerating credible events for $\mathrm{Subs}_{\mathrm{case}}$.

TABLE IV
CRITICAL SCENARIO FOR $\mathrm{Subs}_{\mathrm{case}}$

	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

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

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, $1 \ldots 40$, is shown in Fig. 5. Fig. 5 shows the highest impact scenarios of single and double and multiple substations in the set. Scenarios with ID $1 \ldots 15$ are highly vulnerable combinations of substation outages. They are listed in Table IV. 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 IV.

Fig. 5 shows 40 scenarios in single, double, and multiple substations, the first 18 are above the mean value of vulnerability

TABLE V
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

TABLE VI
CRITICAL 5 SCENARIOS FOR EACH STUDY CASE IN TABLE V

	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

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.

C. Case Study III: Highest Impact Factor Scenarios

Three more cases are selected for this study. In Table V, 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 VI. 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.

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

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.

TABLE VII
NUMBER OF SCENARIOS AND CALCULATION TIME IN TABLE V

	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

As described in Section IV, once vulnerability indices are computed, cybersecurity 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 cybersecurity in substation IEDs. 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 [30]. 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.

VII. CONCLUSION

The proposed framework is intended to improve the cybersecurity of existing substation computer networks. To implement this framework, further efforts are required. The equipment and software deployed at the substations have been equipped with communication technologies. Therefore, the requirements for identifying relevant properties of cybersecurity and performance are crucial. The contribution of this paper 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 system. The proposed anomaly detection algorithm described in Section IV for identification of local anomalies at the substation level can be further extended to perform a large-scale cyber-power system evaluation in a control center environment. At the control center level, the communication links with substations will need to be expanded to incorporate information needed for cybersecurity assessment. The effort for implementation of agent-based extraction from each substation level network can be significant as it requires prototypes corresponding to each substation type. Efficient delivery of information from substations or control centers will be needed to help identify critical messages in the on-line environment. Various techniques are desirable for visualization of the cyber system health, in term of vulnerability level and other critical information.

Acknowledgment

This research is sponsored by Science Foundation Ireland (SFI) through a Principal Investigator Award. The authors would like to thank Mr. A. Ginter, Industrial Security Director at Waterfall Security Solutions, and Mr. S.-S. Wu, University College Dublin, for their great support and useful discussions.

References

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

[2] The Smart Grid Interoperability Panel-Cybersecurity Working Group, "Guidelines for smart grid cybersecurity: Vol. 3, Supportive analyses and references," National Institute of Standards and Technology, U.S. Department of Commerce, NIST Interagency Rep. 7628, Aug. 2010.

[3] 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.

[4] D. Kirschen and B. Wollenberg, "Intelligent alarm processing in power systems," *Proc. IEEE*, vol. 80, no. 5, pp. 663–672, May 1992.

[5] 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.

[6] 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," in *Proc. IEEE PES PSEC*, Mar. 2009, pp. 1–9.

[7] U.S. Computer Emergency Readiness Team (US-CERT), "Quarterly trends and analysis report," June 2009 [Online]. Available: http://www.us-cert.gov/press_room/trendsanalysisQ109.pdf

[8] 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/page.php?cid=2|20

[9] 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," National Institute of Standards and Technology, U.S. Department of Commerce, NIST Interagency Rep. 7628, Aug. 2010.

[10] Cross-Domain Event Detection Analysis, and Response—The Smart Grid Interoperability Panel—Cybersecurity Working Group, "Guidelines for smart grid cybersecurity: Vol. 3, Supportive analyses and references National Institute of Standards and Technology, U.S. Department of Commerce, NIST Interagency Rep. 7628, Aug. 2010.

[11] 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.

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

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

[14] 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.

[15] 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.

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

[17] X. Guan, W. Wang, and X. Zhang, "Fast intrusion detection based on a non-negative matrix factorization model," *J. Network Comput. Appl.*, vol. 32, no. 1, pp. 31–44, Jan. 2009.

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

[19] 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. A, Syst., Humans*, vol. 39, no. 1, pp. 144–159, Jan. 2009.

[20] 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.

[21] D. K. Holstein, "Wi-fi protected access for protection and automation," in *Proc. IEEE Power Syst. Conf. Expo. (PSCE)*, 2006, pp. 2004–2011.

[22] *Configuration Description Language for Communication in Electrical Substations Related to IEDs*, IEC 61850-6 standard, International Electrotechnical Commision, Mar. 2004, 1st ed.

[23] L. Hossenlopp, "Engineering perspectives on IEC 61850," *IEEE Power Energy Mag.*, vol. 5, no. 3, pp. 45–50, May 2007.

[24] Data and Communication Security—Network and System Management (NSM) Data Object Models in IEC/TS 62351-7 Ed. 1.0 (draft).

[25] 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.

[26] 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.

[27] McAfee Foundstone Professional Services and McAfee Labs, Global Energy Cyber-attacks, "Night dragon," MacAfee Feb. 2011 [Online]. Available: http://www.mcafee.com/us/resources/white-papers/wp-global-energy-cyberattacks-night-dragon.pdf

[28] A. Matrosov, E. Rodionov, D. Harley, and J. Malcho, Stuxnet Under the Microscope ESET, "North America", White Paper, Oct. 2010 [Online]. Available: http://www.esetnod32.ru/.company/.viruslab/analytics/doc/Stuxnet_Under_the_Microscope.pdf

[29] M. W. Berry, Z. Drmac, and E. R. Jessup, "Matrices, vector spaces, and information retrieval," *SIAM Rev.*, vol. 41, no. 2, pp. 335–362, 1999.

[30] *IEEE Standard for Substation Intelligent Electronic Devices (IEDs) Cybersecurity Capabilities*, IEC 1686 Standard, IEEE, Feb. 2008.

Chee-Wooi Ten (S'07–M'08) received the B.S.E.E. and M.S.E.E. degrees from Iowa State University, Ames, in 1999 and 2001, respectively, and the Ph.D. degree from University College Dublin, Ireland, in 2009.

He was an Application Engineer with Siemens Energy Management and Information System in Singapore from 2002 to 2006. He is currently an Assistant Professor at Michigan Technological University, Houghton. His primary research interests are cybersecurity of power system infrastructure and power automation applications on SCADA systems.

Junho Hong (S'10) received the B.S.E.E. and M.S.E.E. degrees from Myongji University, Korea, in 2008 and 2010, respectively. He is currently working toward the Ph.D. degree at University College Dublin.

His research interests include substation automation and cybersecurity of the power grid monitoring and control system.

Chen-Ching Liu (F'94) received the Ph.D. degree from the University of California, Berkeley.

He was Palmer Chair Professor of Electrical Engineering at Iowa State University and a Professor of Electrical Engineering at the University of Washington. He is currently a Professor of Power Systems at the University College Dublin, Ireland.

Dr. Liu received an IEEE Third Millennium Medal in 2000 and the IEEE Power and Energy Society Outstanding Power Engineering Educator Award in 2004. He served as Chair of the Technical Committee on Power System Analysis, Computing and Economics (PSACE), IEEE Power and Energy Society.