World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:17, No:4, 2023

Stochastic Edge Based Anomaly Detection for Supervisory Control and Data Acquisitions Systems: Considering the Zambian Power Grid

Lukumba Phiri, Simon Tembo, Kumbuso Joshua Nyoni

***Abstract*—**In Zambia, recent initiatives by various power operators like ZESCO, CEC, and consumers like the mines, to upgrade power systems into smart grids, target an even tighter integration with information technologies to enable the integration of renewable energy sources, local and bulk generation, and demand response. Thus, for the reliable operation of smart grids, its information infrastructure must be secure and reliable in the face of both failures and cyberattacks. Due to the nature of the systems, ICS/SCADA cybersecurity and governance face additional challenges compared to the corporate networks, and critical systems may be left exposed. There exist control frameworks internationally such as the NIST framework, however, they are generic and do not meet the domain-specific needs of the SCADA systems. Zambia is also lagging in cybersecurity awareness and adoption, and therefore there is a concern about securing ICS controlling key infrastructure critical to the Zambian economy as there are few known facts about the true posture. In this paper, we present a stochastic Edged-based Anomaly Detection for SCADA systems (SEADS) framework for threat modeling and risk assessment. SEADS enables the calculation of steady-steady probabilities that are further applied to establish metrics like system availability, maintainability, and reliability.

***Keywords*—**Anomaly detection, SmartGrid, edge, maintainability, reliability, stochastic process.

I. Introduction

RENEWABLE energy sources like water, solar, wind, and biomass are used in Zambia together with fossil fuels like petroleum. Except for petroleum, which is entirely imported into the country, Zambia has the potential to be energy self-sufficient due to its significant untapped renewable resource reserves. Water still serves as Zambia's primary energy source despite the variety of these sources. According to estimates [1], Zambia has over 6,000 MW of hydropower potential, of which roughly 2,354 MW has been generated. It also has 40% of the SADC region's water resources. As of June 30, 2020, the total installed capacity of electricity in the country was 2,981.23 MW. In terms of installed capacity by technology, coal was second at 10.1%, followed by hydro generation at 80.5%. Further generation of heavy fuel oil (HFO) was at 3.7%; diesel and solar generation were at 2.8% and 3.0%, respectively. Over 2,800 MW of significant hydropower projects with current feasibility studies are located on Zambia's principal rivers. It would be wise to create optimal generating plans that are focused on hydropower because of this [1]-[3].

ZESCO (formerly called the Zambia Electricity Supply Corporation) dominates Zambia's energy industry. Electricity is produced, transmitted, distributed, and supplied to local and national markets by ZESCO, a vertically integrated national utility [4]. The Copperbelt Energy Corporation (CEC), a transmission business that buys high-voltage power from ZESCO and distributes it to the mining sector in the Copperbelt region, and the Lunsemfwa Hydro Power Company, are two more significant players [4]. Additionally, there are two rural concessions: Zengamina Hydro Power Company (ZHPC), which manages a remote rural network in the Northern Province, and North West Energy Corporation, which provides energy to a remote mining settlement that is not connected to the ZESCO system [4]. The Energy Regulation Board (ERB) is in charge of regulating the industry. Following the publication of Statutory Instrument No. 6 of 1997, the Energy Regulation Act (Commencement Order) on January 27, 1997 [4], the ERB was established under the Energy Regulation Act of 1995, Chapter 436 of the Laws of Zambia.

An essential component of operational technology is industrial control systems (ICS) [5]. Systems for monitoring and managing industrial operations are included. SCADA systems are industrial systems that record and analyze real-time data using control devices, network protocols, and graphical user interfaces. Hydropower facilities, telecommunications, water and waste management, oil and gas refining, and energy are all monitored and controlled by SCADA systems [6]. A paradigm shift brought about by cloud computing and the Internet of Things (IoT) is fostering innovation, enabling more adaptable resources, and reducing operating costs. ICS is transitioning to cloud computing and IoT to improve operation supervision and control by sharing real-time data among machines, industrial chains, suppliers, and customers. ICS is converting to cloud computing and IoT to improve monitoring and control operations. SCADA systems may present a security concern when connected to the internet since they were designed as air-gapped or isolated systems with unique cyber and physical interactions [7]-[9].

Since Stuxnet's first exposure, there has been a huge global

Lukumba Phiri is with the Department of Electrical and Electronic Engineering, School of Engineering, University of Zambia, Lusaka, Zambia (corresponding author, e-mail: phirilukumba@ gmail.com)

Simon Tembo is The Head of Department Electrical and Electronic Engineering, School of Engineering, University of Zambia, Lusaka, Zambia (e-mail: simon.tembo@unza.zm).

Kumbuso Joshua Nyoni is with the College of Science and Engineering, School of Geosciences, University of Edinburgh, UK (e-mail: kumbusojosh@gmail.com).

surge of cyber security incidents that have impacted electric grids. Black Energy breached the ICS of numerous national critical infrastructures in the United States in 2011. Three-quarters of corporate Windows-based PCs at Saudi Aramco, one of the biggest oil companies in the world, were infected by Shamoon, a self-replicating piece of malware [10]. A similar assault on Saudi Aramco was begun in August 2017. In February 2013, JEA was the victim of a distributed denial-of-service (DDoS) assault, which briefly shut down the online and telephone payment systems [11].

It is becoming more and more obvious that many tiers of intelligent countermeasures are required to protect SCADA infrastructure components and the essential applications they enable. Numerous government assessments have found severe cyber security problems in the electric sector as a result of the emergence of Advanced Persistent Threats (APTs) and the urgent need to protect against them [12]. For example, in the United States, the Department of Energy (DOE) created a cyber security Risk Management Process (RMP) for the electric sector in tandem with the requirements in [13] and [14]. National programs such as the NERC Critical Infrastructure Protection (CIP) [14] and the NIST Interagency Report (NISTIR) 7628 [15] guarantee that suitable standards and safeguards are in place to protect the electric power system from potential cyber vulnerabilities and threats.

We adopt a network-based analysis methodology in this paper and create SEADS, an edge-based multi-level anomaly detection system for SCADA networks. The remote substations, which are the limits of the SCADA network, are where SEADS is situated. To keep track of the condition of SCADA assets, it includes a stochastic anomaly detector. We also present the idea of confidence in the metrics for system evaluation used to measure availability, maintainability, and dependability.

The contributions of this paper are as follows:

1. Implementation of the stochastic reasoning in cyber risk modeling and assessment
2. Use of both transient and steady-state probabilities to evaluate system availability, maintainability, and reliability.
3. Novel modeling of intrusions in SCADA.

In Section II, we describe the structure of the smart grid system and its key benefits as well as its vulnerabilities. Section III summarizes related key anomaly detection and risk assessment frameworks. Then, we propose a stochastic edge-based framework for the SCADA systems in Section IV. In Section V, we compare the proposed framework with the existing management schemes and frameworks. Finally, Section VI concludes this work.

II. SCADA System

A. The Smart Grid Systems

A smart grid is created by fusing information and communication technology with conventional electrical infrastructure. It exchanges data on grid issues and customer requests via networking techniques. Power production and electricity loss reduction are the key goals of this integration. To address the rising electricity demand, the smart grid also incorporates traditional power plants with renewable energy sources. Additionally, the smart grid helps to reduce carbon dioxide (CO_2) emissions and save the environment. To meet the rising electricity demand, more distributed generators (DGs) are being added to smart grids; the bulk of these DGs are renewable resource-based generators, such as wind turbines and solar panels. Furthermore, original techniques, such as microgrids and vehicle-to-grid (V2G) connection, are utilized in smart grids. The micro-grid offers electrical self-sufficiency for a specific area using one or more DGs and storage units and allows the area to be isolated or connected to the main grid according to the current status of the grid; this feature protects the micro-grid in case of a blackout and assists the self-healing of the grid. In addition, the smart grid utilizes the EVs' batteries as temporary storage units for the extra generated power during low demand periods; V2G networks organize the charging/ discharging operations of the EVs' batteries to guarantee a balanced electricity level in the grid [16]-[18].

B. Smart Grid Benefits

A smart grid can improve the efficiency of the maintenance and replacement operations for the involved devices in the grid. For example, there are many deployed sensors in the smart grid for monitoring purposes; they monitor the performance of the different devices and send an alarm message to the control center in case of an error. Finally, a smart grid is a friend to the environment, as it organizes electricity production and uses renewable generation resources. Accordingly, the smart grid plays a significant role in CO_2 emission reduction. To conclude, utility companies are interested in smart grids to assure the optimal usage of electrical power and provide more luxury services to the customers, and consequently, increase their financial profits [16]-[18], [20].

C. Smart Grid Architecture

To accomplish its functions, the smart grid adds new elements and protocols to the electrical grid (see Fig. 1). The reference model for the smart grid, its various layers, and their purposes, and, finally, the systems of the smart grid are all introduced in this part.

Numerous frameworks have been proposed to define the smart grid's structure. The smart grid reference model, according to [19], consists of seven functional domains:

1. Bulk Generation: Electricity is usually generated from non-renewable resources, such as coal and gas generators. In a smart grid, renewable sources, e.g., wind turbines and solar panels, are merged with the traditional ones to satisfy the increased demands and reduce CO_2 emissions.
2. Transmission: Several substations and transmission lines are utilized to transmit the produced power to consumers.
3. Distribution: The distribution domain spreads the electricity to individual customers and communicates with suppliers and users via communication infrastructure.
4. Operation: This domain controls and monitors the transmission and distribution domains to obtain

information about the power system's activities.

5. Market: This domain contains all the parties involved in the electricity-trade operation to sustain the balance between supply and demand.
6. Customer: Customers in the smart grid not only consume electricity but also generate it by DGs and store the extra power in rechargeable batteries.
7. Service Provider: The electricity is provided to customers via a service provider that is responsible for services, such as billing and customer accounts management.

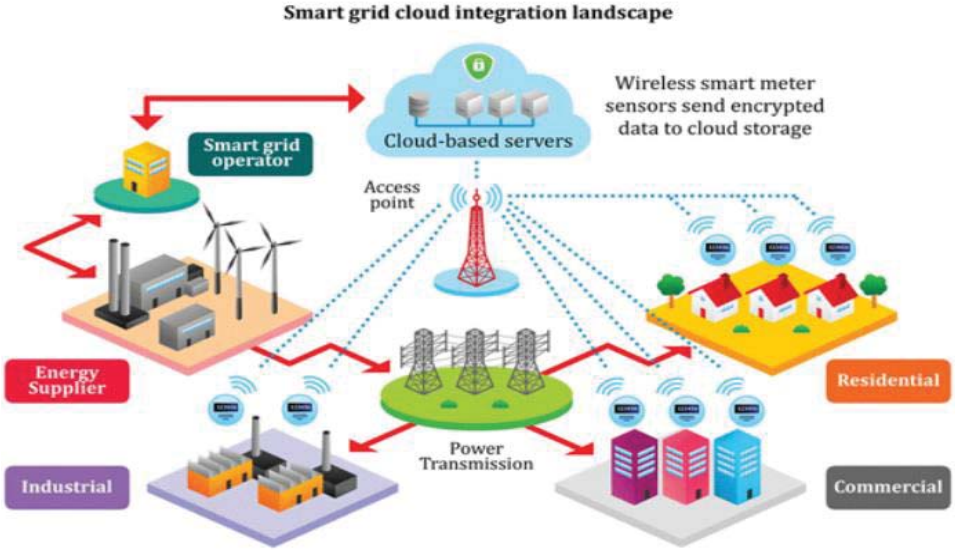

Fig. 1 Structure of the Smart Grid System

D. Assets and Categories

In this section, we describe the assets and categories for *SCADA cyber security*.

TABLE I
ASSET AND CATEGORIES FOR SCADA

Category	Asset	Category	Asset
System	ICCP server	Security	User Account
	HMI		Admin Account
	RTU		Service Account
	Communication front end		Vulnerability
	Alarm	Communication	Firewall
	App Server		Zones
	Database		Router
	Antivirus server		Data diodes
	Directory service		
	Backup server		
	Product		
	DNS Server		
	NIS Server		
	NTP server		
	DE server		
	nHMI server		
	Software		

E. Actors

Table II focuses on actors who primarily work with such assets that can serve as an entrance point for prospective assaults. For simplicity, we only chose actors that are key to a SCADA since they have access to many system components. In this work, we do not focus on all actors in a SCADA system.

TABLE II
LIST OF MAIN ACTORS IN SCADA

Actor	Description
HMI Operator	monitoring and controlling the power transmission network
Admin User	monitoring system health and updating control system configuration
Admin Directory Service User	monitoring service account and access through directory service
Trainer and students	running simulations
Data engineer	updating the power system model
Production planner	viewing historical data and creating plans
Field engineer	sending the data from RTU to the communication front end

F. Attacker Profiles

Rogue actors that we consider for this work are as follows:

- *HMI Operator* can be considered rouge when the accounts to HMI were leaked. This way the attack path can start from HMI.
- *Field engineer* can also be a rogue actor to generate the attacks starting from RTU to the CC.
- *Admin User* of other SCADA that is connected to the current SCADA can be rogue when the first system got compromised.

External SCADA is another system that communicates with the current SCADA through ICCP. Having a connection to another SCADA system is necessary to ensure the guaranteed availability of the service in case of any critical issues. If the secondary SCADA fails to attack, the primary SCADA becomes vulnerable to attacks through ICCP.

Field units include all components of SCADA that collect information about the state of the system in the field. Therefore, we can assume that malicious actors could misconfigure the

data that field units send to CC.

G. Attacks

In this section, we describe attack techniques and vulnerability exploits for SCADA.

1) The attack techniques vary based on the target's key assets. Most importantly, the *availability* of the SCADA is the main priority of CIA properties. This is due to the necessity to continuously preserve the industrial processes putting data loss and confidentiality aside as less priority. Overall, we summarized the goals of attackers into the following categories:

a) *Manipulating sensitive data*,

- Loss of view,
- Manipulation of view,
- Theft of Operational Information

b) Disrupting the safety of operation in SCADA

- Damage to property
- Loss of safety

c) *Disrupting the availability of the system*.

- Loss of Productivity and Revenue
- Denial of control

H. Entry Surface

Entry surface refers to the resources and methods that, either because of the setup of the system (such as front-end web servers or applications for external users) or because of vulnerabilities, are most accessible to hostile actors (e.g. the kernel version, malicious library version found, the possibility of buffer overflow due to weak code, lack of data sanitizing and validation). Regarding SCADA, three entry surfaces are the focus of this work.:

1) External SCADA
2) Field units
3) HMI

III. Related Works

A. SCADA Network-Based IDSs

A SCADA network-based IDS [20], [21] captures the data packets that are communicated between devices such as point-to-point between RTU/PLC, and between RTU/PLCs and the MTU. If a packet is suspicious, the security team will be sent an alarm for further investigation. An advantage of a SCADA network-based IDS is its lower computation costs, as only information in the packet's header is needed during the investigation process, and therefore a SCADA network packet can be analyzed on-the-fly. Consequently, traffic from larger networks can be inspected within a short period [22]. When there is high network traffic, however, a SCADA network-based IDS may experience issues in monitoring all the packets and might miss some attacks. However, the key weakness is that the operational behavior of the underlying SCADA processes cannot be inferred from the information provided at the network level (e.g., IP address, protocol, port, and so on). For example, if the payload of the SCADA network packet contains a malicious message, which is crafted at the application level, the SCADA network-based IDS cannot detect it, particularly when this is not violating the specifications of the protocol being used, or the communication pattern between SCADA networked devices [21].

B. SCADA Application-Based IDSs

SCADA data, which comprise the measurements and control data generated by sensors and actuators, represent the majority of the information. Using these data, the operational behavior of a given SCADA system can be inferred [21]. In contrast to SCADA network-based IDSs that only inspect network-level information, a SCADA application-based IDS can inspect high-level data (i.e., SCADA data) to detect the presence of unusual behavior. For example, SCADA network-based IDSs are often unable to detect high-level control attacks [22] from packet headers; which can be detected by analyzing SCADA data [22].

The following are the several methods to deploy a SCADA application-based IDS since the information source of a SCADA application-based IDS can be acquired from various remote field devices [23]. It can be installed on the historian server because this server receives periodic updates from the MTU server, which collects data and system status for the monitored system using field devices like PLCs and RTUs. When the data and status kept in the historian diverge from the real-time information in the field, this type of deployment poses a security concern. This could occur when the MTU server has been compromised or the data has been changed using False Data Injection attacks [23]; (ii) It can also be deployed in an independent security-hardened server, which from time to time acquires information and statuses from the monitored field devices [24]. Consequently, the large number of requests from this server might increase the network overheads resulting in degraded performance of the IDS; (iii) Similar to the approach suggested in [25] and [26], each neighboring field device can be connected to a server running a SCADA application-based IDS. The main problem, however, is that SCADA data are directly or indirectly connected, thus occasionally an abnormal value in one parameter is caused by an abnormal value in another parameter [21], [22]. The identification and monitoring of associated parameters, such as sensor readings on a single process, would therefore be acceptable.

C. Signature-Based vs. Anomaly-Based SCADA IDS Approaches

The many SCADA-based IDS that have been described in the literature may be roughly categorized into two types based on the detection method: signature-based detection and anomaly-based detection.

A SCADA system's network traffic or application events can be investigated by an IDS that employs signatures to spot malicious activity. This is done by looking for warning signs and comparing patterns to a database of accepted attack signatures or fingerprints. In this kind of IDS, the false positive rate—the percentage of times a regular event is mistakenly classified as an attack—is extremely low and may even be nil. Additionally, since only a matching procedure is used during the detection phase, the detection time can be quick. Despite the benefits of signature-based IDSs noted above, they frequently

miss new attacks (like zero-days) whose signatures are unknown or do not already exist in their database. As a result, the database needs to be updated frequently with new attack patterns [27].

An anomaly-based IDS assumes that the actions of invasive activities may be easily distinguished from regular actions. Using sophisticated mathematical and statistical procedures, the "normal model" is produced using a realistic training set. This model is identified as an anomaly or a potential attack whenever there is a considerable departure from it. To create the regular SCADA network profiles, a modeling technique, for instance, can be used to acquire the normal SCADA network traffic for normal operations. During the detection phase, the deviation degree between the current network traffic and the created normal network profile is computed: if the deviation exceeds the predefined threshold, the current network traffic will be flagged as an intrusive activity. The primary advantage of anomaly-based IDSs compared to signature-based ones is that new or unknown attacks can be detected, although it generally suffers from a higher false positive rate (i.e., detecting normal behavior as malicious) [28].

IV. Framework Design

In this study, we present a unique approach to measuring and analyzing cyber risk that is based on stochastic reasoning. One MTU and many SUB-MTUs, as well as one SUB-MTU and many RTUs, can use the suggested architecture as an interface.

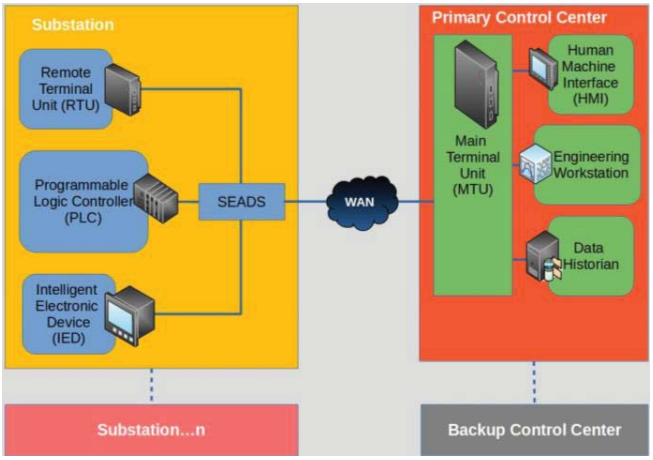

Fig. 2 SEADS Deployment

A. The Framework

The framework's main objectives are to represent all potential attack vectors in the digital control network (smart grid topologies), score the security of the smart grid using security metrics, and evaluate the efficacy of defense measures. The suggested framework, which can be implemented in layer 3.5 of the Purdue design, is seen in Fig. 3. The framework consists of five steps: preprocessing; development of security models; visualization and storage; security analysis; and, changes and updates. Each step is described as follows:

Step1. The security decision-maker offers the inputs required to build a smart network in step 1. The total number of nodes, the network topology, and each node's vulnerability data are the essential inputs. The SG Generator receives the inputs. A smart grid network with a specific network topology made up of levels and nodes with information on their vulnerabilities is created using the SG Generator. After generation, the topology of the network is fixed. The security decision-maker also chooses the security metrics that will be utilized as input in the security analysis phase from a pre-defined metric pool.

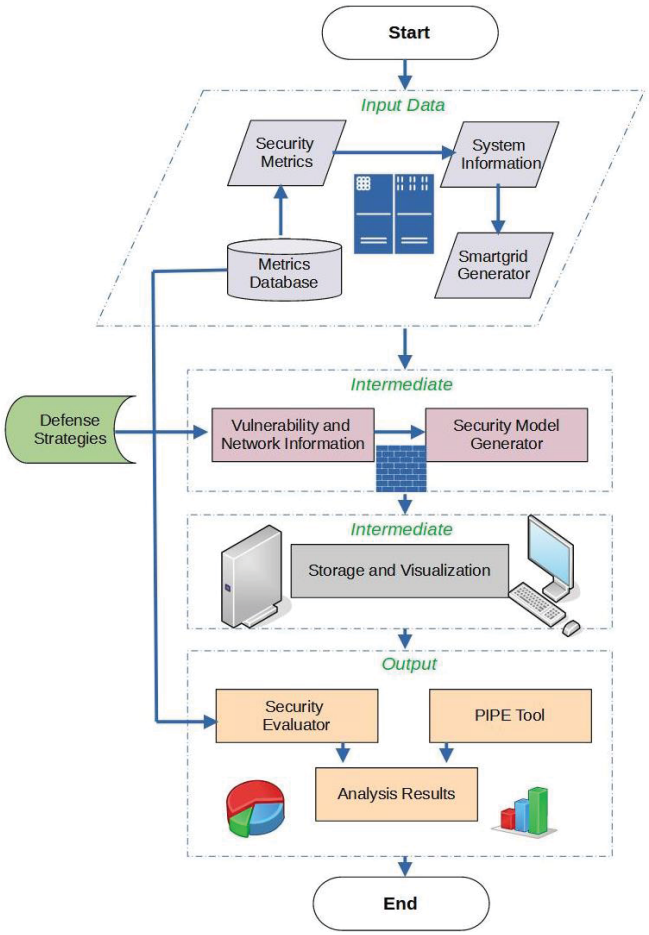

Fig. 3 The Proposed Framework

Step2. The security model has been created and is complete. Our security strategy is built on the architecture of the SCADA network. The constructed network, coupled with inputs about the topology and vulnerabilities, is used by the Security Model Generator to automatically compute all possible attack paths in the SG network.

Step3. The reachability/coverability graph that represents the tangible and vanishing states is used to display the attack paths produced by the security model generator.

Step4. The SG network is the subject of the security analysis. Along with the established security metrics, the attack vectors are entered into the security evaluator. The security analysts can carry out one of the two choices based on the metrics. One option is to output the analysis

results directly, and the other is to create a text file and import it into the Platform Independent Petri net Editor (PIPE) [29] analytical modeling and evaluation tool, which computes the security analysis results. A predefined metric database is used to choose the security metric.

Step5. This step involves updating the model's inputs to reflect any changes brought on by the defense strategies. The security decision-maker can choose the best defense tactics because they are aware of which area of the SG is most vulnerable based on the findings of the security analysis. The implementation of the defense strategy modifies either the topology information, which should be updated and used as input to the Security Model Generator or the vulnerability information (e.g., eliminates a specific vulnerability in a smart grid node or mitigates the effect caused by the vulnerability). The security decision-maker can compare the results of several protection measures utilizing security metrics while selecting the optimal one by assessing each one's effectiveness using the framework.

B. Stochastic Process

A stochastic process (also known as a chance or random process) is a group of random variables that are indexed by a parameter like time [30].

X(t) | t ∈ T, defined on a certain probability space, indexed by the parameter t, where t fluctuates over an index set, T [30], is a family of random variables that make up a stochastic process.

States are the values that the random variable X(t) assumes, and the state space of the process is the set of all possible values. The letter I [30] will stand in for the state space.

A stochastic process is referred to as a discrete-state process, often known as a "chain," if the state space is discrete. The state space in this situation is frequently thought to be {0, 1, 2, . . .} etc. As an alternative, we have a continuous-state process if the state space is continuous. Similar to this, we have a discrete-time (parameter) process if the index set T is discrete; otherwise, we have a continuous-time (parameter) process. The symbol for a discrete-time process, commonly known as a stochastic sequence, is {Xn| n ∈ T} [30]. As indicated in Table III, this results in four different kinds of stochastic processes.

TABLE III
CATEGORIES OF STOCHASTIC PROCESSES

	Index set T (state space)	
Time Parameters	Discrete-time Stochastic chain	Continuous state
Discrete-Time	Discrete-time Stochastic chain	Discrete-time Stochastic process
Continuous Time	Continuous Stochastic chain	Continuous Stochastic process

Classification of Stochastic Processes

For a fixed time $t = t1$, the term $X(t1)$ is a simple random variable that describes the state of the process at time $t1$. For a fixed number $x1$, the probability of the event $[X(t1) \leq x1]$ gives the CDF of the random variable $X(t1)$, denoted by [30].

$$P(X(t1)\ x1) = F(x1;t1) = FX(t1)\ (x1)$$

The first-order distribution of the process $\{X(t) \mid t \geq 0\}$ is denoted as F(x1; t1). X(t1) and X(t2) are two random variables on the same probability space given two-time instants, t1, and t2. The formula for the process' second-order distribution, also referred to as its joint distribution, is $F(x1, x2; t1, t2) = P[X(t1) \leq 1, X(t2) \leq x2$.

In general, we define the nth-order joint distribution of the stochastic process $X(t)$, $t \in T$ by

$$F(x; t) = P[X(t1) \leq x1, \ldots, X(tn) \leq xn] \quad (1)$$

for all x = $(x1, \ldots, xn) \in \_n$ and t = $(t1, t2, \ldots, tn) \in Tn$ such that $t1 < t2 < \cdot\ \cdot\ \cdot < tn$. Such a complete description of a process is no small task. Many processes of practical interest, however, permit a much simpler description. For instance, the nth-order joint distribution function is often found to be invariant under shifts of the time origin. Such a process is said to be a strict-sense stationary stochastic process [30].

Definition (Strictly Stationary Process). A stochastic process $\{X(t) \mid t \in T\}$ is said to be stationary in the strict sense if for $n \geq 1$, its nth-order joint CDF satisfies the condition:

$$F(x; t) = F(x; t + \tau)$$

for all vectors $x \in \_n$ and $t \in Tn$, and all scalars τ such that $ti + \tau \in T$. The notation $t + \tau$ implies that the scalar τ is added to all components of vector t.

We write $\mu(t) = E[X(t)]$ to represent the stochastic process' time-dependent mean. The stochastic process ensemble average is usually referred to as $\mu(t)$. When the strictly stationary process definition is applied to the first-order CDF, we obtain $F(x; t) = F(x; t + \tau)$ or $FX(t) = FX(t+\tau)$ for all τ. It follows that a strict-sense stationary stochastic process has a time-independent mean; that is, $\mu(t) = \mu$ for all $t \in T$.

By restricting the nature of dependence among the random variables $\{X(t)\}$, a simpler form of the nth-order joint cumulative distribution function (CDF) can be obtained.

The simplest form of the joint distribution corresponds to a family of independent random variables. Then the joint distribution is given by the product of individual distributions [30].

A stochastic process $\{X(t) \mid t \in T\}$ is said to be an independent process provided its nth-order joint distribution satisfies the condition:

$$F(x;t)=\prod_{i=1}^{n}(xi; t1) \quad (2)$$

A discrete-time independent process known as a renewal process is denoted by the notation {Xn| n = 1, 2, ...} where $X1$, $X2$, ..., are independent, identically distributed, nonnegative random variables.

As an example of such a process, we consider a system in which the repair (or replacement) after a failure is performed, requiring negligible time. Now, the gaps between subsequent failures could very well be independent random variables

created by a renewal process with identical distributions $\{Xn|\ n = 1, 2, \ldots\}$. Although assuming an independent process makes analysis much simpler, this assumption is frequently unfounded, and we are compelled to take some type of dependence among these random variables into consideration. First-order reliance, also known as Markov dependence, is the most fundamental and significant type of dependence [30].

C. The Markov Chain

The stationery distribution is often discussed in the Markov chain. To obtain a clearer picture, let us assume that substation attacks consist of three states, as shown in Fig. 4 as depicted by [30].

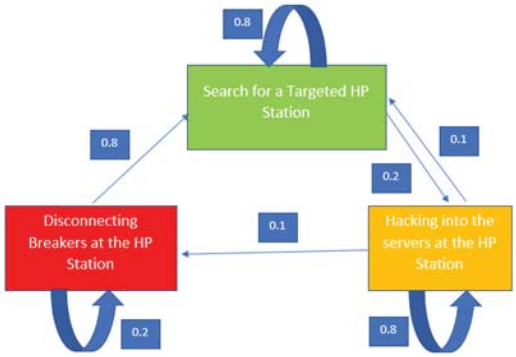

Fig. 2 Example of attack transitions to Hydro Power station network

The state space is shown in this case as $S = \{S_1,S_2,S_3\}$ where,

- S_1: being the search for a targeted hydropower facility
- S_2: hacking into hydropower station servers
- S_3: turning off the breakers at the hydroelectric plant

The probability of this state as well as the likelihood of changing from one state to another is both expressed as $P(X_m)$ and $P(X_{m+1}|X_m)$, respectively. In this example in Fig. 4, $P(X_{m+1} = S_2|X_m = S_1) = 0.1$ and $P(X_{m+1} = S_2|X_m = S_2) = 0.8$.

The Markov chain is defined as (3) using a state at time m, $\{X_m\}$.

$$P(X_{m+1}|X_m,\ldots,X_1,X_0) = P(X_{m+1}) \tag{3}$$

The meaning of this equation can be summarized as two bullet points:

- X_{m+1} is determined by X_m only
- $X_{m-1},X_{m-2},X_{m-3}\ldots$ are nothing to do with X_{m+1}

In this example, it can be stated that disconnecting breakers is nothing to do with searching for the targeted HPstation but has much to do with cracking the server at the Hydro Power station only. These characteristics shown in (4) are called Markov properties. When the Markov chain and its relevant theorems are used, the Markov property for the created Markov chain model needs to be tested first. If the Markov property is not justified, the Markov chain model needs to be further updated, and segmentalizing the states, i.e., increasing the number of states is known as a general countermeasure. Therefore, the Markov chain can be utilized, especially when the action flow or procedure is clarified.

To obtain a clear image, let's use the previous example in Fig. 4. In the transition probability matrix, P is expressed as (5). It can be realized that the summation of each row is always one. In other words, the summation of the probabilities from one state to another (including the same state) needs to be always one. This is an important property that the Markov chain owns.

$$P = \begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0.8 & 0 & 0.2 \end{bmatrix} \tag{4}$$

We assume that our hacker starts at state (search for the target HPstation). In other words, the initial distribution is $\Pi^{(0)} = (1,0,0)$. From discovering the hydropower station, the hacker can go to hacking the server at the HPstation and further disconnect circuit breakers at the HPstation with equal probability, i.e.,

$$\Pi^{(1)} = (1,0,0)\begin{bmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.8 & 0.1 \\ 0.8 & 0 & 0.2 \end{bmatrix} = (0.9,0.1,0)$$

If we analyzed further, the vector $\Pi^{(m)}$ of state probabilities tends to a limit of $m\to\infty$. Even more, one can show that for specific discrete-time Markov chains (DTMCs) the effect of $\Pi^{(0)}$ on the vector $\Pi^{(m)}$ completely vanishes.

D. Modeling in SEADS

We applied the steps outlined in the framework in Section IV *A*. The intrusion is assumed to emanate from the HMI through the RTU and into the IED. The resulting model is depicted in Fig. 5. Table IV describes the parameters applied to the model in Fig. 5.

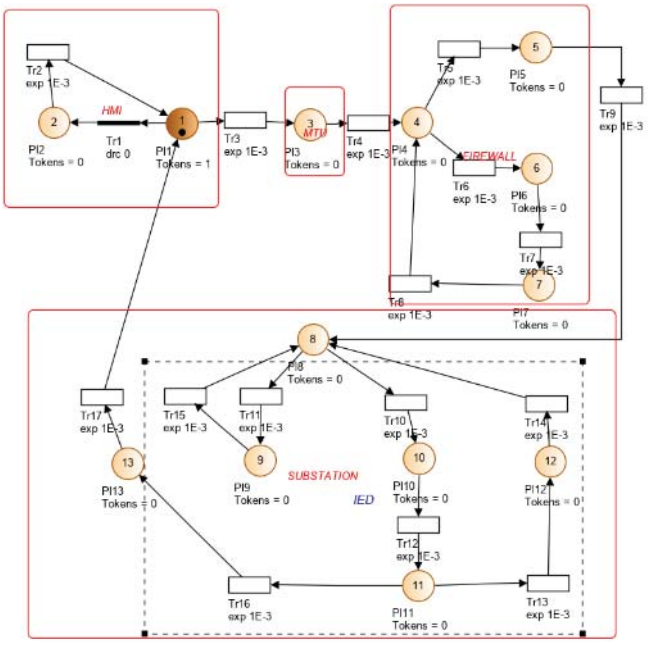

Fig. 3 SEADS Model

TABLE IV
DESCRIPTION OF PLACES AND TRANSITIONS

Places	Description	Rates
PI1	Intrusion attempts begin P_Begin	
P12	Intrusion failed	
P13	Intrusion successful	
P14	Intrusion into the FW begins	
PI5	Intrusion into the FW successful	
PI6	Intrusion into the FW failed	
PI7	FW restored	
PI8	Intrusion into the Substation begins	
PI9	Intrusion into the Substation failed	
PI10	Reconnaissance begins	
PI11	Execution of IED attack	
PI12	IED under attack recovered	
PI13	Intrusion toward the HMI begins	
Transition	**Description**	**Rate**
Tr1	From begin to intrusion failed	(0.01)
Tr2	System recovery and reset after failed attempt	(0.001)
Tr3	transition into MTU succeeded	(0.001)
Tr4	From MTU to FW	(0.001)
Tr5	the transition from MTU into FW succeeded	(0.001)
Tr6	the transition from MTU into FW failed	(0.001)
Tr7	From failed to recovery	(0.01)
Tr8	FW recovery	(0.001)
Tr9	From FW to S/S network	(0.001)
Tr10	S/S attack begins to IED attack successful	(0.001)
Tr11	S/S attack begins to IED attack successful	(0.01)
Tr12	Attack successful	
Tr13	IED attack to recovery	
Tr14	Recovery from an attack	
Tr15	Reset from a failed attack	
Tr16	From IED to HMI	
Tr17	From attacked IED to initial state	

E. Simulation Results

Simulations are conducted following the detailed steps of the framework given in Section IV *A*. The simulation is done using the scenarios in Table VI.

TABLE VI
SIMULATION SCENARIOS

	Number of histories	First random number	Maximum calculation time
Scenario 1	10	12345681	10
Scenario 2	100	12345681	10
Scenario 3	1000	12345681	10

F. Metrics

In this part, we look at three dependability criteria for digital control networks in smart grids, namely reliability, availability, and maintainability. The ability to constantly supply services without interruptions is assessed by reliability [31]. It can be specifically described as the likelihood that the digital control networks function successfully across the period [0, t], or,

$$R(t) = Pr\{X > t\} = e^{-\lambda t} \text{ and } R(0) = 1 \quad (5)$$

When failure is exponentially distributed with a constant failure rate λ and the system is operational at time zero.

$$Rel = R(\infty) = 0 \text{ and } MTTF = 1/\lambda \quad (6)$$

TABLE V
STEADY STATE PROBABILITIES OF THE SCENARIOS

Name	Probabilities Scenario 1	Probabilities Scenario 2	Probabilities Scenario 3
Pl1	0.051196576	0.036051431	0.037315122
Pl2	0.338140172	0.357981292	0.356126014
Pl3	0.0350389	0.060711856	0.073966504
Pl4	0.015546133	0.013348975	0.020544781
Pl5	0	0.072463596	0.045247957
Pl6	0.103760954	0.176213323	0.166132557
Pl7	0.250079113	0.172444476	0.173307633
Pl8	0	0.02125842	0.015960483
Pl9	0	0.168220049	0.116335319
Pl10	0	0.008386059	0.028047419
Pl11	0	0.002931009	0.001981801
Pl12	0	0.015639034	0.011841172
Pl13	0	0	0

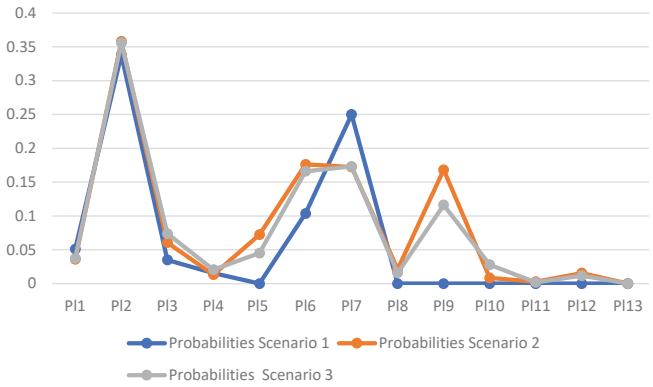

Fig. 4 Comparison of Scenario Probabilities

According to [32], maintainability is the likelihood that a malfunctioning system will be repaired and made operational within a given downtime t.

$$M(t) = 1 - e^{-\mu t} \quad (7)$$

where t denotes the downtime (i.e., time to repair) and the repair distribution is exponentially distributed with a constant repair rate μ. The probability of maintainability (Mnt) as t approaches infinity and the mean time to repair (MTTR) is given by [32]:

$$Mnt = M(\infty) = 1 \text{ and } MTTR = 1/\mu \quad (8)$$

Availability is determined by dependability and maintainability and is defined as the percentage of time the system delivers the right services throughout an observation period [31]. The dependability of each component is measured by MTTF, while the maintainability is measured by MTTR. The MTTF and MTTR should be designed as high and low as possible, respectively, to achieve high steady-state availability. To ensure that the system offers accurate data transmission services from the plant network to the corporate network or vice versa, we are interested in the steady-state availability analysis. Given the steady-state availability:

$$AVL = \sum_j \pi_j \quad (9)$$

where πj is the steady-state solution corresponding to the state j where the system is available, i.e., providing correct services. The steady-state solution π can be calculated by using (1)-(4).

VI. Conclusion

A SCADA system is a significantly important system used in national infrastructures such as electric grids, water supplies, and pipelines. However, the SCADA systems have lots of security vulnerabilities. Any faults or damages to the SCADA system can affect society severely. The study of the security of the SCADA system is essential for that reason.

In this paper, we discussed the cyber-physical security and dependability issues of SCADA systems. We used stochastic processes to model intrusions into digital control networks. The cyber framework that we then suggest is compliant with the NIST framework. Additionally, we assess the steady-state availability using GSPNs and demonstrate the excellent dependability of the suggested framework. In further work, we will model resilience and provide mitigation to enhance the RAM metrics using statistics from a functioning power plant in Zambia, including failure rates, repair rates, failed login attempts, and firewall rates. Additionally, we will offer a better framework based on Generalized Stochastic Petri Nets and Bayesian Nets and compare the results with our current work.

References

[1] "The National Energy Policy 2019." Ministry of Energy Integrated Resource Plan, 21 Oct. 2021, https://www.moe.gov.zm/irp/download/the-national-energy-policy-2019-2/.

[2] Final Report - Moe.gov.zm. https://www.moe.gov.zm/?wpfb_dl=45.

[3] "Home." Ministry of Energy Integrated Resource Plan, 1 Sept. 2021, https://www.moe.gov.zm/irp/.

[4] Energy Sector Report 2020 - Erb.org.zm. https://www.erb.org.zm/reports/esr2020.pdf.

[5] Awad, A.; Bazan, P.; German, R. SGsim: A simulation framework for smart grid applications. In Proceedings of the 2014 IEEE International Energy Conference (ENERGYCON), Cavtat, Croatia, 13–16 May 2014; pp. 730–736.

[6] Al Ghazo, Alaa, "A framework for Cybersecurity of Supervisory Control and Data Acquisition (SCADA) Systems and Industrial Control Systems (ICS)" (2020). Graduate Theses and Dissertations. 17834.

[7] Davis, Katherine R., et al. "A Cyber-Physical Modeling and Assessment Framework for Power Grid Infrastructures." *University of Illinois Urbana-Champaign*, Institute of Electrical and Electronics Engineers Inc., 1 Sept. 2015, https://experts.illinois.edu/en/publications/a-cyber-physical-modeling-and-assessment-framework-for-power-grid-3.

[8] Handa, A., Sharma, A., and Shukla, S. K. Machine learning in cybersecurity: a review. WIREs Data Mining Knowl Discov. 9, e1306.doi:10.1002/widm.1306

[9] Johnson, J., Onunkwo, I., Cordeiro, P., Wright, B.J., Ja-cobs, N. and Lai, C. Assessing DER network cybersecurity defenses in a power-communication co-simulation environment. IET Cyber-Physical Systems: Theory & Applications, 5: 274-282. https://doi.org/10.1049/iet-cps.2019.0084

[10] Li, Beibei & Xiao, Gaoxi & Lu, Rongxing & Deng, Ruilong & Bao, Haiyong. (2019). On Feasibility and Limitations of Detecting False Data Injection Attacks on Power Grid State Estimation Using D-FACTS Devices. IEEE Transactions on Industrial Informatics. PP. 10.1109/TII.2019.2922215

[11] Christopher Baker, by J., & Air Force Base, M.. Cybersecurity for critical infrastructure a Research Report Submitted to the Faculty In Partial Fulfillment of the Graduation Requirements for the Degree of master of operational arts and sciences advisor: wing commander Air Force Base the United States, 2015.

[12] Office of Electricity Delivery and Energy Reliability. Cybersecurity risk management process (RMP). 2011.

[13] North American Electricity Reliability Council (NERC). Critical infrastructure protection (CIP) reliability standards. 2009.

[14] National Institute of Standards and Technology (NIST). Nistir 7628: Guidelines for smart grid cyber security. 2010.

[15] Hassan Bevrani. Robust power system frequency control. Springer, 2014.

[16] Jaime De La Ree, Virgilio Centeno, James S Thorp, and Arun G Phadke. Synchronized phasor measurement applications in power systems. IEEE Transactions on Smart Grid, 1(1):20{27, 2010.

[17] "The Cyber-Physical Security of the Power Grid." IEEE Smart Grid, https://smartgrid.ieee.org/bulletins/november-2019/the-cyber-physical-security-of-the-power-grid.

[18] W. Wang, Y. Xu, and M. Khanna, A survey on the communication architectures in smartgrid," Computer Networks, vol. 55, no. 15, pp. 3604{3629, 2011.

[19] Hamid Gharavi and Bin Hu. Synchrophasor sensor networks for grid communication and protection. Proceedings of the IEEE, 2017

[20] Abdul Mohsen Afaf Almalawi. 2014. *Designing Unsupervised Intrusion Detection for SCADA Systems*. Ph.D. Dissertation. RMIT University, School of Computer Science.

[21] Andrea Carcano, Alessio Coletta, Michele Guglielmi, Marcelo Masera, Igor Nai Fovino, and Alberto Trombetta. 2011. A multidimensional critical state analysis for detecting intrusions in SCADA systems. IEEE Transactions on Industrial Informatics 7, 2 (May 2011), 179–186.

[22] Igor Nai Fovino, Alessio Coletta, Andrea Carcano, and Marcelo Masera. 2012. Critical state-based filtering system for securing SCADA network protocols. IEEE Transactions on Industrial Electronics 59, 10 (October 2012), 3943–3950.

[23] Adnan Anwar, Abdun Naser Mahmood, and Mohiuddin Ahmed. 2014. False data injection attack targeting the LTC transformers to disrupt smart grid operation. In International Conference on Security and Privacy in Communication Systems. Springer International Publishing, Cham, 252–266.

[24] Adnan Anwar, Abdun N. Mahmood, and Zahir Tari. 2017. Ensuring data integrity of OPF module and energy database by detecting changes in power flow patterns in smart grids. IEEE Transactions on Industrial Informatics 13, 6 (2017), 3299–3311.

[25] Cristina Alcaraz and Javier Lopez. 2014. Diagnosis mechanism for accurate monitoring in critical infrastructure protection. Computer Standards & Interfaces 36, 3 (2014), 501–512. DOI:https://doi.org/10.1016/j.csi.2013.10.002

[26] Cristina Alcaraz and Javier Lopez. 2014. WASAM: A dynamic wide-area situational awareness model for critical domains in smart grids. Future Generation Computer Systems 30 (2014), 146–154.

[27] Digitalbond.com. 2013. IDS-signatures/modbus-tcp. Retrieved December, 2018 from http://www.digitalbond.com/index.php/research/ids-signatures/modbus-tcp-ids-signatures/.

[28] Mohiuddin Ahmed, Adnan Anwar, Abdun Naser Mahmood, Zubair Shah, and Michael J. Maher. 2015. An investigation of performance analysis of anomaly detection techniques for big data in SCADA systems. EAI Endorsed Transactions on Industrial Networks and Intelligent Systems 2 (2015), 1–16. Issue 3,e5. DOI:https://doi.org/10.4108/inis.2.3.e5

[29] https://github.com/sarahtattersall/PIPE

[30] F. Bause and P. S. Kritzinger, Stochastic Petri Nets: An Introduction to the Theory, 2nd ed. Braunschweig, Germany: Vieweg, 2002.

[31] Helerea, Elena. "Interconnections between Reliability, Maintenance, and Availability." IFIP Advances in Information and Communication Technology, 19 Aug. 2016, https://www.academia.edu/27901809/Interconnections_between_Reliability_Maintenance_and_Availability.

[32] M. Kim, A Survey on Guaranteeing Availability in Smart Grid Communications," in Proc.ICACT, Korea, February 2012.