Communication papers of the Federated Conference on Computer Science and Information Systems, pp. 323–332 ISSN 2300-5963 ACSIS, Vol. 13
DOI: 10.15439/2017F77

BAGS: A Tool to Quantify Smart Grid Resilience

Yatin Wadhawan
Department of Computer Science
University of Southern California
Los Angeles, USA
ywadhawa@usc.edu

Clifford Neuman
Information Science Institute
University of Southern California
Los Angeles, USA
bcn@isi.edu

***Abstract*—In this paper, we present the Bayesian Attack Graph for Smart Grid (BAGS) tool to quantify smart grid resilience in the presence of multiple cyber-physical attacks. BAGS takes system functions, network architecture, applications and a vulnerability report as input and generates three Bayesian Networks at three different levels of hierarchy. The top level network is called Functional Bayesian Network that defines how smart grid functions are connected. System engineers can select a particular function on a dashboard and view the Network Bayesian Network of that function at the second level. They can also choose a particular network component to see the list of vulnerabilities and the probability of associated compromise at the third level. System engineers can incorporate this functionality into their system and analyze the impact of any compromised component of the smart grid system on its resilience. Furthermore, BAGS helps to identify the failure paths in advance from one power grid function to another so that they can devise secure strategies and deploy resources effectively and efficiently.**

***Index Terms*—Smart Grid; Bayesian Network; Cyber Security; Cyber-Physical Security; Resilience; Industrial Control System**

NOMENCLATURE

S1: Quality of Smart Meter and Electric Vehicle Reads
S2: Quality of Smart Sync Head
S3: Billing System Performance
S4: Performance of Outage Management System
S6: Quality of Data captured by Vendor
S7: Performance of Electricity Energy Control Center
S8: Meter Data Management
R: Resilience

I. INTRODUCTION

CYBER-PHYSICAL Systems (CPS) [16] refer to a new generation of systems where physical processes are controlled and monitored from the cyber domain through advanced computation and communication technologies including humans in the loop. Utilization of advanced Information and Communication Technologies (ICT) in the CPS provides the ability for system engineers to control and monitor the physical processes in real-time. The growing interdependence between the cyber and physical world has opened the doors for various Cyber-Physical Threats (CPT) thus imposes extraordinary challenges on the security of CPS such as Smart Grid (SG). A Stuxnet-style attack on US SG could cost $1 trillion to US government [15]. According to the report, such high-profile attacks on CPS would be used to infect the electricity generation control rooms in some parts of the northern US by exploiting vulnerabilities in network components with a motive to control power generation. Recent multiple attacks on the Ukraine power grid [13] [14] indicate that cyber attacks on such critical infrastructures will be frequent. Hence, identifying, understanding and modeling CPTs and defining system's resilience in the presence of CPTs is now a necessity.

Researchers have focused on describing the resilience of the SG system by analyzing different types of attacks on one of its components/functions [5-9]. Although such efforts provide relevant insights about the security of the SG, they are incomplete in the sense that they do not consider the effect of one compromised function onto the other functions and ultimately on the system's resilience. Furthermore, the current work does not consider the dynamic nature of the vulnerabilities [7-9] and associated attack vectors. The likelihood of system component compromise is changing based on the dynamic nature of the vulnerabilities associated with the components and actions taken by the system engineers. For instance, we know the initial probability of compromising a server and that probability might change if we say that system engineers have applied security controls (patched) to that server. So we revise our belief based on current information, and this belief would be changed when there exists a zero-day attack for that server. Another challenge is that it is misleading to consider the vulnerabilities of the individual components [7]. It is possible that the vulnerability score of the single component is lower than the combination of multiple components, which provide the same functionality [4]. It is necessary to include the causal relationship between various functions, and components during risk analysis. It is insufficient to quantify and analyze Smart Grid Resilience (SGR) based on the static analysis of the system. We have to consider the dynamic nature of the vulnerabilities which is not yet considered by the previous approaches [7] [5]. They have failed to define a metric that can be used in real time to assess the SGR.

In this paper, we propose the **B**ayesian **A**ttack **G**raph for **S**mart Grid (BAGS) tool to quantify the SGR in the presence of multiple CPAs in real time. BAGS takes functions, network architecture, applications and a vulnerability report as input and generates three Bayesian Networks (BN). The top level network is called Functional Bayesian Network (FBN) that defines how SG functions are connected to each other, their probability of failure and connection with the resilience variables. The possibility of a function compromise is the joint

probability distribution of its network components that is based on the vulnerabilities of each component. FBN can be expanded to the second level as Network Bayesian Network (NBN) that can be further expanded to the third level as Vulnerability Bayesian Network (VBN). BAGS provides ease to the system engineers to perform an in-depth study of one of the functions of the SG and evaluates its effect on the overall system resilience. The system engineers can incorporate this functionality into their system, and they can see the impact of any compromised component of the SG on its resilience. The tool enables them to analyze how a failure of a network component controlling a particular power grid functionality propagates from the cyber to the physical domain and its impact on the SGR. It also helps them to identify the failure paths in advance from one SG function to another so that they can devise appropriate secure strategies and deploy resources effectively and efficiently.

The structure of the paper is as follows: Section II describes the related work. Section III discusses the challenges faced during SG risk analysis. Section IV and V describe the Resilience and Bayesian Network respectively. Section VI describes the SG architecture. Section VII explains the design of the BAGS tool. Section VIII describes the tool prototype and Section IX discusses the conclusion and future work.

II. Related Work

Traditionally, researchers have focused on modeling and analyzing the impact of CPTs on various functions and components of the SG system [5-9] [18-19]. Sanjab et al. [10] defined threats targeting the SG infrastructure, challenges involved in the understanding of CPAs and defensive strategies against such attacks. Neuman and Tan [8] described different types of CPAs and how they propagate from cyber to the physical domain and vice versa. Srikantha and Deepa [5] formulated a differential game that describes stealthy strategies for attackers to disrupt transient stability by leveraging control over Distributed Energy Resources (DERs). Researchers in [6] demonstrated an optimal attack scenario using false data injection attacks on Automatic Generation Control (AGC) functionality of the SG system. In [7], the authors defined the systematic approach to quantify the resilience of the SG system through PowerWorld simulation. The authors performed load drop attack by sending remote disconnect commands to smart meters at various locations and analyze its impact on the power system. Although such methods clearly demonstrate how power system gets affected because of changes in one of the functions of the SG due to the cyber attacks, they have failed to capture the dynamic behavior of attacks based on the vulnerabilities associated with the system components.

Findrik et al. [11] present a framework for the development and evaluation of secure and resilient SG control applications. The coupling between the communication system simulator OMNeT++ to the power system simulator PowerFactory using flexible middleware is used to develop the proposed tool. After describing the development of the tool, the paper modeled the cyber attacks in OMNeT++ simulator and analyzed the impact of those attacks on the power voltage control system. One of the limitations of this approach is that it does not include the real time vulnerabilities of the SG system into consideration. Suppose a zero-day vulnerability is discovered or vulnerability is patched in a network component. How this tool would incorporate this information in the modeling and how SGR is affected by that change? A risk assessment approach for power system considering the reliability of the information system is presented in [17]. Through simulations, authors demonstrated that the failure of the information system brings more risk to the power system operation. The line overload and risk of bus voltage out of range are defined and calculated as risk indices. Both software and hardware components are considered to quantify the indices. The paper provides insights into the impact of the information system failure on the SG system which was not performed earlier in the literature.

The concept of defining resilience using BN has frequently been used in securing engineered systems such as in [1-4] [18-19]. Li. et.al [1] described a three-layer framework that assesses the potential risks introduced by the mobile apps within the Android mobile system. The authors formulated three layers called as static, dynamic and behavioral network where risks are identified, and their propagation through each layer is modeled as Bayesian Risk Graph. Similarly, researchers in [4] modeled IT infrastructure using BN that enables system engineers to quantify the chances of network compromise. In [3], researchers defined a BN framework with a motive to compute the resilience of inland waterway network. Nita and Pingfeng [2] outlined a framework using BN to measure the resilience of engineered systems quantitatively. According to them, resilience is the combination of reliability and restoration. And it is similar to what is described in [3] that is absorptive, adaptive and restorative capacity.

We are motivated by the definition of resilience in [3] [4]. We have formulated a BN for the SG with a motive to quantify SGR in the presence of multiple CPAs in real time. Our research is different from current approaches (for assessing SGR) in a way that we have modeled the SG system as a dynamic BN where the probability to compromise system gets updated whenever there is a change in the vulnerabilities of the system. There is a need to develop a tool like BAGS so that system engineers not just can monitor the SG system components, functions and take decisions but can also control and compute the probability to compromise a particular function and analyze its effect on the SGR in real time. Furthermore, they should have the ability to perform an in-depth study from functional level to vulnerability level of any function of the SG system.

Note, this tool is not a replacement for intrusion detection and prevention systems or SCADA system, rather it works along with such systems. This tool provides better attack predicting capability and also enables system engineers to perform containment of compromised components to stop the propagation of attacks to other parts of the system.

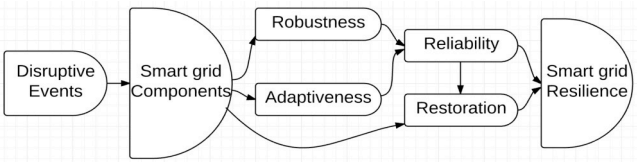

Fig. 1. Smart Grid: Resilience Modeling

III. CHALLENGES IN SMART GRID RISK ANALYSIS

SG is a federated system with multiple complex components. These components are interconnected to each other, and they focus on different functionalities of the SG system. Each federated system consists of various network components that have different attack vectors to compromise. Due to the presence of large number of attack vectors in different SG components and platforms, the risk assessment is challenging. The main challenges faced by the system engineers to perform risk assessment are as follows:

1. *Legacy System:* Traditionally, SG systems were developed not keeping security in mind. This makes them the hub of security vulnerabilities that can be exploited remotely through the internet. Furthermore, it is impossible to patch vulnerabilities altogether since it requires to shutdown the system during the patching process. Thus, it is hard to track and patch all the potential risks present in such federated systems.
2. *Complex Attack Vectors:* Due to large surface area of attack, there are large number of attack vectors present. The cyber attackers can perform multiple attacks by combining different attack vectors. For instance, they combine remote code execution at smart meters or buffer overflow in smart meter head server followed by a DDOS attack to prevent legitimate signals reaching physical units and finally, they send control signals remotely to smart meters to manipulate power supply. To avoid such attacks, it is important to develop a tool that can track vulnerabilities of all the system components in real time and alert in case of emergency.
3. *Risk Monitoring:* The wide variety of vulnerabilities are associated with the SG network components. These vulnerabilities are changing in real time. Thus, it is hard for the system engineers to perform risk monitoring. Each function in the SG consists of network components, which further comprise of various vulnerabilities. There is a need of a tool that combines the risks associated with the network components of different functions into one model and quantifies it.

IV. RESILIENCE

Resilience is the ability of a given system to avoid failures of its functions and components in the presence of the disruptive events and to quickly recover from those failures to an acceptable state without affecting function delivery. Fig. 1 shows the general concept of SGR modeling. To model the resilience of complex engineering systems different variables are defined in the literature [2-4]. We define those variables in the context of the SG system:

1. *Robustness:* It is the ability of the SG to withstand disruptive events. For instance, if there is a shortage of power during the natural disaster such as earthquake or Tsunami, the power grid continues to meet electricity demand by using power storage units. To provide robustness against CPAs, the system must have following capabilities in place: generation units, power storage, fault protection system, peaker power plant, gas storage, backup of the control system and critical workstations on standby mode. Such set of capabilities provides robustness to the infrastructure with a motive to meet power demand even in the presence of attacks.
2. *Adaptiveness:* It is the ability of the SG to adapt in the presence of changing environment so that to overcome disruptive events. For instance, if some generation units in the grid stop producing power, the power capacity is redistributed to other working generation units within the same area and neighboring areas to meet the demand. To provide the adaptive capacity to the SG, the system must have following capabilities: automation feeder and switching control, automatic generation control, automatic voltage control, phase angle regulator, real time load management such as demand response, and real-time load transfer. All such functionalities regulate power generation, transmission, and delivery even in the presence of attacks.
3. *Reliability:* It is the ability of the SG to continue to perform its functions normally. It depends on the Robustness and Adaptive capacity of the system which further depend on the system characteristics and disruptive events that take place on the system components.
4. *Restorative:* It is the ability of the SG to restore/recover from the disruptive events to an acceptable state where system is reliable to deliver its function. The restoration capacity depends on the reliability of the system and system characteristics in the presence of disruptive events. To provide the restorative capacity, the system must have following capabilities: automatic islanding and reconnection, containment of system components, fault current limiting, automatic software patching and hidden networks to operate. All such features enhance resilience of the system against various contingencies.

V. BAYESIAN NETWORK

In this section, we provide brief description of the BN and then, we describe how it is applied to the SGR modeling.

A. Bayesian Belief Network (BBN)

BBN is a probabilistic graphical model based on Bayes' theorem. It represents the conditional dependency between a set of random variables in a form of Directed Acyclic Graph (DAG) $G = (V, E)$. $V = \{V_1, V_2, ..., V_N\}$ is a set of nodes or variables of the system and E is the set of edges representing the dependencies among variables. A link $E_{i,j}$ from V_i to V_j represents the causal dependency between these two nodes. Here, V_j depends on the value of V_i and V_i is called parent (pa) of V_j. The relationship between variables of the BBN is measured using the conditional probability distribution (CPD). The joint probability distribution of N variables is:

$$P(V_1, \ldots, V_N) = \Pi_i^N P(V_i \mid V_{i+1} \ldots, V_N) \quad (1)$$

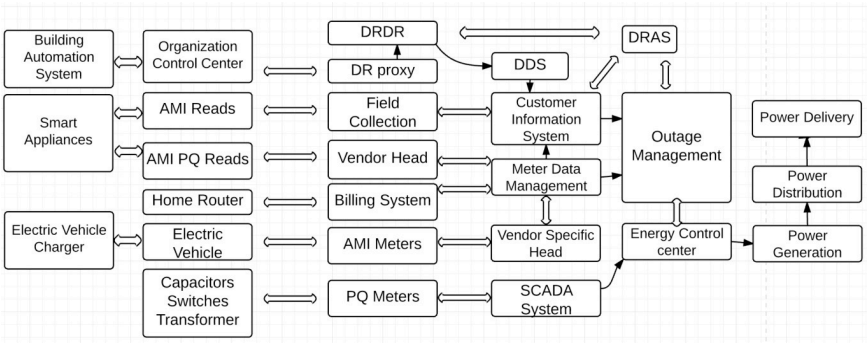

Fig. 2. Smart Grid Architecture

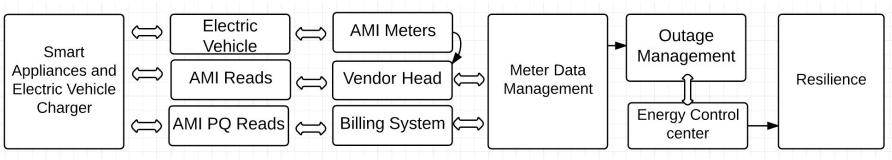

Fig. 3. Test Network

$$P(V_1, \ldots, V_N) = \Pi_i^N P(V_i \mid \text{Parents}(V_i)) \quad (2)$$

To calculate the joint probability distribution, the individual distributions and conditional distributions among parent and children must be predetermined. Such expectations are measured from data analysis, expert knowledge or the combination of both and by performing the simulation. The major advantage of using BBN is to compute posterior probabilities of an event when certain events are observed in real time. This is called belief propagation. For instance, the likelihood of the SG to be in the resilient state is updated when certain disruptive events are observed on some of its components. Now we discuss how this concept is applied to quantify resilience.

B. BBN Applied to Resilience Modeling

The probability of system resilience [2] is expressed regarding the probability of reliability and restorative capacity of the system. The likelihood of restoration depends on the likelihood of reliability and system characteristics when disruptive events are observed. The probability of reliability depends on robustness and adaptiveness of the system. The disruptive events can happen on different components of the SG. It depends on how system components are connected and how an event propagates from one part to other parts of the system. The joint probability distribution of the system according to Fig. 1 is defined as:

P (Resilience) = P (Disruptive Events) * (3)

P (Smart grid Components State | Disruptive Events) *

P (Robustness | Smart grid Components State) *

P (Adaptiveness | Smart grid Components State) *

P (Reliability | Robustness, Adaptiveness) *

P (Restoration | Reliability, Smart grid Components State) *

P (R | Reliability, Restoration)

VI. SMART GRID ARCHITECTURE AND TEST NETWORK

In this section, we provide brief description about the SG architecture and then, we discuss the test network used for experiments.

A. Smart Grid Architecture (SGA)

SGA (see Fig. 2) provides a full vision of the proposed system and ensures that minimum qualification of system requirements such as security management, network deployment, and policy implementation. It also identifies the key domain areas, functions, and their weaknesses. Thus, it is important to understand the SGA so that the system engineers can establish and implement security policies effectively.

Starting from the left-hand side, building automation system, smart appliances, and electric vehicles are the endpoints that consume power from the grid. Building automation system manages the power consumption of the smart buildings and interacts with the central control system which controls other parts of the power network; smart appliances refer to power consumption devices at homes that connect to smart phones, desktop or laptops and provide you more control and information remotely. All such components send readings about the use of power by each appliance and electricity quality delivered to endpoints to AMI meters. Further, AMI meters send the collected data to the field collection system, vendor head end, and billing system. At these places, data is processed and stored in relevant databases. The data will then be sent to Meter Data Management (MDM) system where the data is used for various functions such as power prediction, customer load profile management, power outages, power quality at different places, and billing customers. MDM connects to Customer Information System (CIS) which maintains all the customer's database. All the power quality readings are transferred to the SCADA where it monitors the overall power grid functionality.

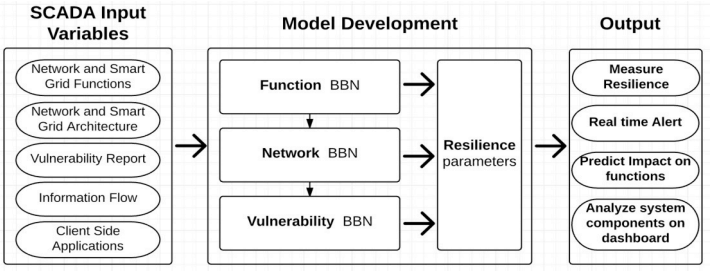

Fig. 4. Tool Design

Energy control center further manages power generation, distribution, and delivery. The data stored in the CIS and MDM, Outage Management System (OMS) predicts power outages and send information to the control center to manage power generation. Demand Response (DR) functionality is controlled through DR proxy. Many independent organizations are connected to DR proxy. DR proxy is further connected to DR Data Repository (DRDR). DRDR connects to DR Decision Support (DDS) and DR Automation Server (DRAS) that takes decision whether to perform DR functions in a given organization. DRAS interact with the OMS to check whether DR is required in response to reduced power generation. By understanding SGA, we know various power grid functions and components and how they are connected to each other. This understanding forms the basis of developing FBN in our model development of system design, which is explained in the Section VII-B.

B. Test Smart Grid Architecture

To develop a mock-up of our proposed design, we consider a test network in Fig. 3 which is a part of the SGA. The system consists of smart appliances and electric vehicle attached to the grid. The power consumption and electricity quality readings from these components are collected by the field collection systems, vendor specific heads, and billing system. Further, this information is communicated to MDM, and via MDM it is transferred to OMS to keep track of reserved power and power outages that might happen shortly. It also manages how much amount of power should be stored to meet the demand during contingencies. OMS interacts with the ECC to take decisions regarding power generation, DR, power delivery, etc. We develop the FBN, NBN, and VBN for this test graph and show how our proposed model is useful in quantifying resilience of the system. We explain the BN of the test network in the next section.

VII. TOOL DESIGN

In this section, we discuss the input variables given to the tool, model development and output of the tool (see Fig. 4).

A. Input Variables

1. *Network and SG functions:* It represents the list of higher level network and SG funcitons such as energy management system, outage management system, power generation, demand response, smart meters, billing function and vendor head end.
2. *Network and SG Architecture:* The network architecture of the system that supports all the functions of the SG system. For instance, MDM consists of servers, workstations, database, and communication network. It also represents the interconnection between all these components.
3. *Applications:* The list of client side and server side applications is also given as input to the system. It also includes vendor side applications which integrate to the SGA.
4. *Vulnerability Report:* To compute the likelihood of a particular system compromise, we must have a list of vulnerabilities that exist in system's components and applications. For instance, billing server has Cross Site Scripting (XSS), local file inclusion vulnerabilities, which can be exploited to gain control over the billing server. We will use Common Vulnerability Scoring System (CVSS) [12] scores to compute likelihood of compromise.

Once we have these inputs to the system, it generates three BN at three different levels of hierarchy, which we describe in the next subsections. The output of the model is displayed on an interactive dashboard where system engineers can select a particular function and can view its system architecture, components, and information flow. Furthermore, they can select a particular component and can view its vulnerability report and likelihood of its compromise. These features give power to the system engineers to perform real-time monitoring and predict the impact of a compromise on different components of the system and finaly, quantify the resilience of the system.

B. Function Bayesian Network (FBN)

FBN represents the causal interconnection between different functions of the SG. Nodes in the FBN describes the functions of the SG and edges represent the information flow from one function to another. It also describes how the impact of a system compromise travels across different functions. We use test network (in Fig. 3) to provide a better understanding of the FBN. Fig. 5 depicts the FBN of the test network. Smart meter function (S1) sends messages to the SG head function (S2) which further send messages to the billing system (S3). Similarly, electric vehicle charging data (consumption and power quality) is captured by the function (S5) and given to the Vendor particular head end (S6).

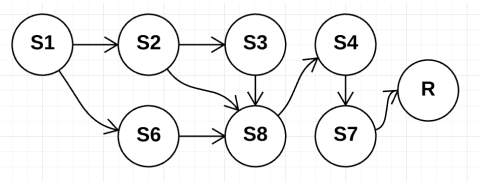

Fig. 5. Test Network: Function Bayesian Network

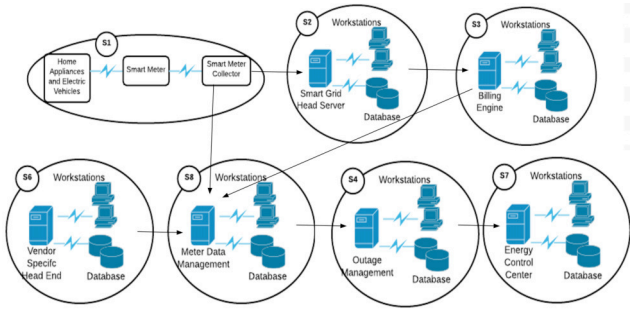

Fig. 6. Test Network: Network Bayesian Network

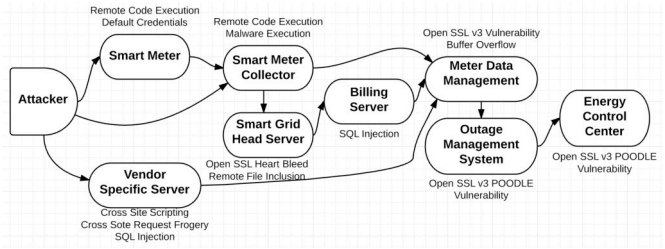

Fig. 7. Test Network: Vulnerability Bayesian Network

The data set from the billing system, head end and vendor specific end is provided to the MDM (S8). MDM connects with the OMS (S4), which is further connected to the ECC (S7). Finally, ECC is connected to Resilience function (R). Suppose an adversary compromises a smart meter by injecting a malware, he sends compromised messages about the power consumption all the way to the billing center and also exploits the vulnerability to gain control over the component. If he gains control over the billing system, he escalates privileges by exploiting vulnerabilities of other systems. FBN provides understanding to the system engineers that how information and control flow works in the SGA. They can also identify various attack paths an adversary may take if a certain function is compromised. In our model, they can choose any function on FBN and view its network components, which we describe in the next subsection.

C. *Network Bayesian Network (NBN)*

NBN represents various network components supporting a particular function. The system engineers can select a specific function at the FBN level on the dashboard and can see its network components and analyze its functional status and information flow. Fig. 6 describes the NBN of the FBN illustrated in Fig. 5. To illustrate with an example, the smart meter functionality (S1) has components: home appliances, electric vehicles, smart meters and smart meter collector server. The communication between these components is mostly wireless. Smart Meter collector server collects data from all the smart meters defined in its zone and sends that information to the smart grid head end (S2) over the wireless mesh network. Further, S2 consists of workstations, database servers and smart grid head server which collects data from various S1 systems. The significance of such modeling is to understand how data flows from one system to another, in other sense, whether there is any vulnerable path from one system to another. By providing such an interface, it provides the ability to analyze every system component and its impact on the overall system. In the next subsection, we describe that how an adversary compromises system vulnerabilities to move forward in the network. We have not shown firewalls and routers in this network diagram. We have displayed components that provide functionality to the SG and have vulnerabilities in their implementation. But we can extend this model to show other network components.

D. *Vulnerability Bayesian Network (VBN)*

The system engineers can select a particular network component from the NBN and view the list of its vulnerabilities by vulnerability report submitted as an input to the system. It also provides the information about the likelihood of network component compromise based on the CVSS score. The possibility of the components combined to calculate the compromise probability of the network component and ultimately of the function to which network components belong to. Also, if there is a change in the vulnerability, the system automatically updates the belief of system compromise and propagate to other parts of the network.

Attack Graph: Fig. 7. describes the attack graph from remote attacker's point of view according to our test case. A remote attacker performs variety of attacks to gain access to the SG functions. According to our test case, the attacker exploits XSS, CSRF or SQL Injection vulnerability to gain access to the Vendor specific server or performs remote code execution on smart meters or smart meter collector. Once the attacker has access to the smart meter collector, he exploits open SSL heart bleed to gain access to the smart grid head server from where he targets the billing engine server. The SQL Injection vulnerability can be exploited to gain remote access through username and password from the database. Once he gains access to the billing engine, he performs the port scan over the network range and identify the MDM server.

Once MDM server is identified, he exploits the buffer overflow vulnerability present in the server operating system. Then, the attacker gains access to the MDM directly without going through billing engine. Once he gains access to the MDM server, he performs the variety of attacks such as integrity attack by changing meter readings, but in this case, he is interested in having access to the energy control center. So he further performs scanning and identifies the OMS server which is connected to the MDM server. He exploits the open SSL v3 POODLE vulnerability and gets root access. Once he has access to the OMS server, he identifies and attacks the ECC by exploiting open heart bleed vulnerability. Table I indicates all the vulnerabilities corresponding to different servers in our test network.

CVSS: Our motive is to compute the local probability distribution of system compromise based on the vulnerabilities of the system. For that, first, we need to calculate the likelihood of success for an attacker to exploit the known vulnerability. We use the process described in CVSS [12] to calculate this probability. CVSS score consists of three different scores: Base, Temporal and Environmental. The base score includes the essential properties of the vulnerability such as Attack Vector (AV), Attack Complexity (AC), User Interaction (UI) required or not, Privileged Required (PR) or not, affecting Confidentiality, Integrity, and Availability (CIA). The temporal score indicates the characteristics of the vulnerability that evolve over its lifetime. It includes variables like exploit code maturity, remediation level, and report confidence. And finally, the environmental score includes characteristics that are dependent on the implementation and environment of the organization. It includes variables such as the requirement of CIA and modified scope. We compute the probability of exploiting vulnerability by considering exploitability score as described in CVSS specification [12]:

$$P(V_i) = (8.22 * AV * AC * UI * PR)/10 \quad (4)$$

Table II represents the CVSS probabilities of compromise corresponding to each vulnerability. The local conditional probability distribution (LCPD) is computed when many exploits are involved. It depends on whether each vulnerability exploitation is a distinct event or not. If it is a distinct event, we compute LCPD using product rule (see eq. 5) otherwise we calculate joint probability for OR case (see eq. 6). Suppose an adversary compromises a known vulnerability of the system (X) and earns a user privilege (Y) to that system. It forms a causal relationship as X -> Y.

TABLE I. TEST SGA VULNERABILITIES

System Name	Vulnerabilities
Smart Meters	Default Password Remote Code Execution
Smart Meter Collector Server	Remote Code Execution
Smart Grid Head End (Windows Server 2008)	Open SSL Heart Bleed
Billing Engine Server (Windows Server 2008)	SQL Injection
Meter Data Management Server (Windows Server 2012)	Open SSL v3 Vulnerability XSS client end
Outage Management Server (Windows Server 2012)	Open SSL POODLE Vulnerability
Energy Control Center Server (Windows Server 2012)	Open SSL POODLE Vulnerability.
Vendor Specific Server (Windows Server 2012)	Cross Site Request Frogery Cross Site Scripting (XSS) SQL Injection
Workstations	Buffer Overflow
MS SQL Server	SQL Injection MS SQL Buffer Overflow

TABLE II. EXPLOITABLE PROBABILITIES BASED ON CVSS

Vulnerabilities	CVSS Probability
Remote code execution	0.84
Buffer Overflow	0.78
Denial of Service	0.74
SQL Injection MS SQL Server	0.72
Open SSL Heart Bleed	0.75
Open SSL POODLE	0.31
Cross Site Scripting	0.61
Cross Site Request Frogery	0.88

$$P(Y \mid Parents(Y)) = \Pi_i^N P(Exploit_i) \quad (5)$$

$$P(Y \mid Parents(Y)) = 1 - \Pi_i^N (1 - P(Exploit_i)) \quad (6)$$

The probability of compromise of Y depends on the vulnerabilities that Y has and the likelihood of X to get compromised. X is the parent of Y. In case of multiple exploits that need to be exploited, the probability is computed via product rule. We compute the product of all the probabilities of exploits that are present in the parent of Y. LCPD is used to calculate the unconditional probability distribution corresponding to each vulnerability, by merging the marginal cases at each node. For instance, the unconditional probability of compromising billing server depends on all the nodes that influence it (smart meter, smart meter collector and smart grid head server). Similarly, the unconditional probability of the ECC server depends on all the nodes shown in the graph Fig. 7. The probability whether remote attacker attacks the SG system components is provided by the system engineers based on their experience and they revise their belief over a period. In our tool, the system engineers can change this probability to see the impact on the resilience of the system.

Posterior Probabilities: The probability to compromise a function changes over a period depending on the vulnerabilities of its network component's and other factors. The posterior probabilities of system components are useful to evaluate the risk in the dynamic environment. For example, if we know that OMS is compromised, we can calculate the likelihood of MDM compromise using Bayes rule (eq. 7):

$$P(MDM/OMS) = P(OMS/MDM)\, P(MDM) / P(OMS) \quad (7)$$
$$= 0.99$$

We already know the value of P(MDM)=0.4786, P(OMS)=0.1484 (see fig. 12.) and P(OMS/MDM) = 0.31. The unconditional probability of MDM getting compromised was 0.4786. But once we know that an attack incident at OMS, the posterior probability becomes 0.99. Similarly, the system engineers can calculate probabilities of successors of MDM and other nodes in response to an attack incident on OMS. Such technique allows the system engineers to see how the effect of an attack propagates to other parts of the system. For

instance, when the individual system gets outdated and has more vulnerabilities exposed to the outer world, the attacker has larger surface area to compromise the system, and that will affect the probability of compromise of other nodes as well. Similarly, the system engineers can also evaluate how attack surface area is reduced when a particular security control is placed on the node or vulnerabilities are reduced by updating the software. They can develop resource allocation algorithm to minimize the risk using posterior probabilities.

Another advantage of this approach is to partition the network by compromised probabilities. Once the likelihood of a component crosses a threshold, that component is either fixed immediately or removed from the network to reduce its effect on other parts of the system. The tool provides the ability to calculate posterior probabilities and show them on the dashboard so that system engineers can monitor the status of each network component.

E. Tool Output

The main motive of this tool is to provide following functionalities:

1. *Measure Resilience:* The primary motive is to measure the resilience of the system in real time. By considering the vulnerabilities of the system components, the likelihood of their compromise is calculated. Using BN, we connect different system components and see how an attack propagates from one system to another. The resilience is computed based on the probability of likelihood of ECC compromise, which controls the power generation and distribution. If the probability of ECC compromise is high, the system is not resilient, and power delivery will be affected in case of an attack. As described in section IV, resilience is dependent on reliability and restorative measures. Due to lack of SG data, we only consider reliability. The probability of the ECC compromise works as an identifier of the resilience of the system.
2. *Alert Mechanism:* The alert mechanism helps system administrator to put check points on the probabilities of the system component compromise. Based on their knowledge of the system, they assign probability thresholds to each component described in the test network. If the probability of the system component compromise crosses threshold, an alarm is raised. This enables system engineers to identify most of the vulnerable components so that they can assign appropriate security controls and perform vulnerability assessment and penetration testing.
3. *Predict Impact:* The system engineers evaluate and predict the impact of a system component compromise on another components by computing posterior probabilities.
4. *View System Architecture:* The system engineers view the whole system on an interactive dashboard. The interactive dashboard provides a view of 1) FBN where all the system functions are logically connected, 2) NBN which is a detailed description of the function components and 3) VBN which describes the list of vulnerabilities associated with the components and probabilities of their compromise (unconditional probabilities) and attack graph. This enables system engineers to analyze the status of the SG system remotely, and they can perform impact analysis by simulating different attack scenarios on different components of the system.

VIII. TOOL PROTOTYPE

In this section, we describe the simulation setup and the mock-up of the tool.

A. Tool Development

We have developed the User Interface (UI) of the tool in Java language using regular window toolkit class. It represents the framework that is visible to the system engineers on the dashboard. We maintain the static files database of the network components, functions and vulnerabilities (described in Section VII-A) as input to the tool. The tool parses the file and generates the FBN (see fig. 8.). In this mock-up, the lines are not directed. But in the real tool, the lines will be directed and marked by the type of information they represent. Note, here, our motive is to demonstrate a mock-up/UI of the tool that shows how powerful it is.

FBN represents the acyclic graph of the connected components according to the test network (see fig. 3 and 5). When a user clicks on a node of FBN, the function's network components are represented and how they are connected (see fig. 9.). When the user clicks on a particular network component, the list of vulnerabilities associated with that component is generated (see fig. 10.). It also contains the details of the vulnerabilities, CVSS score and the probability of compromise. The system engineers can change the system configuration in the database consistently. The same changes are reflected in the tool on refreshing it. The system engineers can add or remove any component, discover or patch any vulnerabilities, and disconnect any component.

In this mock-up, we use the Bayes.jar to represent the unconditional probabilities of the functions. We create nodes corresponding to each functions and link them according to the test network (see fig. 11). Then, we provide the probability of compromise to each component. It gives the unconditional probabilities of compromise of each function as output. The system engineers view the expectations by selecting any function on the dashboard (see fig. 12.).

Fig. 8. Tool prototype: FBN.

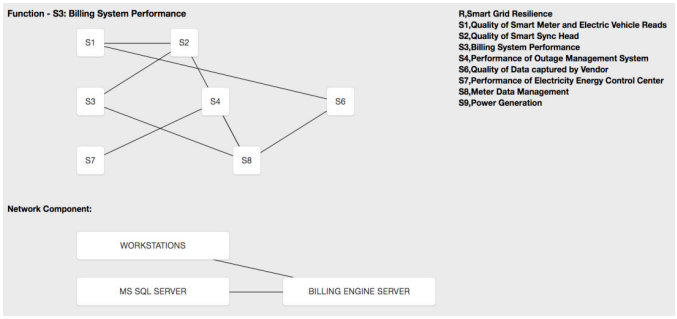

Fig. 9. Tool Prototype: NBN

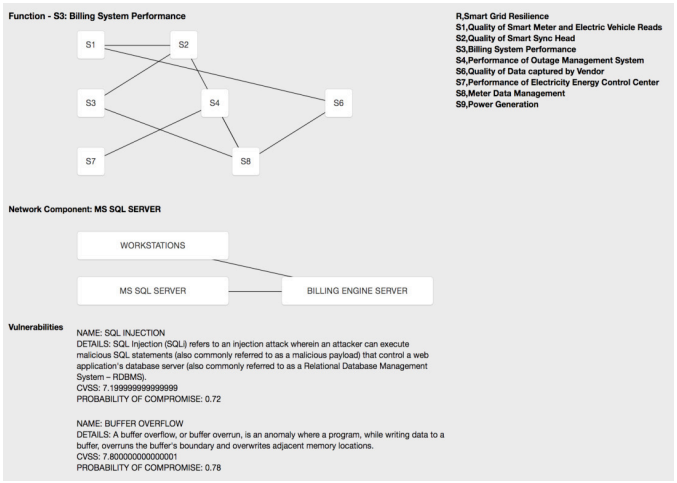

Fig. 10. Tool Prototype: VBN

In the future, we will implement the Bayesian algorithm in our tool. Over here our motive is to show the features of the BAGS tool available to the system engineers. In fig. 8, S7 stands for the performance of ECC. This is the main component of the SG system. If ECC gets compromised, the power system may get shutdown. If this system has high probability of compromise, the overall system is at high risk. The cyber attackers can easily exploit the vulnerabilities and compromise S7. Functions that fall under the category of reliability, restoration, and adoption will be considered in our future work. Due to the limitation of space and the absence of real world dataset, it is difficult to demonstrate how they are modeled.

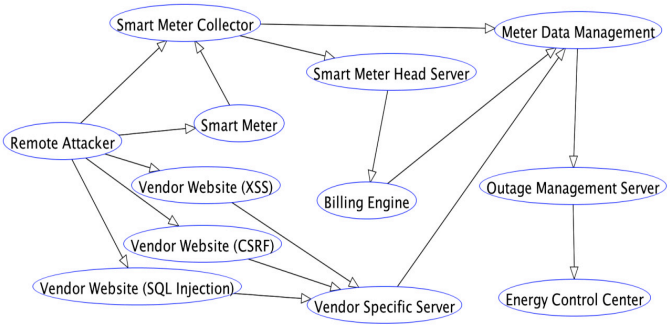

Fig. 11. Bayes.jar tool Function Nodes.

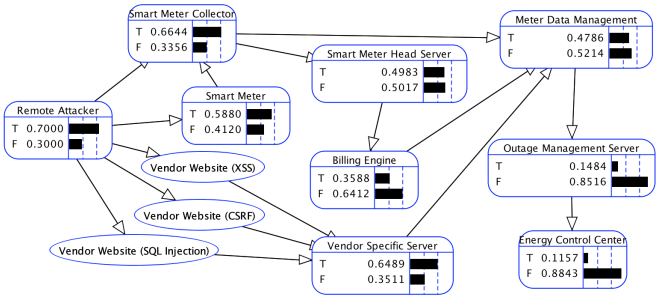

Fig. 12. Bayes.jar Unconditional Probability Distributions when Probability of remote attacker to attack is 0.70.

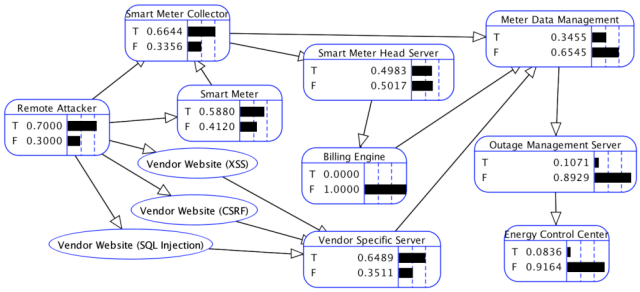

Fig. 13. Bayes.jar Unconditional Probability Distributions when Billing Engine's SQL Injection vulnerability is patched. The effect of such change is propagated to other components.

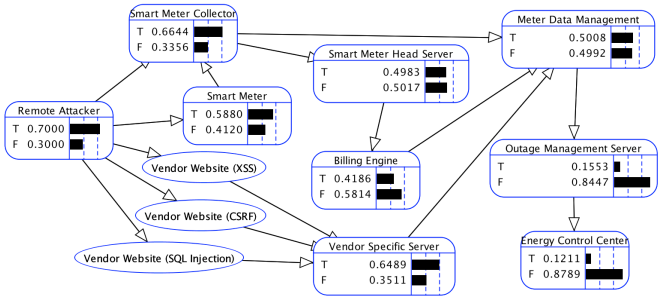

Fig. 14. Bayes.jar: Unconditional Probability Distributions when remote Code Execution is vulnerability is discovered in Billing Engine server. The effect of such change is propagated to other components.

B. *Results*

Fig. 12. represents the unconditional probabilities of all the functions of the test network. Each function is a Bernoulli variable. True (T) variable represents the likelihood of compromise, and False (F) represents the probability of not compromise. They are computed by Bayes rule described in equation 2. Also, they are used to calculate the posterior probabilities of the components to get compromised. The system engineers can easily monitor the status of the components regarding compromise probabilities by analyzing the graph in fig. 5. If any vulnerability is discovered in a component or if any vulnerability is patched, the unconditional probabilities will automatically get change.

Fig. 13. represents the case when system engineers have patched the billing engine server, and its probability of

compromise becomes zero. There is a drastic change in the likelihood of compromise of other components which are children of billing engine server. Similarly, fig. 14 represents when the zero-day vulnerability is discovered in the billing engine server and how the probabilities to compromise of its children change. This is how BAGS enables engineers to evaluate the risk associated with every component and how risk propagates from one component to another in such interdependent network. Based on the experience of the system engineers, they set a threshold on the unconditional probability of any function and create an alert mechanism. This feature will also be implemented as part of the future work.

IX. Conclusion and Future Work

In this research paper, we presented the Bayesian Attack Graph for Smart Grid tool to quantify the resilience of a given smart grid system in the presence of multiple cyber-physical attacks in real time. BAGS takes system functions, network architecture, applications and vulnerability report as input and generates BNs at three different levels of hierarchy: FBN, NBN, and VBN. We have implemented a mock-up/UI of this tool by maintaining a static database of network architecture. The system engineers can incorporate this functionality into their system, and they can see the impact of any compromised component of the smart grid system on its resilience. BAGS enables system engineers to analyze how a failure of a cyber network component controlling a particular power grid functionality propagates from the cyber to the physical domain and its impact on SGR. It also helps them to identify the failure paths in advance from a smart grid function to another so that they can devise relevant secure strategies and deploy resources effectively and efficiently. BAGS works along with the intrusion detection and prevention systems or SCADA system. It provides better attack predicting capability and ability to perform containment of compromised components to stop the propagation of attacks to other parts of the system. The system engineers can feed this input to their dashboard of Security Operations Center to expedite the process of security risk assessment. Note, one can develop this tool to quantify the resilience of other CPSs such as oil and gas systems, nuclear plants, water treatment plants, etc. For that, they need to understand and incorporate the system's cyber and physical infrastructure. Our future work will be to develop this tool using the real world dataset and deploy it in a SCADA system. We are in the process of discussion with companies that may provide SGA data to us. We will also implement Bayesian Algorithm and add features such as alert mechanisms and posterior probabilities as a part of our tool. Furthermore, we will develop an algorithm to allocate security controls to maximize the resilience of the system using game theoretic approaches.

Acknowledgement

This work was conducted with partial funding by Northrop Grumman Information Systems through the Cyber Security Research Consortium, and with support from Schlumberger.

References

[1] Li, S., Tryfonas, T., Russell, G., & Andriotis, P. (2016). Risk assessment for mobile systems through a multilayered hierarchical Bayesian network. IEEE transactions on cybernetics, 46(8), 1749-1759.

[2] Yodo, N., & Wang, P. (2016). Resilience modeling and quantification for engineered systems using Bayesian networks. Journal of Mechanical Design, 138(3), 031404.

[3] Hosseini, S., & Barker, K. (2016). Modeling infrastructure resilience using Bayesian networks: a case study of inland waterway ports. Computers & Industrial Engineering, 93, 252-266.

[4] Poolsappasit, N., Dewri, R., & Ray, I. (2012). Dynamic security risk management using bayesian attack graphs. IEEE Transactions on Dependable and Secure Computing, 9(1), 61-74.

[5] Srikantha, Pirathayini, and Deepa Kundur. "A DER Attack-Mitigation Differential Game for Smart Grid Security Analysis." IEEE Transactions on Smart Grid 7, no. 3 (2016): 1476-1485.

[6] Tan, Rui, Hoang Hai Nguyen, Eddy YS Foo, Xinshu Dong, David KY Yau, Zbigniew Kalbarczyk, Ravishankar K. Iyer, and Hoay Beng Gooi. "Optimal False Data Injection Attack against Automatic Generation Control in Power Grids." In 2016 ACM/IEEE 7th International Conference on Cyber-Physical Systems (ICCPS), pp. 1-10. IEEE, 2016.

[7] AlMajali, A., Rice, E., Viswanathan, A., Tan, K., & Neuman, C. A systems approach to analysing cyber-physical threats in the Smart Grid. In 2013 IEEE International Conference on Smart Grid Communications (SmartGridComm).

[8] Neuman, C., and Tan, K.: Mediating cyber and physical threat propagation in secure smart grid architectures. In Smart Grid Communications (SmartGridComm), IEEE International Conference on (pp. 238-243). (2011)

[9] Yi, Ping, Ting Zhu, Qingquan Zhang, Yue Wu, and Li Pan. "Puppet attack: A denial of service attack in advanced metering infrastructure network." Journal of Network and Computer Applications 59 (2016): 325-332.

[10] Sanjab, A., Saad, W., Guvenc, I., Sarwat, A., & Biswas, S. (2016). Smart Grid Security: Threats, Challenges, and Solutions. arXiv preprint arXiv:1606.06992.

[11] Findrik, M., Smith, P., Kazmi, J. H., Faschang, M., & Kupzog, F. (2016, November). Towards secure and resilient networked power distribution grids: Process and tool adoption. In *Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on* (pp. 435-440). IEEE.

[12] CVSS Score. https://www.first.org/cvss/specification-document

[13] Ukraine's Power outage was a cyber attack: Ukrenergo, 2017. http://www.reuters.com/article/us-ukraine-cyber-attack-energy-idUSKBN1521BA

[14] Analysis of the Cyber Attack on the Ukrainian Power Grid, March 2016. http://www.nerc.com/pa/CI/ESISAC/Documents/E-ISAC_SANS_Ukraine_DUC_18Mar2016.pdf

[15] Stuxnet style attack on US Smart Grid could cost governemnt $1 trillion. https://www.scmagazineuk.com/stuxnet-style-attack-on-us-smart-grid-could-cost-government-1-trillion/article/535452/

[16] Baheti, Radhakisan, and Helen Gill. "Cyber-physical systems." *The impact of control technology* 12 (2011): 161-166.

[17] Lu, D., Liu, Y., & Zeng, Y. (2016, November). Risk assessment of power grid considering the reliability of the information system. In *Smart Grid Communications (SmartGridComm), 2016 IEEE International Conference on* (pp. 723-728). IEEE.

[18] J. Zalewski, S. Drager, W. McKeever, A. Kornecki, B. Czeido, Modeling Resiliency and Its Essential Components for Cyberphysical Systems, Annals of Computer Science and Information Systems (Proc. FedCSIS'2015), Vol. 6, 107–114 (2015)

[19] A. Kornecki, N Subramanian, J. Zalewski, Studying Interrelationships of Safety and Security for Software Assurance in Cyber-Physical Systems: Approach Based on Bayesian Belief Networks, Proceedings of the 2013 FedCSIS Conference, pp. 1381–1387.