Cyber Inference System for Substation Anomalies Against Alter-and-Hide Attacks

Chong Wang, *Student Member, IEEE*, Chee-Wooi Ten, *Senior Member, IEEE*,
Yunhe Hou, *Senior Member, IEEE*, and Andrew Ginter, *Member, IEEE*

***Abstract*—Alarms reported to energy control centers are an indication of abnormal events caused by either weather interruptions, system errors, or possibly intentional anomalies. Although these initiating events are random, e.g., faults on transmission lines struck by lightning, the existence of electronically altered measurements may implicate the process to identify root causes of abnormal events. This paper is concerned with alter-and-hide (AaH) attacks by tampering the actual measurements to normal states with the background of disruptive switching actions that hide the true values of local events from operators at the control center. A cyber inference system (CyIS) framework is proposed to synthesize all sequential, missing, or altered alarms of related substations against AaH attacks. The stochastic nature of such attack events is modeled with probabilities as an integer programming problem with multiple scenarios. The proposed method is utilized to verify alarm scenarios for a conclusion of the potential AaH attacks on the substations.**

***Index Terms*—Alter and hide (AaH), cyber inference system, future control center, situation awareness, tampered alarm events.**

I. Introduction

SINCE the early 1970s, the centralized grid control systems have been computerized with monolithic energy management systems (EMS) to help operators monitor system conditions. Alarms are generated, processed, and presented to the operators for operational readiness under abnormal circumstances [1], [2]. Such disturbance events can be either intentional or unintentional and may require human intervention with a control response to mitigate the risks of system collapse. Due to the sophisticated topologies of busbar arrangement in substations, a single disturbance can result in a large number of alarm events [3], [4]. For two or more independent disturbances, they may result in overlapped alarms, and some might be missed or delayed. These are circumstances that most common alarm analyzers deal with. There has been extensive research that focuses on alarm processing algorithms based on imperfect reported alarms [4]–[7]. Such expert systems were implemented as an online application of fault diagnosis to handle corrupted alarm scenarios [8]–[10]. In addition, there are other methods have been implemented, such as, Petri nets [11]–[13], artificial neural networks [14]–[16] and rough sets [17]. The evolutionary alarm processors in software engineering are used to synthesize the voluminous alerts from substations. A simplified version of the alarm-related information is to help system operators systematically pinpoint the cause of events and possible short circuit locations from supervisory control and data acquisition (SCADA) systems in the control center.

IP-based instrumentation is now the main stream of communication technology deployment. Today's power automation using information communication technologies has become increasingly distributed and networked, which has raised concerns about cyber threats. With Internet protocol (IP)-based systems for wide-area communication, SCADA systems can be vulnerable to a cyberattack [18]. Misconceptions have been addressed with respect to skills and malicious behaviors that would progress slower than the development of infrastructures and technologies [19]. While communications have evolved, the malicious intent of attackers remains. This is one of the critically important priorities for cyber-physical interdependent research [20]. The process of SCADA cyberattacks can be generalized, and a conceptual framework has been introduced [21]. Analyzing incomplete spatiotemporal events with their dependencies can be challenging. More precisely, capturing the anomalies within substation-level networks can be strengthened with additional anomaly detectors in distributed configuration [22], [23]. However, these would cost asset owners to invest additional cybersecurity protection in substations.

Despite all preliminary efforts made to implement protection technologies in substation networks, there remains serious concerns regarding the impacts of attacks using domain-specific knowledge of individuals. Such attacks could aggregate operating conditions as well as mislead operators to make erroneous decisions. Cyber-situation awareness in the control room can be further improved by utilizing already existing metering or alarm information with enhanced inference algorithms. The major contribution of this work is to employ a mathematical model using integer programming to distinguish between potentially malicious incidents of AaH attacks and common disturbances. The rest of the paper is organized as follows. Section II presents the AaH attack model with the proposed cyber inference system (CyIS). In Section III, logical relations among alarms and event hypotheses are introduced,

This work was supported by US National Science Foundation (ECCS1128512), National Natural Science Foundation of China (51277155), Research Grant Council, Hong Kong SAR (ECS739713 and GRF 17202714), and State Key Laboratory of Advanced Electromagnetic Engineering and Technology of China in HUST (2015KF006).

C. Wang and Y. Hou are with the Electrical & Electronic Engineering Department, University of Hong Kong, Pokfulam, Hong Kong, SAR, China. (e-mails: wangc@eee.hku.hk, yhhou@eee.hku.hk).

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

A. Ginter is with Waterfall Security Solutions, Israel (e-mail: andrew.ginter@waterfall-security.com).

and the model is formulated without considering cyberattacks and stochastic issues of alarms. Section IV relates stochastic nature of cyberattacks to the proposed model. Section V verifies simulation results. Section VI concludes with future work.

II. AaH Attack Modeling

A substation communication network is a complex hybrid system. The cyber-physical security problem is often related to the relationship between physical infrastructure and potential cyber manipulation, resulting *directly* or *indirectly* in operational impacts. This may cause overloading damage and may affect any other states of equipment health with both short-term or long-term implications. *Stuxnet* has proven the feasibility of intelligent attacks in the computing world [24], [25]. Local IP-based intelligent electronic devices (IEDs) can be infected and a worm could be programmed with domain-specific knowledge that embeds codes with malicious commands in IEDs while reporting normal states of operations to control centers [26]. This section consists of three parts: (A) AaH Attack, (B) Attack execution plan, and (C) Anomaly inference of the AaH Attacks.

A. AaH Attacks

Alter-and-hide (AaH) attacks refer to an alteration of true values of a partial system, such as, one substation network or more, with an intent to avoid detection by defenders. Malicious activities, such as plotting for a cyberattack in the background covertly, are assumed in a programmed software agent. A successful exploration of critical assets may enable attackers to manipulate metering information intelligently that can mislead operators to action or inaction in a control room. Such as attack can cause serious consequences in system operation. One example of such characteristics is the behavior of *Stuxnet*. The attack agent generally has three malicious parts: (1) a worm executes all routines, (2) an intelligent module propagates the worm to additional machines, and (3) rootkit components hide malicious files and processes to prevent detection [27].

This work is motivated by the credibility of potential intelligent cyberthreats. Unlike the existing conventional attacks focusing on cyber networks, the AaH is a behavioral attack agent that can also be programmed with the intelligence to disruptive switching. Successful implementation by blocking the locally generated alarms at substation level on circuit breakers would prevent alerting operators in the control room. By harnessing system topology and metering integrity, availability as well as topology information, substation anomalies of a possible AaH attack can be inferred. The following describes the assumptions:

1) There may be one or more attackers who can coordinate their attacks from different compromised substations. However, the attackers may not know the relation between one and the other(s), e.g., the connectivity of affected substations if they are closely related in the region.
2) The attackers may not know complete information about a power grid's connectivity. The successful intrusions to multiple substation networks would be restricted to the information the attackers can observe over a short period of time. For example, if the attackers compromise two substations, it is likely that they do not know the relationship between the two compromised substations. Unless, there is information to relate the addresses connecting the local SCADA network to the other compromised network(s).
3) The attackers know the connections between substation(s) and control center, and the mapping addresses of a measurement list. In this case, the attackers would need to understand the DNP communication protocol as well as knowing how to create fake information for sending to the control center. This process happens concurrently to plan for attacks using local substation SCADA.

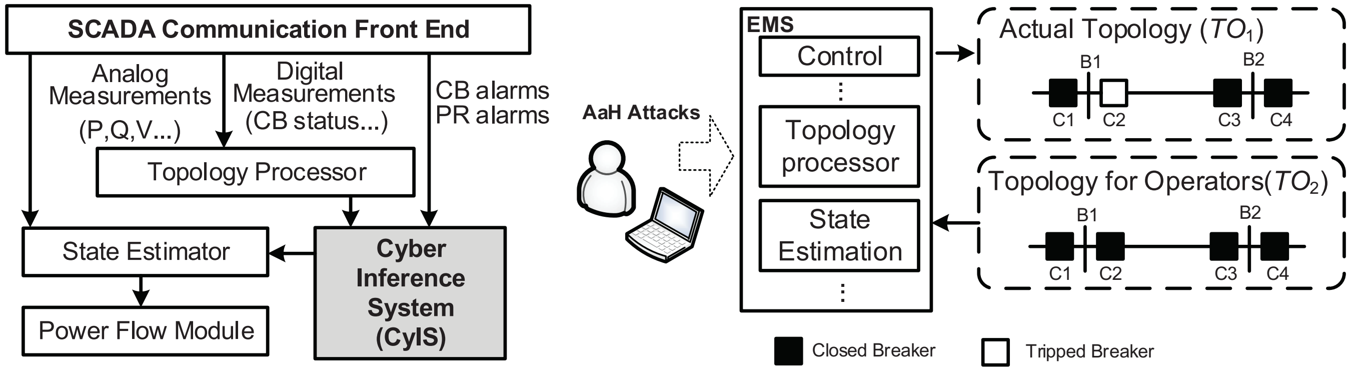
Hypothetically, an attack plan can be executed in an IP-based substation network.

B. Attack Execution Plan

Consider an example of a Distributed Network Protocol version 3.0 (DNP3) replay attack. The following describes generally how an AaH attack can be executed:

1) A controller server is deployed in a substation to communicate across a TCP/IP wide area network (WAN) with a central SCADA server. This is the only controller in a substation that serves as a data concentrator for DNP3 protocol data gathered from all substation devices, and reports this data to the central SCADA server using a single DNP3 address and session.
2) The substation controller is typically set up in a way where it is remotely accessible via a user interface, such as *Telnet*, secure shell (SSH), or remote desktop, for the purposes of system administration and trouble-shooting. To initialize an attack in a substation, attackers require gaining access to this administrative user interface. Such access can be gained, for example, remotely via a stolen password, a hijacked remote access session, or by physically breaking in the substation to access the substation automation controller.
3) Once in control of an administrative session, the attacker transfers, installs, and runs a packet sniffing tool. The tool is written by someone who understands DNP3 packet formats and records the register numbers and values the controller reports to the central SCADA system. After running this tool for a period of time, the attacker stops and disables the legitimate DNP3 software server. The attacker then starts a malicious DNP3 emulator. The emulator impersonates the legitimate DNP3 software, responding to DNP3 queries from the central SCADA server the same way as the legitimate controller responded. But the malicious emulator reports only the recorded register values with faked time stamps.

At this point, the attackers can change any settings on substation equipment without these changes being reported to control center systems by equipment in the compromised substation. Alternatively, if the attackers are physically present, they can execute their attack plan directly to the equipment



(a) CyIS Interaction with State Estimator, Topology and Alarm Processors

(b) Alter-and-Hide (AaH) attacks that reflects real and fake states to both local and control center

Fig. 1. AaH attacks on substation networks with CyIS verification on open circuit breakers and generated alarms blocked by attackers.

panel, while using the emulator to report normal conditions on substation status. Security measures such as encryption, anti-virus, whitelisting and many others increase the difficulty of this attack but are unable to prevent such an attack entirely. Fundamentally, this is a “software attack on a software system.” In practice, all software has defects, and some defects are security vulnerabilities that can be compromised. This includes security software. This relay attack example is only one of many examples of possible attacks, and comparable attacks are possible for other standardized communication protocols.

C. Anomaly Inference of the AaH Attacks

Fig. 1 (a) shows the existing architecture of the modules for *alarm and topology processors*, *state estimator*, and *power flow module*. The topology processor uses the statuses of circuit breakers (CB) to get the system topology. Based on the system topology and analog measurements, e.g., active and reactive power, a state estimation is first performed, following by power flow to determine the best snapshot of a system state. False data injection, e.g., changing measurements of active power, by intruders can be detected in the process of state estimation [28], [29] in the existing architecture. However, the existing architecture cannot deal with AaH attacks, which open substation circuit breakers and falsify substation data reports. For example, the circuit breaker C_2, in Fig. 1 (b), is opened by the attackers, and the attackers change the values reported to an energy management system (EMS), which identifies the system topology as TO_2. With this topology, the state estimation and the power flow calculation would be inaccurate. Therefore, before performing the state estimation, the proposed cyber inference system (CyIS) will determine anomalies before passing the information to state estimator and power flow module, as shown in Fig. 1 (b). The CyIS mainly gathers all datasets from system topology, received CB alarms and PR alarms to infer if a substation might be under an AaH attack, and then update the system topology.

III. Cyber Inference System

The cyber inference system (CyIS) is a spatiotemporal anomaly agent that synthesizes all topologies and alarm information to detect possible AaH attacks. Fig. 2 details the architecture of CyIS and how it gathers and correlates information inconsistencies. Typically, SCADA systems present all alarms to system operators following a disturbance event. These alarms include analog measurements, such as, current, voltage, active and reactive power, and digital measurements, e.g., protective relay (PR) operation and circuit breaker (CB) tripping. We generalize the digital (binary) measurement alarms associated with protective relays and circuit breakers, i.e., PR and CB alarms, in the CyIS module. These are classified in accordance to the IEC61850 standard. Based on the system topology, logical rules between PR operation and CB tripping are modeled in an engineering workstation before transitioning to online applications. The missed/delayed/tampered alarms are attributed with probabilities in relations among all events, e.g., disturbances, PR operation and CB tripping. These probabilities and the basic rules are considered as an index of evaluation with integer constraints representing logical relations between possible events with reported alarms. Typically, the size can be practically large and scenario reduction is employed to handle a larger set of increasing alarms with event hypotheses. The integer programming is applied to determine irregularities.

A. Anomaly Modeling

The irregularity of incoming alarms is determined based on combinations of telemetered measurements from substation automation networks. This section establishes a generalized alarms with event hypotheses and expected arrival of alarms at the control center. The following PR and CB definitions are given to establish notational consistency for the proposed model:

1) Reported Alarms: Following an event of disturbance, whether it is a fault or potential cyberattacks, alarms are generated and transmitted to control centers. These reported alarms can be delayed, missed, or electronically tampered. We consider two types of alarms for the development of the proposed inference system.

- *PR alarms*: Alarms are related to PR operation.

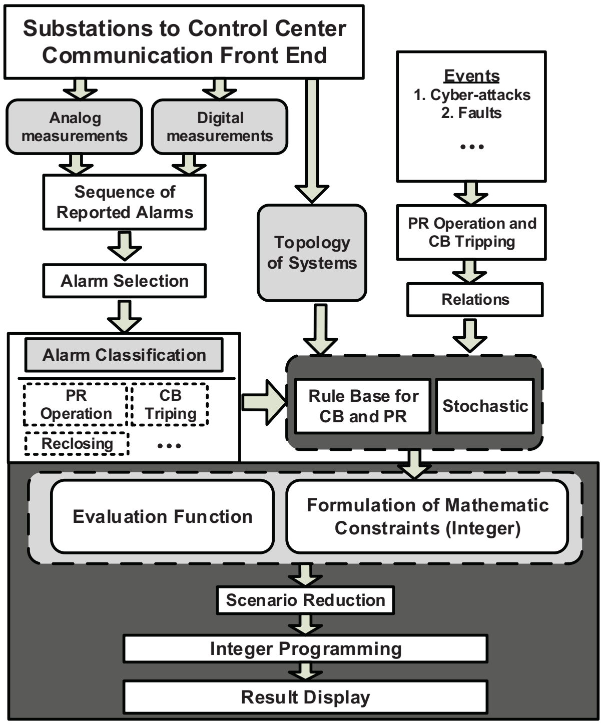


Fig. 2. Components of Proposed Cyber Inference System (CyIS) and Relations with Existing Power Control Center Applications.

- *CB alarms*: Tripping alarms CB by the PR operation.

The RA_i represents a single received alarm and $\mathbf{\Omega}^{(RA)}$ denotes a set of reported alarms, including both PR alarms and CB alarms at the control center end.

2) Event Hypotheses EH: These hypotheses represent hypothesized events based on the received alarms. Possibilities of hypothesized events are inferred based on the reported alarms in the control center. The $\mathbf{\Omega}^{(EH)}_{RA_i}$ denotes a set of event hypotheses corresponding to the alarm RA_i. In real systems, the relationship between alarms and event hypotheses are determined by the topology of the system.

3) Expected Reported Alarms: The hypothesis, following a disturbance event EH_k, is assumed that would generate a series of alarms that are reported to a control center, such as the PR and CB alarms that are expected. However, a minority of expected alarms may “slip through the cracks” due to unknown reasons, resulting those alarms to be missed, delayed, or tampered with by deceptive users. We denote $\mathbf{\Omega}^{(ERA)}_{EH_k}$ as a set of all expected reported alarms, i.e., expected PR and CB alarms that corresponds to event hypotheses EH_k.

B. Protective Relay Schemes and Circuit Breaker Tripping

Relational schemes of protective relays can be proprietary. In this model, we generalize three main protective schemes that are associated with the tripping of circuit breakers:

- Main Protective Relays ($MPRs$): This scheme is set up to be responsible for detecting and isolating failures on the corresponding devices. During fault occurrence, the schemes $MPRs$ are supposed to act quickly, and if this scheme performs correctly, a tripping command will be sent to the corresponding circuit breakers.
- Backup Protective Relays ($BPRs$): If the $MPRs$ cannot act promptly, BPR installed on devices are expected to act. Similarly, a tripping command will be sent to the circuit breakers that are associated with this scheme.
- Breaker Failure Protections ($BFPs$): This scheme determines if the designed breakers are out of service. If not, the scheme will react when there is a malfunction of a circuit breaker.
- Circuit Breakers (CBs): The circuit breakers associated with the aforementioned three schemes will trip and isolate the transmission circuits electrically upon receiving the trip commands from the protective relays.

These are the alarms from the schemes of $MPRs$, $BPRs$, $BFPs$ and CB tripping considered in the modeling for the CyIS module as the input of PR and CB.

C. Logical Relations of PR Operation and CB Tripping

The logical relations of PR operation and CB tripping are important to determine the anomaly. This section first analyzes scenarios without cyberattacks under an uncertainty of received alarms in the control center. This is to distinguish the baseline of potential abnormal condition. In the next section, hypothetical cyberattack scenarios are elaborated, a distinct irregularity would be sufficiently conclusive to be inferred as anomalous. Below is an example to show the formulation of modeling under one time window with reported alarms at the control center. The substation topology considered in this example is shown in Fig. 3.

1) The main PR of L_1 operates (RA_1);
2) The backup PR of L_1 operates (RA_2);
3) C_5 trips (RA_3);
4) C_4 trips (RA_4).

Based on reported alarms that are successfully received at the control center, the relations among event hypotheses and expected alarms is represented in Fig. 4. The relational diagram denotes that different events can result in different expected reported alarms. For example, under ideal condition when an event EH_1 occurs, the expected alarms transmitted to control center should be RA_1 and RA_8.

In the relational diagram shown in Fig. 4, RA_5 denotes that the backup PR of the device for T_1 to operate, RA_6 represents that the main PR of the backup device for T_1 to operate, RA_7 denotes for the main PR of L_1 if it fails to react, and RA_8 represents that C_1 to trip. These four alarms are possibly generated that may not be reported to control center. The detailed event of hypotheses are shown in Table I.

The integer constraints without AaH consideration are formulated that represent the relations between possible events and alarms. An evaluation index represents the matching degree between received reported alarms as well as expected reported alarms inferred by event hypotheses. Based on the relations in Fig. 4, a set of $\mathbf{\Omega}^{(RA)}$, $\mathbf{\Omega}^{(EH)}_{RA_i}$, and $\mathbf{\Omega}^{(ERA)}_{EH_k}$, defined previously, is defined as follows:

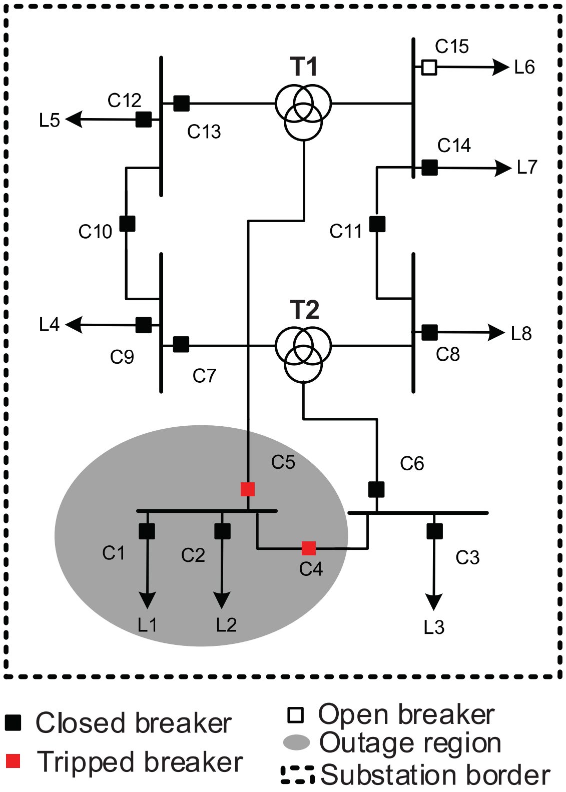


Fig. 3. Functional Diagram of a Substation with Detailed Topology.

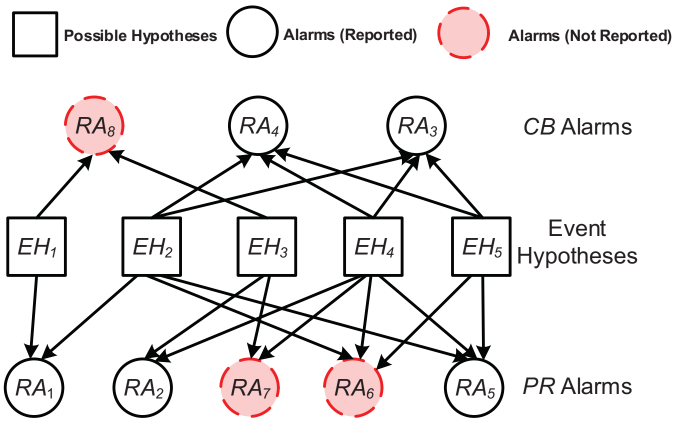


Fig. 4. Logical Relations between Event Hypotheses and Alarms.

TABLE I
EVENT HYPOTHESES

EH No.	Content
EH_1	A fault on L_1; Main PR of L_1 act; C_1 trips.
EH_2	A fault on L_1; Main PR of L_1 acts; C_1 fails to trip.
EH_3	A fault on L_1; Main PR of L_1 fails to act; Backup PR of L_1 act, C_1 trips.
EH_4	A fault on L_1; Main PR of L_1 fails to act; Backup PR of L_1 act, but C_1 fails to trip.
EH_5	A fault on B_1.

1) Definitions of Variables and Sets: Based on received reported alarms, the set of all possible events can be obtained, as shown in (1).

$$\mathbf{\Omega}^{(EH)} = \left(\bigcup_{RA_i \in \mathbf{\Omega}^{(RA)}} \mathbf{\Omega}^{(EH)}_{RA_i} \right) \tag{1}$$

where $\mathbf{\Omega}^{(RA)}$ is a set of received alarms, $\mathbf{\Omega}^{(EH)}_{RA_i}$ is a set of event hypotheses with respect to the received alarm RA_i. A set of all expected reported alarms, including PR and CB alarms, is expressed as

$$\mathbf{\Omega}^{(ERA)} = \bigcup_{EH_k \in \mathbf{\Omega}^{(EH)}} \mathbf{\Omega}^{(ERA)}_{EH_k} \tag{2}$$

Decision Variables: Each $EH_k \in \mathbf{\Omega}^{(EH)}$ corresponds to a decision variable X_{EH_k} is to determine which event has happened. It is a binary value can be either 1 or 0. The value 1 denotes the corresponding event that has happened, and the value 0 represents the event has not happened.

State Variables: As each possible event can cause by different PR and CB alarms, the matching degree between received reported alarms and expected alarms caused is an index to evaluate each event hypothesis. Therefore, each alarm $RA_i \in \mathbf{\Omega}^{(ERA)}$ corresponds to a state variable X_{RA_i}. Its value can be either 1 or 0. The value 1 denotes this alarm should be existent, and the value 0 indicates otherwise.

2) Constraints of Alarm Events: Based on power system protective schemes, one event hypothesis can generate several PR and CB alarms. Even though parts of alarms might be not reported to system operators. These missed, delayed, or tampered alarms are the relations between one event and its associated alarms are determined as follows:

$$\begin{aligned} X_{EH_k} \cdot \prod_{RA_i \in \mathbf{\Omega}^{(ERA)}_{EH_k}} X_{RA_i} \cdot \prod_{RA_i \notin \mathbf{\Omega}^{(ERA)}_{EH_k}} (1 - X_{RA_i}) \\ + (1 - X_{EH_k}) = 1, \quad EH_k \in \mathbf{\Omega}^{(EH)} \end{aligned} \tag{3}$$

The first part of the left side of the equation ensures the coincidence of each event and its corresponding alarms, and the second part of the left side of the equation guarantees that the corresponding alarms may be also generated by other events. It means that $X_{RA_i} \in \mathbf{\Omega}^{(ERA)}_{EH_k}$ can be 0 or 1 when $X_{EH_k} = 0$. Take EH_4 as an example. The constraint can be written as

$$\begin{aligned} X_{EH_4} \cdot X_{RA_2} \cdots X_{RA_7} \cdot (1 - X_{RA_1}) \cdot (1 - X_{RA_8}) \\ + (1 - X_{EH_4}) = 1 \end{aligned} \tag{4}$$

3) Constraints of Different Events: Based on a series of related alarms, event hypotheses are exclusive when they are assumed as follows:

$$\sum_{EH_k \in \mathbf{\Omega}^{(EH)}} X_{EH_k} = 1 \tag{5}$$

Take EH_1, EH_2, EH_3, EH_4 and EH_5 for example,

$$X_{EH_1} + X_{EH_2} + X_{EH_3} + X_{EH_4} + X_{EH_5} = 1 \tag{6}$$

4) Model Objective: The matching degree serves as the objective to optimize the problem. The matching degree determines a measure of the degree to which expected reported alarms caused by an event hypothesis agree with the reported alarms received at the control center. The objective function is represented as follows:

$$\max \quad IN = IN_1 + IN_2 + IN_3 + IN_4 \tag{7}$$

The first part IN_1 is

$$IN_1 = \frac{\sum\limits_{RA_i \in \Omega^{(RA)}} X_{RA_i}}{N} \tag{8}$$

where the denominator N is the number of received reported alarms to system operators. For each event hypothesis, it results in its own expected reported alarms. Some overlap with other received alarms. The numerator denotes the number of overlapping alarms. For received reported alarms in the example, IN_1 of EH_1 and EH_4 are $1/4$ and $3/4$ respectively.

The second part IN_2 can be represented as:

$$IN_2 = 1 - \frac{\sum\limits_{EH_k \in \Omega^{(EH)}} \left(X_{EH_k} \cdot \sum\limits_{RA_i \in \Omega^{(ERA)}_{EH_k}} \left(X_{RA_i} - X^{(0)}_{RA_i} \right) \right)}{N} \tag{9}$$

where $X^{(0)}_{RA_i}$ is the original value of the alarm RA_i. If the alarm RA_i is in the set of received reported alarms, the value of $X^{(0)}_{RA_i}$ is 1; if RA_i is in the set of expected reported alarms but not in the set of received reported alarms, the value of $X^{(0)}_{RA_i}$ is 0. For example, $X^{(0)}_{RA_1} = 1$ and $X^{(0)}_{RA_8} = 0$. For a certain event hypothesis, some of its expected reported alarms might not be in received reported alarms. The numerator in (9) denotes the number of those expected reported alarms that are not in received reported alarms. For example, IN_2 of EH_1 and EH_4 are $3/4$ and $1/4$ respectively.

The third part IN_3 can be written as:

$$IN_3 = \frac{\sum\limits_{RA_t \in \Omega^{(RA)}} \left(X_{RA_t} \right)}{\sum\limits_{EH_k \in \Omega^{(EH)}} \left(X_{EH_k} \cdot \sum\limits_{RA_t \in \Omega^{(ERA)}_{EH_k}} X_{RA_i} \right)} \tag{10}$$

where the denominator denotes the number of expected reported alarms caused by a certain event hypothesis. Even though the expression of the denominator is the sum of expected reported alarms for all $EH_k \in \mathbf{\Omega}^{(EH)}$, the exclusive between different event hypotheses guarantees that the denominator is just the number of expected reported alarms caused by one event. For example, IN_3 of EH_1 and EH_4 are $1/2$ and $3/6$ respectively.

The fourth part IN_4 can be expressed as:

$$IN_4 = 1 - \frac{\sum\limits_{EH_k \in \Omega^{(EH)}} \left(X_{EH_k} \cdot \sum\limits_{RA_i \in \Omega^{(RA)}} \left(X^{(0)}_{RA_i} - X_{RA_i} \right) \right)}{\sum\limits_{EH_k \in \Omega^{(EH)}} \left(X_{EH_k} \cdot \sum\limits_{RA_t \in \Omega^{(ERA)}_{EH_k}} X_{RA_i} \right)} \tag{11}$$

where the denominator is identical as in (10). Some received reported alarms though may not in the expected reported alarms caused by a certain event hypothesis. The numerator denotes the number of those reported alarms that are not in the expected reported alarms. For example, the values of IN_4 of EH_1 and EH_4 are $-1/2$ and $5/6$ respectively. With constraints, this can be optimized using integer programming.

IV. Model Considering AaH Attacks and the Uncertainties

This section extends the modeling from previous the section with a consideration of AaH attack. The influences of potential AaH attack on tampered alarms are analyzed with its randomness and possible scenario reduction.

A. Influences of Cyberattacks on the Alarms

Consider a circumstance where attackers have successfully hacked into a substation network, and have gained administrative privilege to perform unauthorized operations, such as, open circuit breakers. If these generated alarms related with unauthorized operations are send to the control center, it can be inferred that the system is under a cyberattack. For a sophisticated attack like AaH, intruders might tamper with some alarms, e.g., delete some existent alarms and add some other non-existent alarms, to hide, suppress, or modify certain measurements. To maximize their attack outcomes, covertly planning for a cyberattack is necessary in order for attackers to understand what functions are available on the local substation SCADA system. This is a critical step that intruders have the capacity to open all CBs that are inferred in the logical relations. Because PRs are designed to respond to faults, it is believed that unauthorized CB tripping will not cause PR actions. Therefore, there are only CB alarms if reported alarms are correct when a system is possibly under attack. For example, under a perfect situation, alarms should include RA_3, RA_4 and RA_8 if there is an anomaly in the given case. However, the received alarms are actually RA_1, RA_2, RA_3 and RA_4. In this case, if the system is subject to cyberattacks, the intruders should delete RA_8 and add RA_1 and RA_2. Considering characteristics of the communication system, intruders can tamper alarms with certain probabilities. Fig. 5 shows the relational rules under a cyberattack scenario.

The probability that received reported alarms under an attack scenario can be represented as

$$P_{EH_{CA}} = \prod_{PA_i \in \mathbf{\Omega}^{(PA)}} P^{(D)}_{PA_i} \cdot \prod_{CA_i \in \overline{\mathbf{\Omega}^{(CA)}} \bigcap \mathbf{\Omega}^{(ECA)}} P^{(D)}_{CA_i} \tag{12}$$

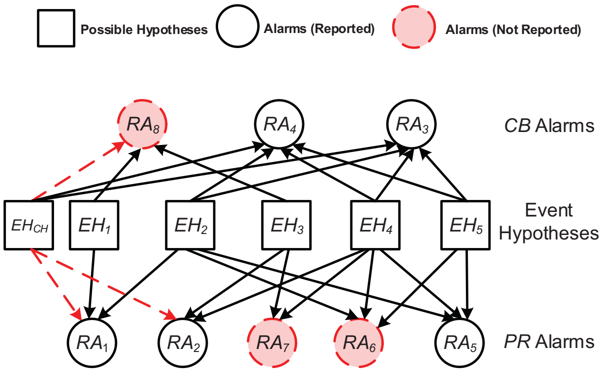


Fig. 5. Omission of Basic Relations under Attack.

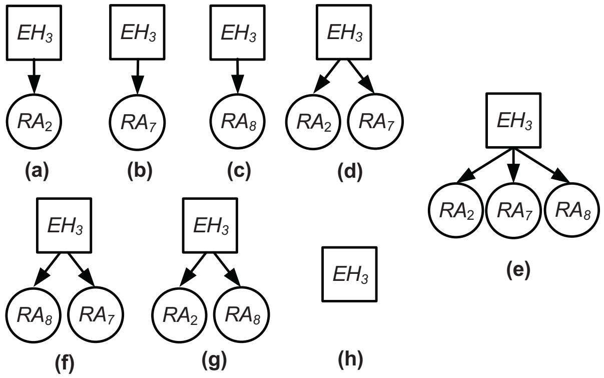


Fig. 6. Multiple Scenarios for One Event Hypothesis.

where $P_{PA_i}^{(D)}$ is the probability of adding an extra PA_i alarm, $P_{CA_i}^{(D)}$ is the probability of deleting an existent CA_i alarm. $\boldsymbol{\Omega}^{(PA)}$ and $\boldsymbol{\Omega}^{(CA)}$ are the sets of received reported PR and CB alarms respectively, and $\boldsymbol{\Omega}^{(PA)} \bigcup \boldsymbol{\Omega}^{(CA)} = \boldsymbol{\Omega}^{(RA)}$. $\boldsymbol{\Omega}^{(ECA)}$ are sets of expected reported CB alarms. $\overline{\boldsymbol{\Omega}^{(CA)}} \bigcap \boldsymbol{\Omega}^{(ECA)}$ is the set of possible unreceived CA alarms. For example, this set is $\{CA_8\}$ in Fig. 5.

Considering influences of cyberattacks, the constraint (5) and the matching degree index with cyberattacks (ICA) is revised as follows:

$$\max \quad ICA = [IN]_{CA} \cdot (1 + P_{EH_{CA}} \cdot X_{EH_{CA}} - X_{EH_{CA}}) \tag{13}$$

$$\sum_{EH_k \in \boldsymbol{\Omega}^{(EH)} \bigcup \{EH_{CA}\}} X_{EH_k} = 1 \tag{14}$$

where $X_{EH_{CA}}$ denotes the decision variable of a cyberattack incident. The value 1 means there is an attack incident, 0 is otherwise. $[IN]_{CA}$ is the matching degree with considering common event hypotheses and cyberattacks. The calculation is similar with (8), (9), (10), and (11). The only difference is $EH_k \in \boldsymbol{\Omega}^{(EH)} \bigcup \{EH_{CA}\}$.

B. Stochastic Issues of Relations among Events and Alarms

In real systems, some alarms might be delayed and missed in communication. Therefore, system operators may receive imperfectly reported alarms. Take EH_3, RA_2, RA_7 and RA_8 for example. Without considering unstable communications, the scenario that includes alarms RA_2, RA_7, and RA_8, as shown in Fig. 6, is reasonable. However, when considering delayed or missed alarms, the received reported alarms of other scenarios in Fig. 6 are also reasonable. These are the general scenarios that exist with certain probabilities and the sum of probabilities of all scenarios using (15) is 1, shown in Fig. 6.

$$\sum_{AS_j \in \boldsymbol{\Omega}_{EH_k}^{(AS)}} P_{EH_k}^{AS_j} = 1, \quad EH_k \in \boldsymbol{\Omega}^{EH} \tag{15}$$

where $\boldsymbol{\Omega}_{EH_k}^{(AS)}$ is the set of all reasonable alarm scenarios corresponding to EH_k, and $P_{EH_k}^{(AS_j)}$ is the probability of the alarm scenario to EH_k.

Fig. 4 shows the scenario of reasonable alarm rules without delayed and missed alarms. If considering delayed and missed alarms, there are multiple scenarios of reasonable alarm rules. The number of scenarios is the number of combinations of possible alarm scenarios of each event hypothesis. The set of possible alarm scenarios $\boldsymbol{\Omega}_{EH}^{(AS)}$ can be represented as follows.

$$\boldsymbol{\Omega}_{EH}^{(AS)} = \boldsymbol{\Omega}_{EH_1}^{(AS)} \times \boldsymbol{\Omega}_{EH_2}^{(AS)} \times \cdots \times \boldsymbol{\Omega}_{EH_k}^{(AS)} \times \cdots, \quad EH_k \in \boldsymbol{\Omega}^{(EH)} \tag{16}$$

where × denotes Cartesian product. The probability of each scenario, i.e., one element in $\boldsymbol{\Omega}_{EH}^{(AS)}$, is notated by $P_{EH}^{(AS_j)}$, $AS_j \in \boldsymbol{\Omega}_{EH}^{(AS)}$. This probability is the product of corresponding probability $P_{EH_k}^{(AS_j)}$, $EH_k \in \boldsymbol{\Omega}^{(EH)}$. This problem now becomes a problem subject to multiple scenarios with certain probabilities.

1) Model Objective: The objective function considering AaH attacks is modified as follows:

$$\max \sum_{AS_j \in \boldsymbol{\Omega}_{EH}^{(AS)}} \left(P_{EH}^{(AS_j)} \cdot ICA|_{AS_j} \right) \tag{17}$$

where $ICA|_{AS_j}$ denotes the matching degree index subject to the scenario $AS_j \in \boldsymbol{\Omega}_{EH}^{(AS)}$. The formulation of this index is similar with (8), (9), (10), (11), and (13). For an example, variables such as X_{EH_k}, $X_{EH_{CA}}$, and X_{RA_i}, with only one scenario, can represent multiple scenarios with the superscript AS indicating all scenarios, i.e., $X_{EH_k}^{AS_j}$, $X_{EH_{CA}}^{AS_j}$, and $X_{RA_i}^{AS_j}$, $AS_j \in \boldsymbol{\Omega}_{EH}^{(AS)}$.

2) Constraints of Event Hypotheses: Based on (14), the constraints can be expressed as

$$\sum_{EH_k \in \boldsymbol{\Omega}^{(EH)}} X_{EH_k}^{AS_j} + X_{EH_{CA}}^{AS_j} = 1, AS_j \in \boldsymbol{\Omega}_{AS}^{(EH)} \tag{18}$$

The constraints denote those variables for each scenario should be exclusive.

3) Consistent of Decision Variables: Due to multiple scenarios, each variable in multiple scenarios are different. However, the decision variables in those scenarios should be consistent.

$$X_{EH_k}^{AS_l} = X_{EH_k}^{AS_j}, AS_l, AS_j \in \boldsymbol{\Omega}_{AS}^{(EH)}, EH_k \in \boldsymbol{\Omega}^{(EH)} \tag{19}$$

$$X_{EH_{CA}}^{AS_l} = X_{EH_{CA}}^{AS_j}, \quad AS_l, AS_j \in \boldsymbol{\Omega}_{AS}^{(EH)} \tag{20}$$

4) Constraints of Events and Alarms: The constraints in (3) should be expended to all scenarios.

$$X^{AS_j}_{EH_k} \cdot \prod_{RA_i \in \mathbf{\Omega}^{(ERA)}_{EH_k}|_{AS_j}} X^{AS_j}_{RA_i} \cdot \prod_{RA_i \notin \mathbf{\Omega}^{(ERA)}_{EH_k}|_{AS_j}} \left(1 - X^{AS_j}_{RA_i}\right) + \left(1 - X^{AS_j}_{EH_k}\right) = 1, EH_k \in \mathbf{\Omega}^{(EH)}, AS_j \in \mathbf{\Omega}^{(EH)}_{AS} \tag{21}$$

where $\mathbf{\Omega}^{(ERA)}_{EH_k}|_{AS_j}$ is the set of expected reported alarms of EH_k subject to the scenario $AS_j \in \mathbf{\Omega}^{(EH)}_{AS}$.

C. Scenario Reduction

The proposed model can be computationally intensive as the event hypotheses and expected incoming alarms increase. Scenario reduction is introduced to handle the complexity of computation.

1) Scenario Combination Reduction: Total scenario combination in (16) can guarantee accurate results; however, calculations may be huge. Although missed and delayed alarms exist in SCADA systems, the probabilities of them are very small. Based on this common view, the expected reported alarm scenario of each event hypothesis can be reduced. We set up the scenarios such that at most two alarms can be missed or delayed, and only one analyzed time window during the communication is considered.

2) Event Hypothesis Reduction: When there are many possible event hypotheses inferred based on received reported alarms, we first sort the event hypotheses by the matching degree index without considering stochastic issues. Then, several event hypotheses with larger index values are chosen to consider stochastic issues.

V. Simulation Results

This section has four examples to demonstrate the validation feasibility of the proposed cyberattack inference system by synthesizing all substation alarm-related events.

A. The First Scenario

The example in the Section III is employed as the first scenario. RA_1, RA_2, RA_3 and RA_4 are received reported alarms. Based on these four reported alarms, five event hypotheses, i.e., EH_1, EH_2, EH_3, EH_4 and EH_5, and one cyberattack hypothesis are assumed. Fig. 7 (a) shows the inferential results with probabilities of missed and delayed alarms and probabilities of tampered alarms caused by cyberattacks. Fig. 7 (b) shows index values of several event hypotheses. One event hypothesis with a high index value is the most possible event that has happened.

B. The Second Scenario with Approximate Index Values

In real systems, some events might have similar index values which increase the difficulty of identifying them when missed alarms and delayed alarms exist. This scenario shows that the proposed inference system can identify them. Fig. 8 shows the partial changed topology in Fig. 3.

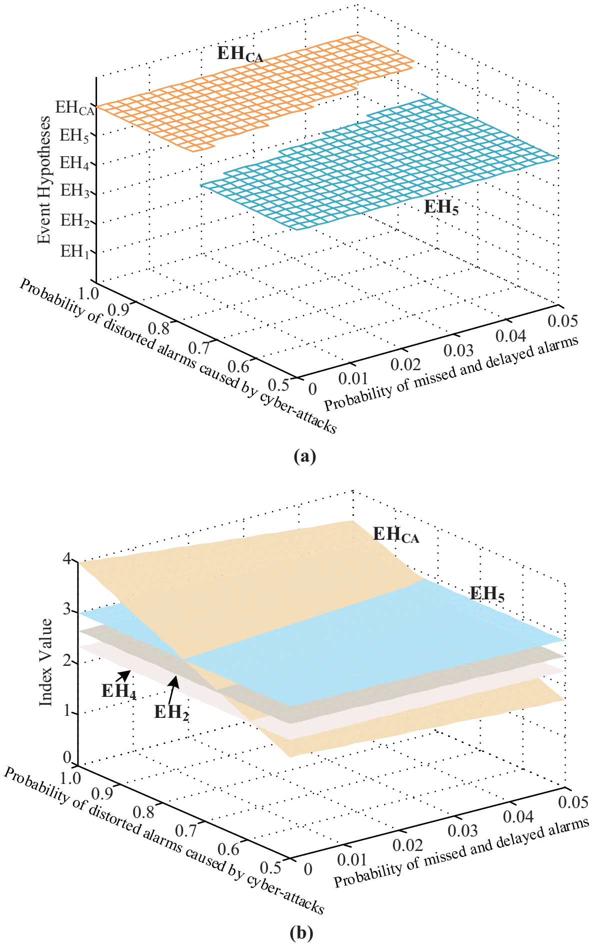


Fig. 7. (a) Inferential Results based on Alarms RA_1–RA_4 from One Substation. (b) Index Values of Event Hypothesis based on Alarms RA_1–RA_4 with Different Probabilities of Distorted, Missed, or Delayed Alarms.

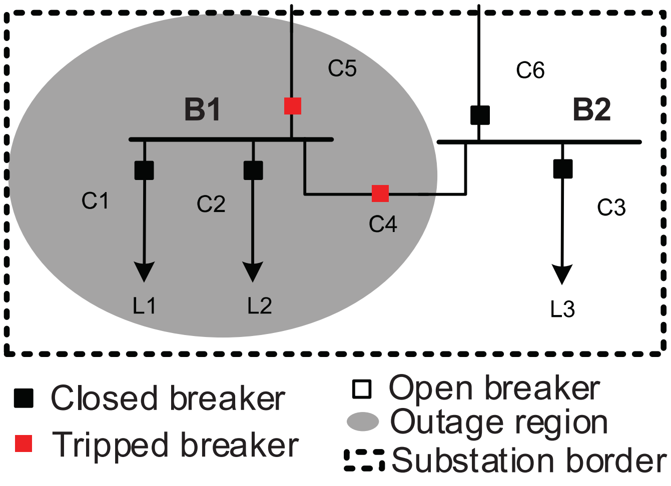


Fig. 8. Partial System Topology (Case 2).

In this new scenario, the circuit break C_2 is closed. Received reported alarms include RA_1, RA_2, RA_3, RA_6, RA_7, RA_8,

RA_{10} and RA_{12}. RA_1 to RA_8 denote the same meanings in the first case. RA_{10} denotes that C_2 trips and RA_{12} represents that the backup PR of L_2 operates. The logical relations are shown in Fig. 9. EH_1 to EH_5 denotes the same event hypotheses as in the first case. Table II shows other event hypotheses and expected reported alarms.

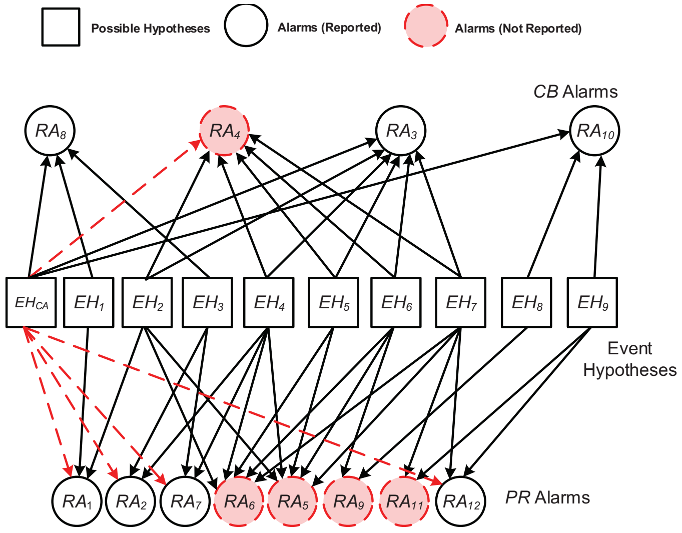


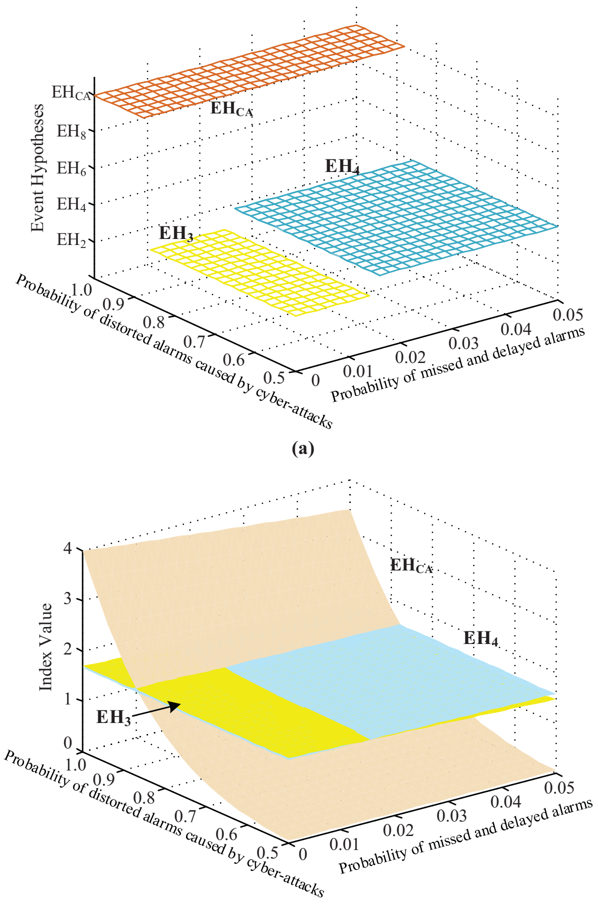
Fig. 9. Logical Relations of Case Two.

TABLE II
EVENT HYPOTHESES AND ALARMS

EH No.	Contents
EH_6	A fault on L_2; Main PR of L_2 acts; C_2 fails to trip.
EH_7	A fault on L_2; Main PR of L_2 fails to act; Backup PR of L_2 act, but C_2 fails to trip.
EH_8	A fault on L_2; Main PR of L_2 act; C_2 trips.
EH_9	A fault on L_2; Main PR of L_2 fails to act; Backup PR of L_2 act, C_2 trips.
RA_9	The main PR of L_2 operates.
RA_{11}	The main PR of L_2 fails to act.

Figs. 10 (a) and (b) show the inferential results and index values of corresponding event hypotheses in Fig. 9. According to the results, with different probabilities of missed and delayed alarms, there can vary from incident to incident based on received alarms.

Figs. 11 (a) and (b) show the inferential results and index values subject to the scenario with different received reported alarms that are RA_4, RA_7, RA_8, RA_9 and RA_{10}. For example, when the probability of missed and delayed alarms (MDA) is 0.005 and the probability of tampered alarms caused by cyberattacks (CA) is 0.6 or much less, the event is EH_8. When the probabilities of MDA and CA are 0.03 and 0.7, the event is EH_3. When the probability of CA is 0.8 or much larger, this can infer a possible AaH attack.



(a)

(b)

Fig. 10. (a) Inferential Results based on Alarms RA_1–RA_3, RA_6–RA_8, RA_{10}, RA_{12} from One Substation. (b) Index Values of Event Hypothesis based on Alarms RA_1–RA_3, RA_6–RA_8, $RA_{1}0$, $RA_{1}2$ with Different Probabilities of Distorted, Missed, or Delayed Alarms.

C. The Third Scenario with Possible Multiple incidents

For the two above scenarios, we do not consider multiple incidents that are happened at the same time during the analyzed time window. In the real system, multiple incidents may occur during the analyzed time window. Fig. 12 shows the topology. Reported alarms include RA_3, RA_9, RA_{10}, RA_{13}, RA_{14} and RA_{15}. RA_{13} denotes the main PR of L_3 operates; RA_{14} and RA_{14} represent that the main and backup device for T_2 operates respectively. Besides EH_1, EH_2, EH_3, EH_4 and EH_5, event hypotheses EH_{10}, EH_{11}, EH_{12}, EH_{13} and EH_{14} can be possible based on the reported alarms.

For above given event hypotheses, some of them can be happened simultaneously during one analyzed time window and cause similar reported alarms. For example, $EH_1 \cdot EH_{14}$, $EH_1 \cdot EH_{10}$, $EH_1 \cdot EH_{11}$, $EH_1 \cdot EH_{12}$, $EH_1 \cdot EH_{13}$, $\cdots$ can also work as event hypotheses. The "$\cdot$" denotes two events can occur at the same time during the analyzed time window. Fig. 13 shows the inferential results and index values of corresponding event hypotheses.

D. Case Study With Multiple Substations

We further extend the case study with multiple substations. This is a realistic setup extracted from a utility configuration.

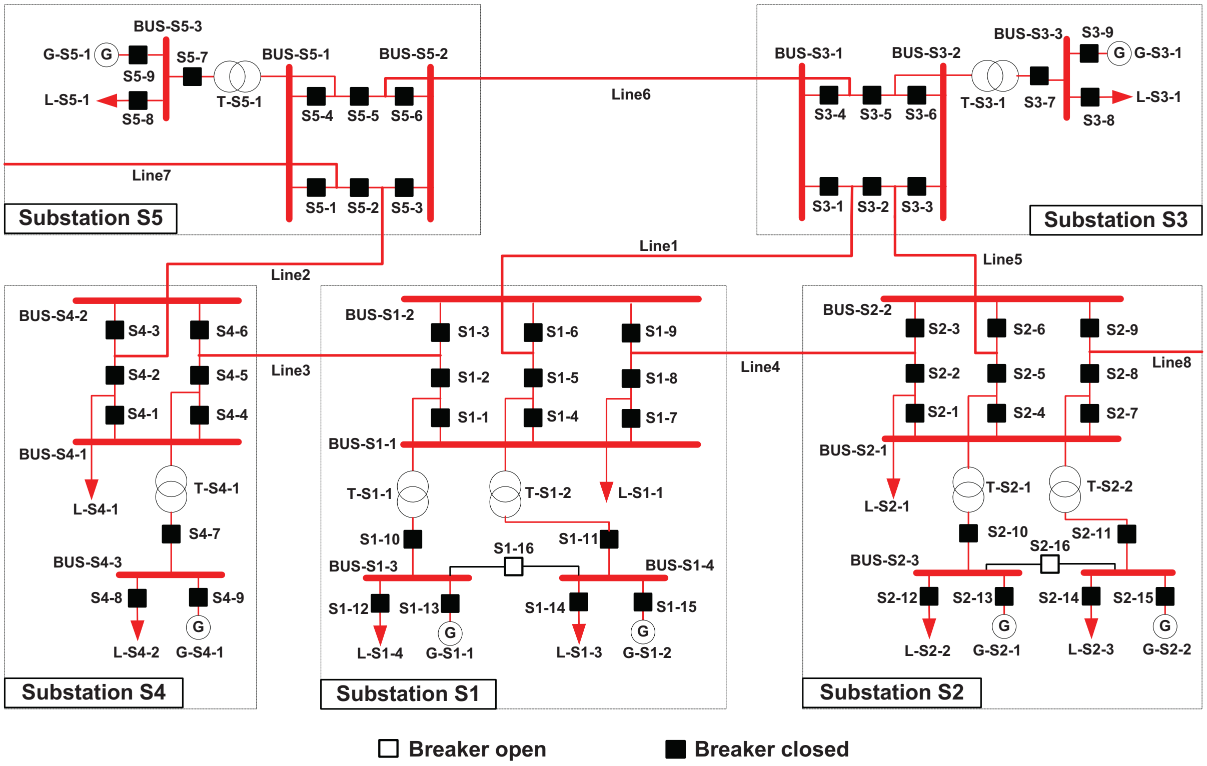


Fig. 14. Setup of Five Realistic Substations in Breaker-and-a-Half Configurations.

TABLE III
EVENT HYPOTHESES

EH **No.**	**Content**
EH_{10}	A fault on L_3; Main PR of L_3 act; $C3$ trips.
EH_{11}	A fault on L_3; Main PR of L_3 acts; $C3$ fails to trip.
EH_{12}	A fault on L_3; Main PR of L_3 fails to act; Backup PR of L_3 act, C_3 trips.
EH_{13}	A fault on L_3; Main PR of L_3 fails to act; Backup PR of L_3 act, but C_3 fails to trip.
EH_{14}	A fault on B_3.

The system has five substations configured in breaker-and-a-half schemes shown in Fig. 14. Statistically, the probability of missed/delayed alarms is approximately 0.05. This study is divided into 3 catagories: (1) Fault occurrence (without AaH attacks), (2) AaH attacks (without fault occurrence), and (3) AaH attacks during fault conditions.

1) Fault Occurrence (without AaH attacks): Due to the lengthy list of generated alarms during a fault situation, a relevant snapshot of the list for this study is shown in Table IV. Generated alarms include the statuses of breakers, protective relays, and other analog measurements such as power flow information between lines and breakers. The setup of two independent fault events are included where one is at fault on the bus BUS-S1-2 where the backup protective relay reacted to the disturbance. The other fault occurs on Line 6 as well as the responses from associated backup protective relay. Fig. 15(a) shows the index values regarding fault event hypotheses, and Fig. 15(b) shows the index values corresponding to AaH attack scenarios with a variation of the probabilities for distorted alarms. It can be observed from Fig. 15(a) that the largest index value of the fault event hypotheses is 3.01, which is mapped to the same value in Fig. 15(b) when the probability of distorted alarms caused by cyberattacks is larger than 0.98. In operational reality, it is unlikely that a probability of distorted alarms larger than 0.98 will be concluded to be an AaH attack. This infers the likelihood of fault occurrence even with a largest index value. This is a study without AaH attack consideration but the next two cases will involve such attacks.

2) AaH Attacks (Without Fault Occurrence): Contrary to the last case setup, this case is revised consistently to the reported alarm events with major modifications of telemetered measurements from substations. Similarly, Fig. 16(a) shows the index values regarding fault event hypotheses, and Fig. 16(b) corresponds to the index values of AaH attacks with a variation of the probabilities for distorted alarms. Using the proposed method, the largest index value of the fault event hypotheses is 1.02, which was reached when the probability of distorted alarms caused by the attack was larger than 0.67. Considering small index values of the fault event hypotheses and rapidly increasing trend of index values of AaH attacks, it can be inferred that the substation S3 is under attack.

3) AaH Attacks During Fault Conditions: This case is set up to assess concurrent event occurrences, both intentionally and unintentionally. Fig. 17(a) shows the index values regarding fault event hypotheses with another different group of

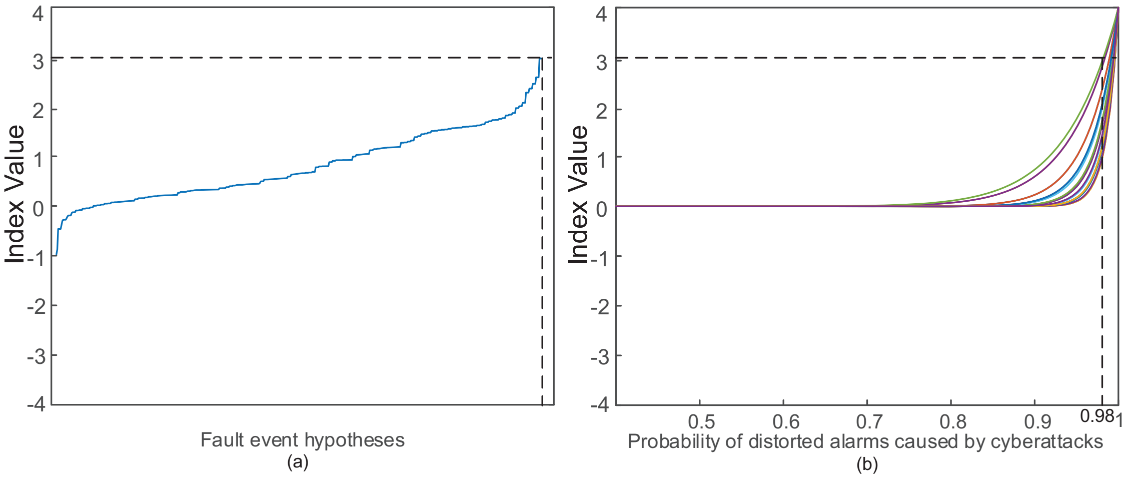


Fig. 15. Simulation Results of Fault Occurrence (Without AaH Attacks).

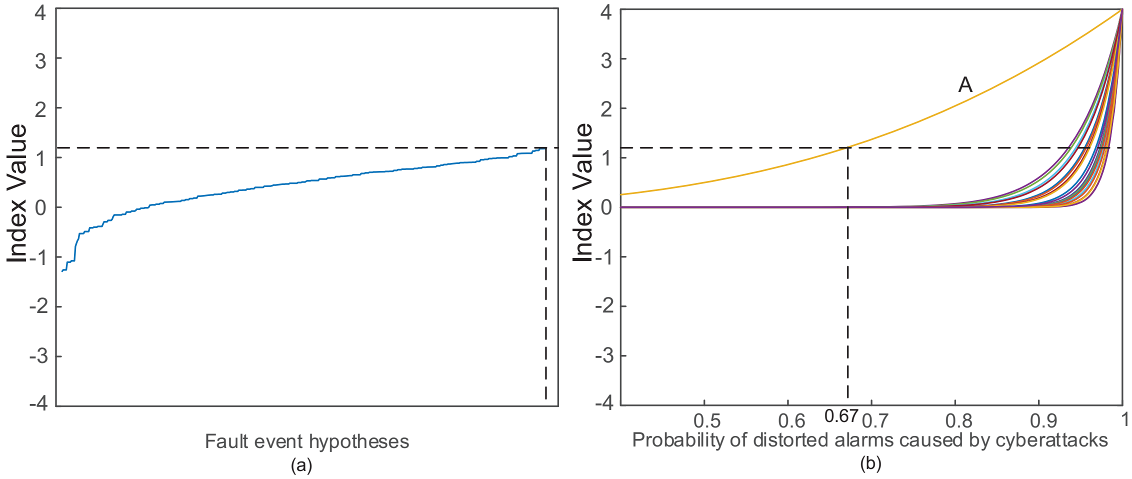


Fig. 16. Simulation Results of AaH Attacks (Without Fault Occurrence).

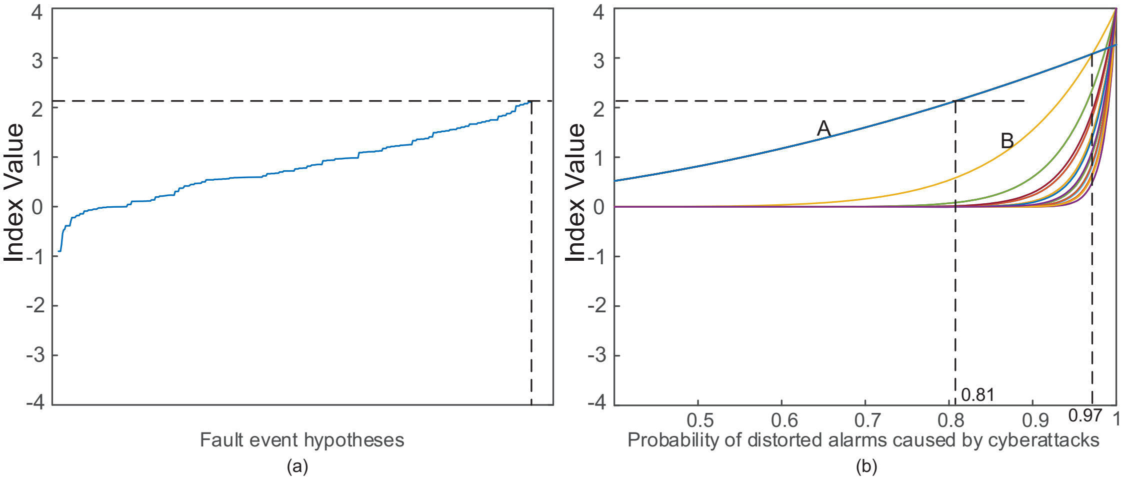
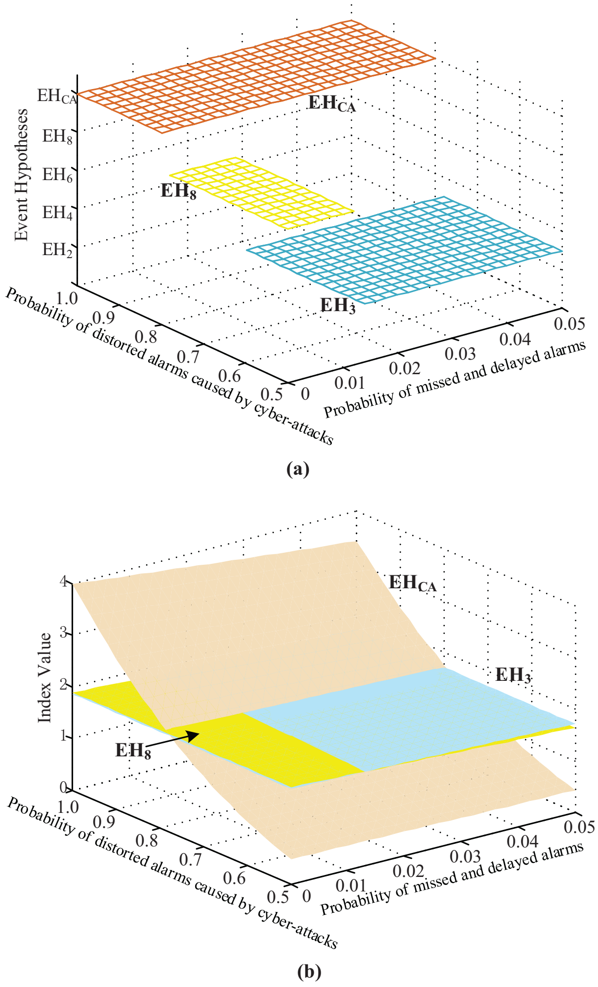


Fig. 17. Simulation Results of AaH Attacks During Fault Conditions.



(a)

(b)

Fig. 11. (a) Inferential Results based on Alarms RA_4, RA_7-RA_{10} from One Substation. (b) Index Values of Event Hypothesis based on Alarms RA_4, RA_7-RA_{10} with Different Probabilities of Distorted, Missed, or Delayed Alarms.

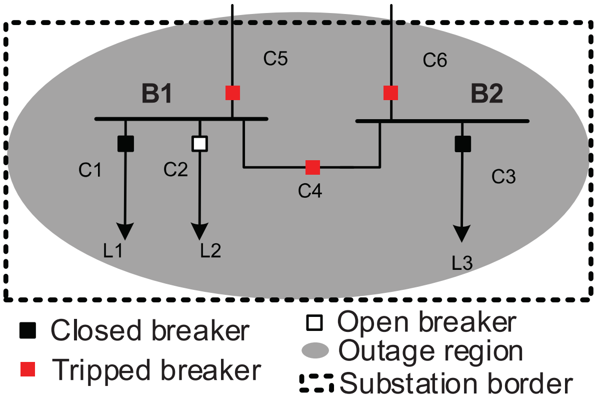
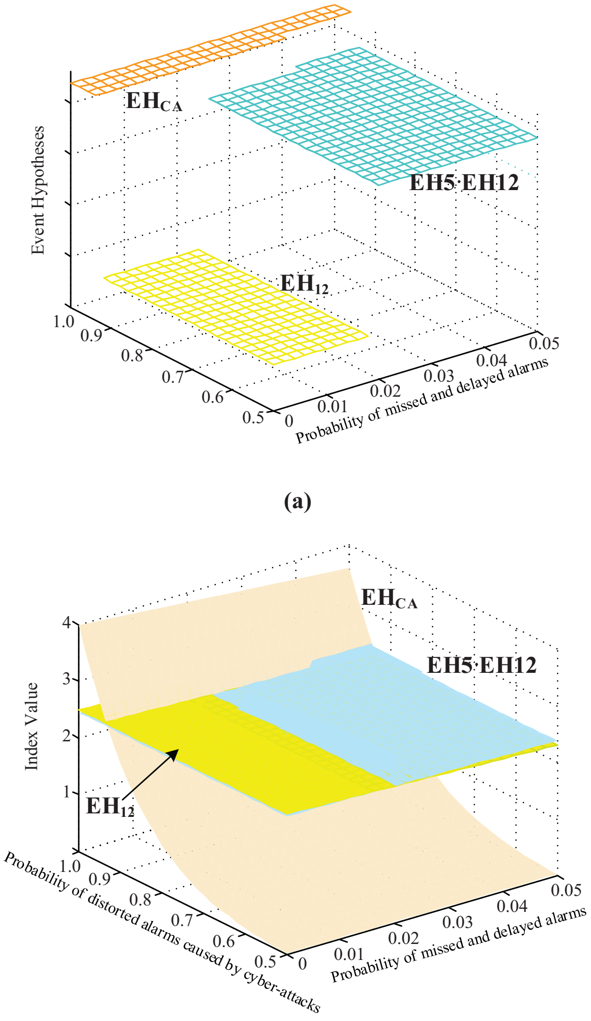


Fig. 12. Partial System Topology (Case Three).



(a)

(b)

Fig. 13. (a) Inferential Results based on Alarms RA_3, RA_9, RA_{10}, RA_{13}–RA_{15} from One Substation. (b) Index Values of Event Hypothesis based on Alarms RA_3, RA_9, RA_{10}, RA_{13}–RA_{15} with Different Probabilities of Distorted, Missed, or Delayed Alarms.

revised alarms. Fig. 17(b) shows the index values of different AaH attacks with a variation of the probabilities for distorted alarms. The disturbance event with the largest index value shown in Fig. 17(a) has a fault on Line 6. As depicted in Fig. 17(b), the curve A corresponds to the AaH attack on the substation S3 with fault on Line 6. With the consideration of its rapidly increasing trend of index values, substation S3 can be observed with a higher probability of an AaH attack. The curve B corresponds to AaH attacks on the substation S3. The probability corresponds to the intersection point of the curve A and B is about 0.97. This implies intentional anomaly with the possibility of distorted alarms that is larger than 0.97. In this case, this can be concluded that the substation S3 is under an AaH attack.

VI. Conclusion and Future Work

The proposed method is developed to synthesize substation alarm-related events for verification of AaH attacks. The inference system provides an approach to identify cyberattacks or other anomalous incidents. In the paper, it is assumed that successful cyber intrusions can execute disruptive switching actions associated with the substations, which are in the logical relations inferred by received reported alarms. Intruders may tamper with generated alarms and send fake alarms to confuse

TABLE IV
SNAPSHOT OF PARTIAL ALARMS RECEIVED AT CONTROL CENTER

No.	Received Alarms
1	S1-9 trips
2	Line 1 main protection fail to act
3	BUS-S1-2 differential protection device abnormal
4	BUS-S3-1 differential protection act
5	BUS-S1-2 differential protection CT abnormal
6	S1-3 trips
7	S1-9 reclosing
8	S1-3 over current protection
9	S1-6 over current protection
10	Power flow through S3-5: Q=0
11	Power flow through S3-5: P=0
12	S3-4 trips
13	S3-2 trips
14	S3-3 trips
15	Power flow through S1-6: P=0
16	Power flow through S1-3: Q=0
17	Power flow through S1-9: Q=0
18	Lin5 main protection fail to act
19	S3-3 reclosing
20	BUS-S5-2 differential protection act
21	S5-5 trips
22	Line 6 main protection communication broken
23	S5-6 trips
24	Power flow through S5-3: P=0
…	……

operators at a control center. For missed alarms and delayed alarms during the communication, stochastic problems are incorporated in the model. The entire problem is formulated as an integer programming problem with multiple scenarios, and scenario reduction is employed to handle increasingly larger event hypotheses and received reported alarms. With the cases introduced in the simulation study, the test systems show that the proposed inference system demonstrates systematic identification of potential anomalous events in within substations. Future work includes verifying the proposed method in an online SCADA environment. There may require additional information as well as to infer the probabilities of substations under a control area to determine accurately the event occurrence of AaH attack, short circuit faults, or combination of both.

REFERENCES

[1] "Final report on 1965 blackout," July 1967. [Online]. Available: http://blackout.gmu.edu/archive/pdf/fpc_65.pdf

[2] F. F. Wu, K. Moslehi, and A. Bose, "Power system control centers: Past, present, and future," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1890–1908, Nov. 2005.

[3] W. R. Prince, B. F. Wollenberg, and D. B. Bertoignolli, "Survey on excessive alarms," *IEEE Trans. Power Syst.*, vol. 4, no. 3, pp. 950–956, Aug. 1989.

[4] B. F. Wollenberg, "Feasibility study for an energy management system intelligent alarm processor," *Power Engineering Review, IEEE*, vol. 6, no. 5, pp. 54–55, May. 1986.

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

[6] Z. A. Vale and A. M. 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.

[7] H.-J. Lee, B.-S. Ahn, and Y.-M. Park, "A fault diagnosis expert system for distribution substations," *IEEE Trans. Power Del.*, vol. 15, no. 1, pp. 92–97, Jan. 2000.

[8] H. J. Miao, M. Sforna, and C.-C. Liu, "A new logic-based alarm analyzer for on-line operational environment," *IEEE Trans. Power Syst.*, vol. 11, no. 3, pp. 1600–1606, Aug. 1996.

[9] L.Wei, W. Guo, F. Wen, G. Ledwich, Z. Liao, and J. Xin, "An online intelligent alarm-processing system for digital substations," *IEEE Trans. Power Del.*, vol. 26, no. 3, pp. 1615–1624, Jul. 2011.

[10] W. X. Guo, F. S. Wen, Z. W. Liao, L. H. Wei, and J. Xin, "An analytic model-based approach for power system alarm processing employing temporal constraint network," *IEEE Trans. Power Del.*, vol. 25, no. 4, pp. 2435–2447, Oct. 2010.

[11] K. L. Lo, H. S. Ng, and J. Trecat, "Power systems fault diagnosis using petri nets," *Inst. Eng. Technol. Gen., Transm. Distrib.*, vol. 144, no. 3, pp. 231–236, May. 1997.

[12] X. Luo and M. Kezunovic, "Implementing fuzzy reasoning petri-nets for fault section estimation," *IEEE Trans. Power Del.*, vol. 23, no. 2, pp. 676–685, Apr. 2008.

[13] K. L. Lo, H. S. Ng, and J. Trecat, "Extended petri-net models for fault diagnosis for substation automation," *Inst. Eng. Technol. Gen., Transm. Distrib.*, vol. 146, no. 3, pp. 229–234, May. 1999.

[14] G. J. Cardoso, J. G. Rolim, and H. H. Zurn, "Application of neural network modules to electric power system fault section estimation," *IEEE Trans. Power Del.*, vol. 19, no. 3, pp. 1034–1041, Jul. 2004.

[15] A. P. A. da Silva, A. H. F. Insfran, P. M. da Silveira, and G. L. Torres, "Neural networks for fault location in substations," *IEEE Trans. Power Del.*, vol. 11, no. 1, pp. 234–239, Jan. 1996.

[16] G. Rigatos, P. Siano, and A. Piccolo, "Neural network-based approach for early detection of cascading events in electric power systems," *Inst. Eng. Technol. Gen., Transm. Distrib.*, vol. 3, no. 7, pp. 650–665, Jul. 2009.

[17] C. L. Hor and P. A. Crossley, "Unsupervised event extraction within substations using rough classification," *IEEE Trans. Power Del.*, vol. 21, no. 4, pp. 1809–1816, Oct. 2006.

[18] "Supervisory control and data acquisition (scada) systems, national communications system, technical information bulletin 04-1," 2004. [Online]. Available: http://scadahacker.com/library/Documents/ICS_Basics/SCADA/20Basics/20/20NCS/20TIB/2004-1.pdf

[19] L. Pietre-Cambacedes, M. Tritschler, and G. N. Ericsson, "Cybersecurity myths on power control systems: 21 misconceptions and false beliefs," *IEEE Trans. Power Del.*, vol. 26, no. 1, pp. 161–172, Jan. 2011.

[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. Wei, Y. Lu, M. Jafari, P. M. Skare, and K. Rohde, "Protecting smart grid automation systems against cyberattacks," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 782–795, Dec. 2011.

[22] 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.

[23] Y. Zhang, L. Wang, W. Sun, R. C. Green, and M. Alam, "Distributed intrusion detection system in a multi-layer network architecture of smart grids," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 796–808, Dec. 2011.

[24] "Siemens: Stuxnet worm hit industrial systems," Sep. 2010. [Online]. Available: http://www.computerworld.com/article/2515570/network-security/siemens--stuxnet-worm-hit-industrial-systems.html

[25] N. Falliere, L. Murche, and E. Chien, "W32.stuxnet dossier," Feb. 2011. [Online]. Available: https://www.symantec.com/content/en/us/enterprise/media/security_response/whitepapers/w32_stuxnet_dossier.pdf

[26] "A declaration of cyber-war," Apr. 2011. [Online]. Available: http://www.vanityfair.com/news/2011/04/stuxnet-201104

[27] D. Veluz, "STUXNET malware targets SCADA systems," Oct. 2010. [Online]. Available: http://www.trendmicro.com/vinfo/us/threat-encyclopedia/web-attack/54/stuxnet-malware-targets-scada-systems

[28] O. Kosut, L. Jia, R. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Trans. Smart Grid*, vol. 2, no. 4, pp. 645–658, Dec. 2011.

[29] G. Hug and G. A. Giampapa, "Vulnerability assessment of ac state estimation with respect to false data injection cyber-attacks," *IEEE Trans. Smart Grid*, vol. 3, no. 3, pp. 1362–1370, Sep. 2012.

Chong Wang (S'10) received his B.E. (2009) and M.E. (2012) degrees from Hohai University, China. He is currently pursuing the Ph.D. degree in the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong SAR, China.

Chee-Wooi Ten (SM'11) received the BSEE and MSEE degrees from Iowa State University, Ames, in 1999 and 2001, respectively. He was an Application Engineer with Siemens Energy Management and Information System (SEMIS) in Singapore from 2002 to 2006. He later received the Ph.D. degree in 2009 from University College Dublin (UCD), National University of Ireland. He is currently an Assistant Professor at Michigan Technological University. His primary research interests are power infrastructure cybersecurity, interdependency modeling for critical cyberinfrastructures, and SCADA automation applications for power grid.

Yunhe Hou (M'08) received the B.E. and Ph.D. degrees from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2005, respectively. He was a Post-Doctoral Research Fellow at Tsinghua University, Beijing, China, from 2005 to 2007, and a Post-Doctoral Researcher at Iowa State University, Ames, IA, USA, and the University College Dublin, Dublin, Ireland, from 2008 to 2009. He was also a Visiting Scientist at the Laboratory for Information and Decision Systems, Massachusetts Institute of Technology, Cambridge, MA, USA, in 2010. Since 2011, he has a joint appointment of professorship with the State Key Laboratory of Advanced Electromagnetic Engineering and Technology, Huazhong University of Science and Technology. He joined the faculty of the University of Hong Kong, Hong Kong, in 2009, where he is currently an Assistant Professor with the Department of Electrical and Electronic Engineering.

Andrew Ginter is the Vice President of Industrial Security at Waterfall Security Solutions. He spent the first part of his career developing systems level and control system products for a number of vendors, including Honeywell and Hewlett-Packard. He led development of middleware products connecting industrial control systems at Agilent Technologies to the SAP enterprise resource planning systems. As Chief Technology Officer at Industrial Defender, Mr. Ginter led the development of the core industrial security product suite. Andrew currently represents Waterfall Security Solutions on standards bodies and works with customers to incorporate Waterfall Unidirectional Gateways into their industrial network designs. Andrew holds degree in Mathematics and Computer Science from the University of Calgary, as well as Industrial Security Professional (ISP), Information Technology Certified Professional (ITCP), and Certified Information Systems Security Professional (CISSP) accreditations.