A Detection and Response Architecture for Stealthy Attacks on Cyber-Physical Systems

Tawfeeq Shawly a, *

a Electrical Engineering Dept. at Rabigh, King Abdulaziz University, Jeddah, 21589, Saudi Arabia
Corresponding author:* tshawly@kau.edu.sa

Abstract — There has been an increased reliance on interconnected Cyber-Physical Systems (CPS) applications. This reliance has caused tremendous growth in high assurance challenges. Due to the functional interdependence between the internal systems of CPS applications, the utilities’ ability to reliably provide services could be disrupted if security threats are not addressed. To address this challenge, we propose a multi-level, multi-agent detection and response architecture built on the formalisms of Hidden Markov Models (HMM) and Markov Decision Processes (MDP). We have evaluated the performance of the proposed architecture on one of the critical smart grid applications, Advanced Metering Infrastructure (AMI). This paper utilizes a simulation tool called SecAMI for performance evaluation. A Stealthy attack scenario contains multiple distinct multi-stage attacks deployed concurrently in a network to compromise the system and stop several critical services in a CPS. The results show that the proposed architecture effectively detects and responds to stealthy attack scenarios against Cyber-Physical Systems. In particular, the simulation results show that the proposed system can preserve the availability of more than 93% of the AMI network under stealthy attacks. A future study may evaluate the effectiveness of various stealthy attack strategies and detection and response systems. The high availability of any AMI should be protected against new attack techniques. The proposed system will also determine a distributed IDS’s efficient placement for intrusion detection sensors and response nodes within an AMI.

Keywords — Security; detection; response; artificial intelligence; machine learning; CPS; AMI.

I. INTRODUCTION

A Cyber-Physical System (CPS) is a physically based system in which computer networks and algorithms monitor, manage, and control connected components. A well-known example of an application for CPS is the Smart Grid (SG), which can analyze data on power demands simultaneously to decide where to produce electricity, how to generate and distribute it, and also how to deal with a utility customer; for instance, who requires minimum power in the peak hours. Advanced Metering Infrastructures (AMIs), sometimes known as Smart Meters (SMs), are critical elements of the smart grid that greatly impact people’s lives.

In Saudi Arabia, the Smart Metering Project (SMP) is a significant AMI initiative carried out to fulfill Vision 2030 objectives of energy conservation and emission reduction [1]. Regarding initial deployment, it is one of the largest smart meter initiatives ever, with approximately 10 million smart meters deployed in one year. This project offers a variety of capabilities that enhance and enrich customers’ experiences with electricity services. There are many benefits, such as:

- The meter readings can be obtained remotely or on demand from power companies.
- Customers can look at specific consumption data and change their consumption habits as necessary (e.g., the ability to change the thermostat or turn off certain appliances during periods of high demand).
- When a consumer is behind on payments, for example, smart meters can be utilized to disconnect a residence from the utility network centrally.

These smart meters use a Neighborhood Area Network (NAN), a smart meter network that typically covers a full neighborhood. Based on parameters like distance, and other variables, each node in these mesh networks may link with a certain subset of other nodes. Each smart meter acts as a router, tying together all the nodes it talks with [2]. Each NAN contains a Data Concentrator Unit (DCU), a node that collects information from the NAN's SMs and sends it to the energy supplier.
Any smart meter may also request instructions to disengage from the utility remotely from the DCU. An example of an AMI and its essential elements is shown in Figure 1. Despite the potential smart meters and NANs have of becoming the primary inputs for crucial decisions on energy efficiency, optimization, and operational dependability, these technologies are limited by the utility network's vulnerabilities. For instance, by altering the data delivered to the utility through processes like False Data Injection (FDI), a malicious person who gets access to the DCU and controls it might be able to control power loads and cause blackouts [4]. If an attacker or multiple attackers gain access to the DCU, they can completely disconnect the NAN. According to a 2013 evaluation from NESCOR [5], a malicious person might trigger a mass disconnect. A smart meter from a specific vendor can be reverse engineered in the scenario described in AMI.27, disclosing a remotely exploitable vulnerability that might be used to control a specific meter.

AMI breaches often fall into one of two categories: control flow attacks or data flow attacks. When the attack is from the control flow category, the intruder moves across the network, getting control of each node she comes across until she approaches the DCU and is able to issue a mass disconnect order. A data flow attack involves the attacker injecting false data about the load at each meter into the network, ultimately resulting in system failure [4]. Among the biggest and most destructive attacks on the data integrity of energy management systems are those utilizing FDI [6]. The control and data flow attacks become stealthier in multiple interleaved networks where an assailant can exploit different vulnerabilities and entry points.

A reliable detection and response mechanism for general and AMI attacks on CPS is crucial. This paper proposes an integrated cyber-physical intrusion detection and response architecture to detect and respond to stealthy attacks. This generic architecture can utilize and evaluate multiple AMI data sources to detect stealthy attacks while reducing false positives. The proposed system, similar to McLaughlin et al. [7], can combine evidence gathered from various types of AMI-specific information sources, including 1) host- and network-based intrusion detectors on the cyber part; and 2) through non-intrusive load monitoring (NILM), power measurement-based abnormal consumption detectors are identified [7].

The remainder of the paper is structured as follows. The author introduces some of the related works. We present the system model and the proposed architecture in Section II and the performance evaluation and results in Section III. In Section IV, the author concludes the paper and presents future work.

A. Related Work

In [4], R. Andersen expresses worry about the potential for a malicious attempt to misuse the smart meters’ remote connect/disconnect feature, leading to a widespread blackout. The authors of [8] suggest a plan for integrating IDS sensors into AMI smart meters. This system's traffic monitoring coverage is extensive compared to the conventional centralized solution, which employs the utility server as an IDS. Attacks on meters that pass via these nodes on their route to the DCU will thus be detected early. In [9], Berthier et al. introduced Amilyzer, a NESCOR failure scenario-based AMI centralized specification-based intrusion detection system.

A connection is made between the rate of an attack's spread, the potential for its detection, and the effects on the total number of meters connected to the NAN introduced by Shawly et al. [5]. They also suggest SecAMI, an AMI open-source modeling and simulation tool [10]. Percentage describes how rapidly an attack may spread throughout NAN. In order to address the issue of how soon an IDS must react to an attack, they put forth an approach based on the features of NAN networks. In order to understand how the attack and response phases interact at a high level, they determine how much of the network can still be accessed from the DCU after an attack.

Many related works have recently used machine learning techniques to detect and respond to attacks on Cyber-Physical systems [11]-[23]. From a detection point of view, Hidden Markov Model (HMM) is a pioneering strategy in Machine Learning (ML) approaches for multi-stage attack detection and prediction. This method models the many phases of an attack, as HMM states [24]. For many reasons, Ourston et al.
[25] claim that HMM is the best detection method for these types of attacks. In order to analyze input and output interactions and create transition probability matrices based on training data, HMM first uses a manageable mathematical formalism. Then, it monitors the development of a multi-stage attack since it is tailored to cope with sequential data by utilizing transition probabilities between states. Unfortunately, current HMM-based detection systems mostly concentrate on single multi-stage attacks. None of the research tackles the issue of interleaving multi-stage attacks in CPS and how this interleaving affects detection and response performance.

Utilizing Markov Decision Process (MDP)-based intrusion response systems has grown in popularity in the last decade for network security. A game-theoretic definition of defense mechanisms for network intrusion detection has been presented [26]. However, the work does not present an automatic response mechanism based on damage containment.

A probabilistic IRS based on the MDP paradigm has been proposed [27]. Intending to ensure long-term system security, the model can plan the best set of response actions. An intrusion response method for cyber networks is suggested for utilizing the Partially Observable Markov Decision Process (POMDP) [28]. The suggested paradigm slows down an attacker's advancement while reducing the harm to system availability. One may track an attacker's development by embedding a state space on the condition dependency graph, an attack graph that represents dependencies between an attacker's skills and exploits. The defender's response is to prevent the attacker from using some of the exploits by blocking them. The attacker's hidden strategy, which relies on the defender's actions, directs how it moves through the network. The fundamental issue with all MDP-based models is the production of a vast state space, which makes the optimal response actions computationally exhaustive. Additionally, the aforementioned MDP-based intrusion response models do not address the threat containment issue for stealthy cyberattacks.

II. MATERIALS AND METHOD

This research presents an architecture for effective multi-level, multi-agent detection and response to stealthy attack scenarios on CPS. The suggested architecture, which is based on the HMM [24] and MDP [29] formalisms, can detect at any given time the presence of multiple organized attacks and provide insights into their dynamics, determining which ones are active and which attacks are inactive, the speed of each attack progressing, and the security state of each attack at any time. Knowing these results can help create suitable response mechanisms that lower the security risk of the network. Before the HMM processing subsystem, the proposed detection architecture de-interleaves mixed warnings from several assaults. It modifies the HMM model's parameters to detect multi-stage assaults when mixed alerts are based on the targeted network services.

By creating a database of $K$ HMM templates [30], [31], it is possible to generalize the current single-attack architecture, such as the one in Ourston et al. [25], to be able to detect multiple multi-stage attacks, $K$ attacks, and respond to them efficiently. Each of the HMM templates is connected with one of the $L$ MDP templates, $K \leq L$. In Fig. 2, we introduce the architecture for the threat detection and response processes that use such components. Each HMM-based template used here is intended to identify a certain kind of multi-stage attack that begins at an entry point (a node with an exploitable vulnerability) and ends with a specified critical node. Each $L$ MDP template is created offline and from all possible attack states and response actions within a service-based dependency graph.

The response level of this architecture uses a partly offline computation of MDP and the estimated current states from the HMM to determine network isolation boundaries or, from a different point of view, safe zones for critical components in the CPS (i.e., a set of vulnerabilities to be disabled). Fig. 3 illustrates the problem of the high dimensionality of safe-zones response, which prevents us from providing an online precise solution.

Fig. 2 Detection and Response Architecture
The design also contains an Intrusion Detection System (IDS), which uses network traffic under predetermined criteria to identify anomalous activity and provide real-time attack-related alerts. The IDS is a crucial component of the suggested design (e.g., Snort software [32] and Amilyzer [9]). Typically, an IDS generates a stream of alerts chronologically ordered by their timestamps. The set of IDS rules assists in lowering both the high volume of alerts and the number of false positive detections. The IDS can provide interleaved alerts for a single attack or several multi-stage attacks. These alerts can be pre-processed to generate observations that can be transmitted to the HMM database.

In order to improve predictions, the alerts are correlated and grouped by the alert’s clustering and pre-processing module, which also assigns different levels to the incoming alerts based on their severity, i.e., the severity of the alert increases with level, indicating that a continuous multi-stage attack is moving closer to an advanced stage. The detecting system groups the incoming alerts into a sequence of observations of length $T$.

Cyber Physical-Based System (CPBS) models generate Service Dependency Graphs (SDG) for a whole system and its missions. A Vulnerability Dependency Graph (VDG) component generates VDG templates for all SDGs and contains all known and possible multi-stage attacks. A VDG is created by integrating the SDG generated from a Cyber-Based System model with its vulnerabilities, as generated from the output of vulnerability scanners such as the Nessus vulnerability scanner. VDG templates can be generated from Nessus vulnerability scanner output and service dependency graphs. For each potential attack scenario, a Hidden Markov model (HMM) based template is produced from each VDG template. Based on the previous set of triggered IDS alerts, these templates are utilized to predict the attacker’s most likely attack path at each time instant.

A. Threat Model

The author considers $K$ distinct multi-stage attacks deployed concurrently in a network to compromise the system and stop $k$ services. The idea of potential vulnerabilities is captured by an edge coming out of previously exploited vulnerabilities in a VDG template.

Each attack begins at an entry point in the network (i.e., exploiting an initial exploit in a VDG template) and attempts to advance through the network by exploiting these vulnerabilities.

IDS-generated alerts about attacks are sent as an interleaved stream of alerts to the HMM database. These alerts may emerge from the intentional blending of numerous multi-stage attacks launched by a single attacker or may be produced at random by various attackers. The author assumes the HMM system processes $N$ alerts for each observation length ($T$). In particular, these $N$ alerts could, at any given time, come from a single attack or, at most, a combination of $K$ attacks.

![Fig. 3 Offline Computation of Safe Zone (SZ) Response vs. Online](image)

![Fig. 4 An Example of a CPBS Model](image)
III. RESULTS AND DISCUSSION

The author uses SecAMI [3], an open-source simulator to assess cyber-attack’s impact on AMI, to model stealthy attacks. SecAMI can be used to carry out two tasks: first, it can provide an undirected graph representing the topology of an AMI with data concentrators and smart meters; second, it can simulate an attack on the topology that compromises each node (smart meter) one at a time using three simulation parameters, namely hop time ($h$), compromise time ($c$), and detection time ($d$) of a compromised meter ($cm$). In the simulation, these settings control how quickly an attack spreads, concerning how quickly it is detected. The author presents detection experimental findings based on a stream of alerts produced by the SecAMI simulator based on Amilyzer IDS alerts [3], [9].

This paper expanded SecAMI to include a new module to construct the safe-zone-based response. The isolation module built on top of SecAMI creates multiple safe zones for a certain AMI topology using a parameter $k$, which represents the number of required safe zones, and a topology graph as inputs. The isolation module maintains both member information and the safe-zone components. The proposed response method uses the safe-zone information to pinpoint the safe zones with vulnerable components. The communication between the components of the attack safe zones and the rest of the AMI topology is immediately terminated.

A. Simulation

The attack and response models may be applied in one of two ways using the SecAMI simulator. SecAMI is used to determine whether there is a relationship between network availability, attack time, and detection time. In order to do this, a set of specified attack, detection, and reaction periods is input. The response actions in simulations start from the DCU since the author believes its security and resiliency are more than other meters. Some sample SecAMI attack and defense strategies are shown in Table 1.

<table>
<thead>
<tr>
<th>Sample Attack Strategies</th>
<th>Sample Response Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distributing malware.</td>
<td>Disconnect a suspect meter(s) from the network.</td>
</tr>
<tr>
<td>Getting root access on a meter.</td>
<td>Sending a remote re-keying command.</td>
</tr>
<tr>
<td>Compromising the DCU.</td>
<td>Rebooting a suspect meter(s).</td>
</tr>
<tr>
<td>Sending a command from the DCU to remotely disconnect a smart meter.</td>
<td></td>
</tr>
</tbody>
</table>

There are two levels in the simulation. The first level repeatedly cycles over the graphs, attackers’ entry points, and attack-to-detect-time ratios. The simulation is run using these settings in the second level to produce a dataset. It keeps track of upcoming activities, like when a specific meter is reached or compromised or when the simulation will learn that a node has been compromised (implemented as a priority queue). The first thing that should happen is for the attack’s entrance point to be breached. A spread event is scheduled for the node after it happens, and a detection event is prepared for the node where it occurs. There are no further occurrences after this point since either the attack has been stopped or the DCU has been hacked.

Below are the guidelines for the attack scenario and our reaction to an ongoing attack. Once attackers have gained access to a node, they can visit its neighbors and choose at random or specifically which neighbor to attack next. The attacker successfully compromises a new node after a specific number of hops. Note that the author modified SecAMI by calculating the likelihood that a susceptible node would be successfully compromised. Be aware that the severity of the vulnerability is also determined by the chance of compromising a node via the VDG. The attackers can choose a different node to compromise while still at the original one. The attacker can attack any node in range once a node has been compromised. An attack is identified after several hop times, the quickest path to the DCU, and a continuous detection duration. When a node is found to have been hacked, the NAN removes it from the service. Aside from that, any attack currently in the process from that node is halted, defending its intended target. The DCU must first acknowledge that one node has been hacked to stop the attack before it may corrupt further nodes. This is more likely to happen as the attack approaches the DCU. In this paper, the author looked into one SecAMI response to stealthy attacks in which a set of nodes is disconnected from the network using MDP to create a safe zone.

B. Results

SecAMI evaluation’s major objective is to answer issues such as whether a specific attack scenario is under control from a mitigation point of view. These answers are given in terms of detection and reaction times. In order to reduce dependence on a specific topology and obtain a more general conclusion on the class of graphs concerned, the author simulated many network topologies and averaged the findings across all graphs. In addition, the author averages the findings for a specific topology after simulating the attacker starting at every conceivable meter in the NAN.

With 100, 200, and 300 nodes, the author makes graphs. The author made 10 graphs with a maximum of 3.5, or 10 additional connections for each node level. The author particularly focused on the influence of the ratio between the time it takes to compromise a node and the time it takes to notice and react to a node compromise. The author launched 1, 3, and 5 attacks from 1, 3, and 5 random nodes in the graph simultaneously. For each of the degree combinations, the author produced results.

First, the author demonstrated the effects of the 1, 3, and 5 attacker(s) starting close to or far from the DCU. Removing a set node from the graph protects against these attacks. There are two factors at work here: the DCU can respond to attacks more quickly the closer they are to it because of the shorter transmission distance. However, if the attack starting point is near the DCU, it could do greater damage as more of the graph becomes disconnected. With a ratio of 1:1 related to attack to detection, the network availability is, on average, 24.3 percent greater for attacks starting near the DCU (Fig. 5). The author discovers that the benefit of mitigation time outweighs the increased risk. It is challenging for any node, especially one close to the DCU, to disconnect significant segments of the
graph in the case of removal because of the network’s ability to self-reconfigure (basically change connection).

After determining the threat’s location, the author considers the stealthy attacks starting at every meter for each multi-stage attack, and then the results are averaged (Fig. 6). On a single graph, the author combines the data for three of the nine scenarios. For example, 100 nodes and three connections, 200 nodes and five connections, and 300 nodes and ten connections. These were picked in order to display results that are typical of each node and connection category. No matter how big or linked the network is, the graph below demonstrates that for a significant amount (over 93%) of the network to be preserved, the attack time to detection time ratio must be at least 3:1. In other words, it must take no more than one-third of the time to detect and react as it takes an attacker to compromise a node. If the detection and response are quicker than this, a huge portion of the network is available without a strong spike occurring at this ratio.

This paper proposes a multi-level, multi-agent detection and response architecture for stealthy attacks on CPS. The main goal of the architecture is to detect stealthy attacks at an early stage and to stop the spread of damage efficiently. The proposed architecture is based on the formalizations of Hidden Markov Models (HMM) and Markov Decision Processes (MDP). The author evaluated the architecture’s performance using Advanced Metering Infrastructure, a critical CPS application (AMI). The simulation program, SecAMI, is used in this paper’s performance evaluation. The simulation results show how well the proposed architecture works to detect stealthy attacks on CPS and respond to them. In particular, the results show that the proposed system can preserve the availability of a high amount (more than 93%) of the CPS network in stealthy attacks.

Many obstacles must be overcome to create a strong security solution for CPSs. One challenge is to deal with IDS performance in real-time in order to modify, train, and compute online the HMM and MDP templates. Assessing the proposed detection and response mechanism for more sophisticated and stealthy attacks is another challenge the author plans to address.

**REFERENCES**


