A Framework for Evaluating Intrusion Detection Architectures in Advanced Metering Infrastructures

Alvaro A. Cárdenas, Robin Berthier, Rakesh B. Bobba, Jun Ho Huh, Jorjeta G. Jetcheva, David Grochocki, and William H. Sanders

***Abstract*—The scale and complexity of Advanced Metering Infrastructure (AMI) networks requires careful planning for the deployment of security solutions. In particular, the large number of AMI devices and the volume and diversity of communication expected to take place on the various AMI networks make the role of intrusion detection systems (IDSes) critical. Understanding the tradeoffs for a scalable and comprehensive IDS is key to investing in the right technology and deploying sensors at optimal locations.**

This paper reviews the benefits and costs associated with different IDS deployment options, including either centralized or distributed solution. A general cost-model framework is proposed to help utilities (AMI asset owners) make more informed decisions when selecting IDS deployment architectures and managing their security investments. We illustrate how the framework can be applied through case studies, and highlight the interesting cost/benefit trade-offs that emerge.

***Index Terms*—AMI, threat model, intrusion detection, architecture.**

I. Introduction

The protection of power grid infrastructures against computer attacks is a matter of national security, public safety, and economic stability, but in many countries, the majority of these critical assets are owned and operated by private companies with pressing operational requirements, tight security budgets, and aversion to regulatory oversight. For most of these private stakeholders, creating a business case for improving computer security and supporting long-term security research is a difficult task, partly because cybersecurity risk is challenging to quantify. As stated in the Roadmap to Achieve Energy Delivery Systems Cybersecurity released by the U.S. Department of Energy (DoE) [1], "Quantifying risk is problematic when the energy sector faces rapidly changing threats that are difficult to predict and have consequences that are hard to demonstrate."

In order to help utilities better manage their cyber-security risks, DoE, in cooperation with the National Institute of Standards and Technology (NIST) and the North American Electric Reliability Corporation (NERC), has developed a cyber-security Risk Management Process (RMP) tailored to smart grids [2]. In this paper we leverage information from RMP and other security-risk management frameworks for Intrusion Detection Systems (IDSes) [3] and apply them to the specific case of AMI networks in which asset owners are evaluating the use of an IDS as part of their security controls.

We focus on AMI networks because security investments in this field need to consider multiple potential threats. Smart meters are low-cost commodity devices, operating in physically insecure locations and with an estimated lifetime of several decades [4]. Therefore, while some basic protective measures have been developed (tamper-evident seals, secure link communications), they may not be enough to prevent successful attacks during the lifespan of smart meters. As the scope of applications using AMI networks increases, it has become critical to design and deploy efficient security monitoring solutions.

Although there has been a great deal of work on intrusion detection, including distributed IDSes, for wireless networks (e.g., [5] and references therein), there has been little work studying the practicality of deploying distributed IDS devices, and in comparing the possible architectures for this deployment.

We consider the point of view of a utility company that needs to create a business case for improving their security posture by introducing an IDS in their AMI network. We formulate the problem with the following sequence of steps:

1) First, the threat to AMI systems has to be understood. This has been well-explored in the literature in [5], [6], [7], [8], and [9].
2) We offer the first consideration of the trade-offs associated with centralized and distributed IDS, which we further subdivide into IDS sensors embedded in smart meters, and IDS sensors based on dedicated devices. We then analyze the practical advantages and disadvantages of each possible IDS deployment.
3) Finally, we construct a decision framework cast as a risk-assessment problem. We first evaluate the cost-effectiveness of each possible deployment relative to AMI network density and coverage area. Those general conclusions do not rely on specific values. We end our analysis by considering specific case studies of attacks and estimated costs.

We find that while a centralized IDS architecture provides sufficient protection against some security threats, a distributed IDS is essential for some AMI applications. In addition, a distributed IDS is an effective tool for timely discovery and thus rapid and low-cost recovery from attacks.

This paper is structured as follows. Section II provides an overview of AMI. Section III discusses related work. We describe three different IDS architectures for AMIs in Section IV, and discuss the network visibility they provide as well as their deployment costs. Section V then describes a risk-assessment framework that can help utilities make well-informed decisions when choosing an IDS architecture to install. Section VI

Alvaro A. Cárdenas and Jorjeta G. Jetcheva are with Fujitsu Laboratories of America, Inc. and can be contacted at {alvaro.cardenas-mora,jjetcheva}@us.fujitsu.com

Robin Berthier, Rakesh Bobba, Jun Ho Huh, David Grochocki, and William H. Sanders are with the University of Illinois at Urbana-Champaign and can be contacted at {rgb,rbobba,jhhuh,dgrocho2,whs}@illinois.edu

illustrates use of the framework. Conclusions and future work are summarized in Section VII.

II. AMI Overview

The role of an AMI is to enable communication between utility companies and electricity meters, including remote electricity usage readings (on-demand and periodic), electricity price information, alerts about outages, and upgrades of meter firmware, among other communications. Some messages require real-time delivery, while others can be buffered and delayed without negative consequences. In addition, AMIs have security and privacy requirements, since sensitive customer information is frequently exchanged, and some of them provide a remote disconnect feature. To accommodate the aforementioned requirements and also a wide range of meter deployment topologies, e.g., from dense urban settings to sparse rural environments, meter manufacturers have designed highly flexible network architectures that can include different communication media. Those architectures usually follow the same network hierarchy, with a wide area network (WAN) connecting utilities to a set of gateways in the field, and then neighborhood area networks (NANs), also called *field area networks* (FANs), connecting gateways to meters. Meters themselves can be used as gateways to access the home area network (HAN) deployed within customer premises to connect to thermostats and smart appliances. A WAN uses long-range and high-bandwidth communication technologies, such as long-range wireless (e.g., WiMAX), cellular (e.g., 3G, EVDO, EDGE, GPRS, or CDMA), satellite, or Power Line Communication (PLC). NANs typically have shorter range requirements and can be deployed using wireless (e.g., IEEE 802.11, IEEE 802.15, or proprietary communication stacks) or PLC-based technologies. In some cases, meters can directly include cellular capabilities or even use the customer's home Internet connection to bypass the need for separate WANs and LANs. In this paper, we focus on NANs that use a wireless mesh network. The mesh topology brings robustness to the network, since communication routes can automatically adapt when failures occur. However, they also represent a challenge for the deployment of an efficient security monitoring solution, due to their distributed nature, and their use of wireless communication technologies.

III. Related Work

The continuously growing threat landscape of AMI has attracted research on threat characterization and mitigation. [10] presents the design of a firewall to secure wireless communication in energy delivery systems. [8] examines attacks targeting energy theft in AMIs; the authors later used that analysis to motivate a new methodology for penetration testing in AMIs [7]. [5] identifies a set of IDS requirements for AMI and briefly mentions different sensor deployment locations, including dedicated and meter-level sensors. [6] expands on that work to develop a specification-based IDS specifically targeting attacks in HANs. Faisal et.al. [11] studied the effectiveness of stream mining algorithms for intrusion detection in the context of the limited resources available in smart meters (memory and space). Their main objective was to evaluate stream mining in an embedded architecture. They did not consider a dedicated or a centralized architecture in their approach, nor did they study the cost-benefit tradeoffs between these different architectures.

Zhang et al. [12] proposed a hierarchical IDS framework, consisting of detection modules placed strategically throughout the smart grid to monitor HAN, NAN, and WAN communications. Lower-level modules would first attempt to mitigate detected attacks before elevating alerts to a high-level nodes, which could be more effective having increased awareness over the network. In Mohammadi et al. [13], the authors discussed smart grid communication requirements and security concerns in regards to the unique aspects of smart grid architectures. They also proposed a hierarchical IDS solution for NANs that utilized a combination of anomaly and signature-based approaches to intrusion detection. Unlike the previous work, their architecture relied on a central IDS to make decisions when malicious activity is detected. However, both of these works also did not consider cost in their analysis, and while they do both present a possible solution to intrusion detection in AMI, they fail to consider other deployment options. In this paper, we build on the contributions from [9], which covers threats prevalent in the AMI mesh network, along with several sensor deployment ideas, suggesting that a hybrid approach would provide the widest monitoring coverage.

This paper relates also to the concept of cost model for selecting different IDS architectures depending on the deployment characteristics. [3] proposes a methodology for analyzing the cost benefit trade-offs in network IDSes. Their cost model, however, is implemented on a specific cooperative principle-based network IDS, and does not consider different sensor deployment locations. Some characteristics that are important to the AMI, like the coverage area, are also not considered. [14] introduces a cost model for AMI in the context of automated response and recovery actions. The model focuses on converting impacts of attacks and responses on the integrity, availability, and confidentiality attributes into financial values through the analysis of service level agreements. The main difference with our work is on the level of granularity of our cost model. [14] looks at system-wide costs while we focus on detailed attack steps and we study detection time. While these efforts have been important in shedding light on the security issues and solutions surrounding wireless mesh networks and AMIs, to the best of our knowledge, there has not been a detailed and extensive guide to the design of a comprehensive security monitoring solution.

IV. IDS Deployment Schemes

A. Information Required for Detection

An understanding of the kind of information required for detecting intrusions is crucial to the design of a comprehensive and cost-efficient monitoring and intrusion detection solution as the type and placement of sensors is dependent on this. In [5], Berthier et al. organize the information required for detecting attacks against AMIs into the following three categories:

- *System information*: health reports from meters, and gateways (CPU, battery consumption), firmware and software integrity of AMI devices, clock synchronization.
- *Network information*: NAN collision rate, packet loss, node response time, traffic rate, health and integrity of routing table, associations between physical addresses and node identity.
- *Policy information*: Authorized AMI protocols, authorized AMI devices, authorized traffic patterns, authorized route updates, authorized firmware updates.

The above categorization reveals that different types of data must be collected and from different locations in the infrastructure. For example, the need for information on health and integrity of routing tables requires routers (in this case, meters) to be instrumented so that they can send periodic health reports or at least be remotely queried for health and integrity checks. However, instrumentation of all routers in the network may be too expensive, and it could more cost-effective to rely on attack manifestations at other locations in the system instead of routers for detection. In the rest of this section we will discuss different intrusion detection architectures that result from trade-offs among the types and placement of sensors based on information needed for detection.

B. Centralized IDS Trade-Offs

A centralized monitoring architecture can be located at the utility data center, where smart meter data are processed and stored. In that type of deployment, the IDS scheme will only be able to analyze network traffic to and from the AMI network; peer-to-peer traffic between nodes in the AMI network will not be visible to it. This architecture is in line with typical enterprise security controls, in which data transmitted between two different networks (Intranet-Intranet or Intranet-Internet) can be monitored at a central location, since communication typically has to go through a small number of routers that serve as points of traffic aggregation in the network.

While a centralized architecture will capture most of the traffic in the network, and has the potential to detect attacks originating from the AMI network and going towards the utility servers (or vice versa), it will not be able to detect attacks within the AMI network, such as attacks against the routing protocol of the mesh network, MAC or PHY layer attacks, and end-to-end application layer attacks between peer to peer AMI nodes. A more *distributed* way of monitoring is necessary to detect those network level attacks. A sufficient number of IDS sensors would have to be deployed in various places in AMI networks – in a way that every network traffic and every network device integrity can be measured and monitored. In the next two sections, we explain two such distributed monitoring architectures and discuss their trade-offs.

C. Embedded Sensing Infrastructure Trade-Offs

In an embedded sensing infrastructure, every smart meter node is instrumented with intrusion detection capabilities.

There are a number of trade-offs to consider with such an architecture. On the one hand, with an embedded architecture it is not necessary to acquire permits in order to install the infrastructure (beyond whatever was needed to install the AMI network itself), and no specialized staff is required, beyond the meter installers (already needed to install the AMI network). Thus there are savings in terms of time, cost, and installation complexity. On the other hand, smart meters have limited processing, storage, and communication capabilities. Limited processing capabilities at each meter would cause gaps in security coverage, as meters that are busy with their regular functions would not be able to spare the CPU cycles to perform IDS operations.

While meter vendors can sell more powerful meters that can handle intrusion detection functions, utilities might be unwilling to pay the additional price. Most utilities need to purchase millions of smart meters, so a small increment in price for each meter (e.g., a few dollars) would result in an additional investment of millions of dollars. In addition, meters have to abide by power consumption constraints specified by ANSI Standard C12.1 and IEC 62053-61, which limit the maximum energy consumption of a meter to 5W. Increased computational requirements due to additional security processing lead to difficulties in maintaining the meter's energy consumption within the overall power budget [15]. Thermal constraints impose additional constraints on the hardware components that can be included in each meter as well as their level of utilization during processing.

D. Dedicated Sensing Infrastructure Trade-Offs

In a dedicated sensing infrastructure additional devices (in addition to smart meters and other communication-relaying hardware) are deployed throughout the network for the purpose of monitoring the infrastructure. These dedicated intrusion detection systems can be used to monitor not only security events, but also the health of the network (e.g., routing topology of an AMI network).

The most important benefit of a dedicated sensing infrastructure is the availability of processing and storage to perform complex monitoring and processing functions, as the dedicated devices will be more powerful than smart meters (although there will also be fewer dedicated devices, when compared to smart meters). On the other hand, there are a number of challenges to deploying this kind of infrastructure, related to the cost and complexity of installation and maintenance.

In particular, while the meters have reserved sockets where they are installed, the dedicated sensor equipment has to be sited elsewhere, e.g., on light poles or rooftops. That kind of installation would typically require a site survey, obtaining of permits, renting of installation sites, and hiring of highly specialized personnel, especially when the installations are in places that are difficult to reach. Similarly, maintenance of the dedicated sensors would require the dreaded *truck roll*, whereby specialized personnel and equipment are needed to gain physical access to the IDS sensor. For example, a deployment guide for residential wireless broadband networks puts the cost of a basic truck roll in the neighborhood of $300 per household [16]. More complex installations which require mounting equipment on light polls or roofs would incur costs that are several times higher.

c	Centralized Architecture
d	Dedicated Devices Architecture
e	Embedded Architecture
C_i	Capital expense for architecture i
n_i	Number of devices needed for architecture i
s_i	Marginal cost for devices in architecture i
R	Risk = likelihood * cost
p_i	Likelihood of attack i on AMI network
d_j^i	Probability of architecture j detecting attack i
A_j^i	Cost of attack i before it is detected by architecture j
A_m^i	Cost of attack i if it goes undetected

TABLE I
LIST OF SYMBOLS

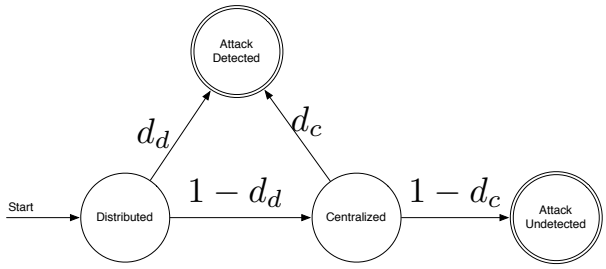

Fig. 1. The joint probability distribution between the probability of detection of a centralized architecture and a distributed architecture is modeled as a Markov chain. With probability one the process starts in the distributed state, and the only two absorbing states are the Detection state or the Attack Undetected state. A similar joint distribution can be used for the embedded architecture.

In addition, in multi-channel networks, different sets of nodes or even different pairs of nodes may communicate simultaneously on different channels, and channel selection may even change on a per-packet basis. That would require that the sensor be able to decode traffic on multiple channels simultaneously. Off-the-shelf hardware with that kind of functionality is not currently available for all PHY/MAC layers in use in smart grid networks, increasing the cost of the IDS sensor.

V. DECISION-MAKING FRAMEWORK

As explained above, owners and operators have several options for deploying an IDS in AMI mesh networks. Creating the business case for each of those options, however, is not straightforward. Now, we provide a risk-assessment formulation that would help AMI asset owners make more informed decisions when faced with the question: *which IDS deployment is best-suited to my network?*

Despite the difficulties associated with estimating probabilities and cost, risk-assessment models are arguably the most useful tools that currently exist for making investment decisions. The common alternative is to make decisions based on intuition, in which case assumptions are not explicitly discussed or analyzed. By leveraging the risk-assessment framework, we are able to analyze decisions based on clear assumptions: assumptions that can be discussed, critiqued, and improved by future researchers. Thus, this paper takes the first step in creating a more systematic way of evaluating IDS deployment options in AMI networks.

Furthermore, while describing the framework, we also demonstrate some basic principles of IDS deployment that are independent of specific probability and cost values; for instance, in rural and small AMI deployments, having an embedded IDS sensor is the most cost-effective solution, whereas in large, high-density deployments, dedicated IDS sensors are the best approach. These examples validate our risk-assessment formulation.

A. Risk Assessment and Cost Model

In this section we use the notation defined in Table I. We assume that utilities want to minimize their risk by studying which technology is best for their AMI deployment:

$$\arg_{i\in\{c,d,e\}} \min R_i, \tag{1}$$

where c indicates a centralized monitoring architecture, d is a dedicated architecture, e is an embedded architecture, and R_i represents the traditional notion of risk ($\sum likelihood \times cost$). We can decompose the risk into a static investment cost to deploy the monitoring architecture (with a probability of 1) and a dynamic event cost that changes over time with respect to the threat environment and the detection capabilities of the IDS sensors deployed.

The capital expense C_c for a utility that is considering deploying a centralized monitoring solution would include the fixed cost of deploying security monitoring appliances in the utility network. In the case of a dedicated architecture in the field, the investment C_d would depend on the number of dedicated sensors n_d and the cost of each sensor s_d (including compound operational cost, like the rental of new places to install these devices), so $C_d = n_d * s_d$. Finally, the investment C_e of instrumented meters with IDS sensors would depend on the number of meters n_m and the per-unit extra cost s_m, so $C_e = n_m * s_m$.

The risk related to dynamic events depends on the likelihood of attacks and their cost if successfully detected or not. Let p_i be the likelihood of an attack i on the AMI network. The cost of an attack depends on how soon the attack is detected: the longer it takes to detect an attack, the more expensive it is to undo or fix the damage. Given that centralized, dedicated, and embedded IDS deployments will likely detect an attack at different points in its lifecycle, we differentiate A_c^i, A_d^i, and A_e^i as the costs of attack i incurred by the utility before it is detected by the centralized, dedicated, and embedded IDSes, respectively. If the attack is left undetected, its cost would reach a value A_m^i larger than any of the costs of detected attacks (i.e., $A_m^i > A_c^i,\ A_d^i,\ and\ A_e^i$). We then name d_c^i, d_d^i, and d_e^i as the probabilities of detection of this attack i by a centralized IDS deployed at the head-end, by a set of dedicated sensors, and by an embedded IDS deployment, respectively. Distributed sensors have more information on attacks, and might be able to detect attacks earlier and with higher probability than centralized detectors. Therefore we assume $A_c^i \geq A_d^i, A_e^i$ and $d_c^i < d_d^i, d_e^i$.

Based on those properties, we present three equations that can be used by utilities to calculate the risks associated with different IDS deployment architectures. First, the risk of

investing in a centralized solution can be calculated by:

$$R_c = C_c + \sum_i p_i(d_c^i A_c^i + (1 - d_c^i)A_m^i), \tag{2}$$

which says that if a utility relies on a centralized IDS that costs C_c to deploy, the risk to the utility is the likelihood of an intrusion times the cost of such an intrusion summed over all intrusions considered by the threat model. The cost of a specific intrusion has two components, corresponding to detection of the intrusion and failure to detect the intrusion.

Even if a utility invests in a distributed IDS infrastructure, it would still require a central point for receiving and managing alerts. We note that the distributed infrastructure refers to intrusion detection sensors deployed within the AMI. The alerts reported by those sensors, however, would be managed in a centralized fashion; therefore, in our framework, we assume that investments in distributed sensors also require a basic investment in a centralized solution. Being mindful of those properties, we present the second equation, which can be used to calculate the costs expected when a utility invests in a *dedicated* distributed IDS (in addition to a centralized IDS):

$$R_d = C_c + n_d s_d + \sum_i p_i(d_d^i A_d^i + (1 - d_d^i)d_c^i A_c^i + (1 - d_d^i)(1 - d_c^i)A_m^i), \tag{3}$$

where the risk of an intrusion is reduced by the probability of detection d_d^i by the dedicated infrastructure, assuming that the cost A_d^i is much lower than A_c^i, which is in turn lower than A_m^i. We assume that the joint probability distribution among the alerts generated by the centralized detector d_c and the distributed detector d_d are such that those detections precede the detection by a centralized IDS, therefore, for the A_d (and A_e below) case, we assume $d_d^i(1-d_c^i) = d_d^i$ (and $d_e^i(1-d_c^i) = d_e^i$). This is exemplified in figure

Similarly, if a utility invests in an *embedded* distributed IDS (in addition to the centralized IDS), the risk can be calculated using:

$$R_e = C_c + n_m s_m + \sum_i p_i(d_e^i A_e^i + (1 - d_e^i)d_c^i A_c^i + (1 - d_e^i)(1 - d_c^i)A_m^i), \tag{4}$$

VI. Framework Application

A. Coverage Area Analytical Study

Comparison of the trade-offs between the two distributed architectures (embedded vs. dedicated) will depend on the network topology and the type of deployment (urban, suburban, or rural AMI), as described in the NIST Guidelines for Assessing Wireless Standards for Smart Grid Applications [17]. We now detail how to incorporate that information into the risk equations to enable calculation of risks for a variety of network deployments.

Among the various inputs required by the risk assessment equations, C_c, s_d, and s_m are easy to estimate, since utilities can obtain the prices of security appliances and sensors from

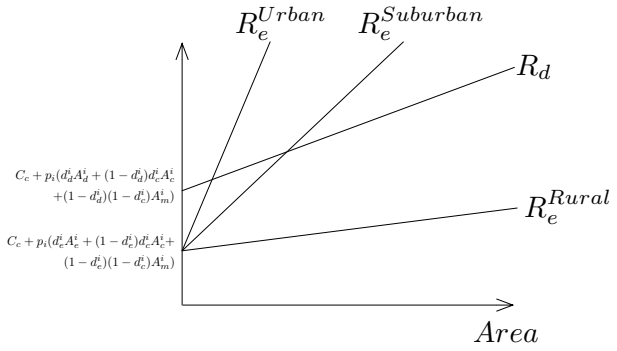

Fig. 2. The network density determines the cost effectiveness of dedicated sensors. In rural AMI deployments and in small areas, having an embedded sensor minimizes the risk, whereas in large, high-density deployments, it becomes cost-effective to deploy dedicated sensors instead.

vendors. However, the number of dedicated sensors n_d requires study of the notion of coverage. Coverage of a wireless network by dedicated sensors requires that the sensors be deployed such that they can overhear any packet sent by a smart meter within the AMI network area. A typical density used as a guideline for the deployment of Wi-Fi access points is in the range of 50-60 nodes per square mile, given a range of 300 feet per node [18]. We adopt this guideline for our dedicated sensor evaluation and express the number of dedicated nodes in terms of the area that needs to be covered, as follows:

$$n_d = Area * Sensor\_Density = Area * 50 \tag{5}$$

The number of meters n_m is well-known by utilities, since it matches their customer base. We can express the number of smart meters in terms of area through the use of several basic meter densities derived from U.S. Census data on household densities [19], assuming one meter per household. In particular, the densities for urban, suburban, and rural environments are 972.5, 128.7, and 7.5, respectively. Therefore, $n_e = Area * 972.5$ for urban environments, $n_e = Area * 128.7$ for suburban environments, and $n_e = Area * 7.5$ for rural environments.

Substituting those values (n_d and n_e) in the corresponding risk equations (R_d and R_e), we can identify the trends for coverage area depending on the type of deployment, as shown in Fig. 2. By analyzing how R_d and R_e change as a function of the area covered by the AMI deployment, we find that the slope of the risk equations for embedded deployments will depend on the density of meters. Therefore, the slope for R_e^{Urban} (972.5) is larger than the slope for $R_e^{Suburban}$ (128.7), which in turn is larger than the slope for R_e^{Rural} (7.5). In addition, we can compare each of the AMI deployments with the deployment requirements of the dedicated infrastructure. Because the deployment of dedicated IDS sensors does not differ significantly among different types of AMI network, the slope of R_d is fixed (50). From those results, we can conclude that as the area covered by the AMI network increases, the cost for deploying embedded IDS sensors will be higher than that of a dedicated IDS-sensing infrastructure for urban and suburban environments; however, for rural deployments, the cost of a dedicated IDS infrastructure would be too expensive

relative to that of embedding IDS sensors in the smart meters themselves.

Those results are confirmed by our experience in deploying AMI networks. Some rural AMI networks consist of a large chain of smart meters forming a line. (Each smart meter only has two neighbors in its routing table; one neighbor relays packets towards the collection unit, and the other relays packets away from the collection unit.) Deployment of dedicated IDS devices to monitor this network would require at least one dedicated IDS sensor for every two smart meters, if not one for every smart meter. In contrast, in urban deployments, several hundred (or even a thousand) smart meters fall within the same wireless communication range; therefore, a single dedicated IDS device can potentially monitor hundreds of smart meters, making a dedicated infrastructure a cost-effective alternative to an embedded solution in large, high-density deployments.

The framework allows utilities to make use of information about the network topologies and deployment types that are relevant to their AMIs. Such information, as shown above, would improve the accuracy of the cost and risk values being computed.

B. Likelihood of Attack Analytical Study

A similar study can be done to consider the impact of the likelihood of attack p_i on the cost-effectiveness of each IDS deployment option. If a utility is certain that it will not be attacked at all ($\sum p_i = 0$), then $R_c = C_c < R_d = C_c + n_d s_d < R_e = C_c + n_m s_m$. (The last equation holds if we assume a large, dense area as depicted in Fig. 2 with $n_m > n_d$ such that $n_m s_m > n_d s_d$.) As the likelihood of attack $\sum p_i$ increases (any of the possible attacks the model considers becomes more likely), then we need to consider the right-hand side of equations (2), (3), and (4).

We look at the cost of an attack i in equations (2), (3), (4). First, we show how the cost of an attack i in the centralized model is greater than in the dedicated model:

$$\begin{aligned} &d_c A_c + (1-d_c)A_m > \\ &d_d A_d + (1-d_d)d_c A_c + (1-d_d)(1-d_c)A_m \end{aligned} \quad (6)$$

After some algebraic manipulations, we observe that the previous inequality holds if and only if

$$d_c A_c d_d + (1-d_c)d_d A_m > d_d A_d. \quad (7)$$

Assuming $d_d > 0$, we only need to show that $d_c A_c + A_m(1-d_c) > A_d$, but this relation holds because

$$d_c A_c + A_m(1-d_c) > d_c A_c + A_c(1-d_c) = A_c > A_d, \quad (8)$$

since under our model assumptions $A_m > A_c > A_d$.

A similar analysis can be done for the comparison between an embedded sensor and a dedicated sensor. In general, we assume that the probability of detecting an attack with an embedded sensor is higher than with a dedicated sensor, and that the damage done when an attack is detected through a dedicated sensor is similar to, or higher than, the damage done when it is detected through an embedded sensor (i.e., $A_d \geq A_e$). That implies:

$$\begin{aligned} &d_d A_d + (1-d_d)d_c A_c + (1-d_d)(1-d_c)A_m \geq \\ &d_e A_e + (1-d_e)d_c A_c + (1-d_e)(1-d_c)A_m \end{aligned} \quad (9)$$

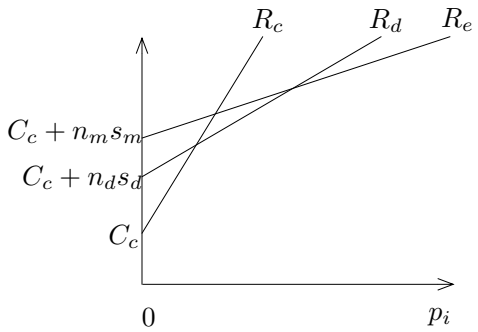

Fig. 3. In a large, and dense (e.g. urban) area, as the likelihood of an attack increases, the most cost-effective IDS deployment option changes from a centralized deployment, to a dedicated sensor deployment, and finally when attacks are very likely, to an embedded intrusion detection deployment.

The final result can be seen in Fig. 3.

The likelihood of attack will increase with the longevity of the AMI deployment (i.e., if all other variables are equal, a device deployed for only one year is less likely to be attacked during its lifetime than the same device deployed for 30 years). Based on those observations, we argue that distributed IDS architectures will be more cost-effective in a long-lived deployment, and that the risks associated with each type of deployment scheme will converge to the trends shown in Fig. 3 when a more extensive set of attacks are taken into account, along with the lifespan of each deployment. If a utility finds it hard to estimate the likelihood of attacks, our framework recommends using the AMI deployment lifespan to choose a suitable IDS architecture.

C. Leveraging Historical Data

One of the main challenges, in applying a risk assessment framework is in the estimation of the parameters required to compute quantitative results. In this subsection and the next, we present two different approaches to illustrate how utilities can leverage our framework, whether they have access to estimated parameters or not. A first set of two examples describes the case when parameter values can be extracted from empirical data; a second example shows how to conduct a sensitivity analysis to enable informed decisions even when some important parameter values are not available.

1) Denial-of-service against data collection unit: In this scenario, we assume that a utility is interested in investing in an IDS to monitor its AMI after experiencing a cyber incident that disrupted the communication network between meters and a data collection unit (DCU). In particular, this utility was the victim of a DDoS attack against one of the core DCU that affected some demand response applications and distribution automation devices. Such attacks can lead to demand supply imbalances in the grid, which could cause brownouts or blackouts. In addition to investing in new protective measures to prevent recurrence of such incidents, security administrators are interested in deploying a comprehensive IDS that could

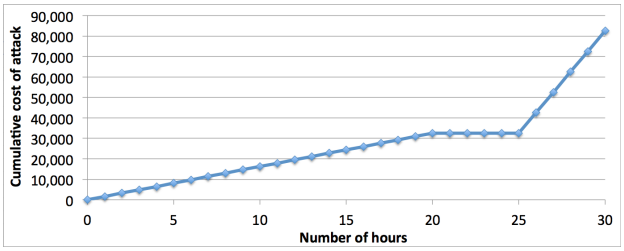

Fig. 4. Estimated cost of a DDoS attack against a DCU as it progresses without being detected.

have alerted them when this incident occurred. The question is, which IDS architecture would be the best investment from both cost and security perspectives. We first review the attack process in detail. We assume that the utility conducted a detailed forensics investigation that led to the following understanding. Adversaries proceeded in three sequential steps: 1) they compromised 50 meters over the course of 20 hours by exploiting a vulnerability in the optical communication port; 2) they coordinated communications among those compromised meters in order to create a botnet within the AMI; and 3) they synchronized the launch of a massive number of attack payloads against the DCU. Those attack payloads were C12.22 requests sent at a high frequency and crafted to maximize the computational power required by the DCU to process them.

The first step in applying the risk assessment framework consists in assigning attack cost over time for each attack step. For the first step, we assume that the damage cost increased linearly with the number of meters compromised. In particular, 50 meters were compromised over the course of 20 hours. Based on vendor information, we estimate the cost of a meter device to be $100, the cost of replacing it $500, and finally the cost of losing measurements from compromised meters $50. As a result, each compromised meter cost a total of $650 for the utility; after 20 hours, the cost added up to $32,500 for the 50 meters. The second step had no impact on the infrastructure, so the cost of the attack stayed at $32,500 between hour 21 and hour 25. At hour 26, the distributed denial of service attack commenced and prevented core AMI applications (e.g., demand response and distribution automation) from working. Based on financial losses due to failure to execute demand response curtailments, the utility estimated that the cost of this attack phase was $10,000 per hour. As a result, the total cost of such an attack, if it is not detected, adds up to $82,500 (5 hours of denial of service in addition to the attack cost at hour 25). The evolution of this cost over time is shown in Fig. 4.

The next step is to estimate the detection probability. Table II shows the ability of the different monitoring architectures to detect each attack step. For example, the first step that involves compromise of meters is a host-based attack that would be out of the reach of network-based monitoring solutions, so only the embedded IDS could detect it. However, the second step, which involves coordination of compromised meters, requires some command-and-control communications in the mesh, so it could be detected by both the embedded and distributed IDSes, but would likely be invisible to a central monitoring solution. Finally, in flooding the network and preventing the DCU from properly responding, the third step would be visible in all three of the IDS architectures.

TABLE II
COVERAGE OF THE DIFFERENT IDS ARCHITECTURES TO DETECT THE VARIOUS STEPS REQUIRED FOR A DDOS ATTACK

Attack steps for DDoS	*Centralized*	*Distributed*	*Embedded*
1. Meter compromise			✓
2. Meter coordination		✓	✓
3. Attack payload	✓	✓	✓

Based on the detection capabilities of each architecture and knowledge about the true and false positive rates of each IDS architecture, we estimate the probabilities of detection to be $d_e = 90\%$ for the embedded IDS, $d_d = 80\%$ for the dedicated IDS, and $d_c = 70\%$ for the centralized IDS. (Different value ranges are explored for those three parameters in the next subsection.) We then estimate that it would take 8 hours for the embedded IDS to detect the first step of the attack, which leads to an attack cost before detection of $A_e = \$13,000$. (20 meters would be affected during the first 8 hours, and the cost of each is $650.) The cost for the second step of the attack that can be detected by the dedicated IDS is fixed: $A_e = \$32,500$. We also estimate that it would take 2 hours for the centralized IDS to detect the third step of the attack, so $A_c = \$52,500$ (the cost of the second step of the attack along with $10,000 per hour for the third step of the attack). Finally, the cost if the attack is never detected adds up to $A_m = \$82,500$ (cost after step 3 along with 2 additional hours of denial of service at $10,000 each).

The final phase of our approach is to factor in the cost associated with each infrastructure. As the AMI under consideration is a suburban environment covering 500 square miles, we can estimate the total number of meters to be $n_e = 500 * 128.7 = 64,350$. The total number of dedicated sensors would be $n_d = 500 * 2 = 1,000$. Finally, if we assume that the investment and management cost of the centralized monitoring system is $C_c = 200,000$, the cost of a dedicated sensor is $50, and the cost of an embedded sensor is $1. Figure 5 show the risks associated with the three deployment architectures as the probability of the DDoS attack varies from 0 to 1. The resulting risk indicates that the centralized architecture is the most cost-efficient infrastructure to deploy for this attack. However, one must keep in mind that the attack would last longer in the case of a centralized IDS than for a distributed IDS. It is interesting to see from Figure 5 that if the attack has a high probability of occurrence, the dedicated IDS architecture becomes more cost-efficient than the embedded IDS architecture.

2) Abuse of a Distribution Automation System: The goal of distribution automation (DA) application is real-time monitoring and maintenance of electricity grid health (e.g., by managing voltage levels at various points in the grid). Distribution automation elements can leverage an AMI for their communication needs, allowing a utility to have a single communication infrastructure that enables many applications [20]. Having a multi-purpose communication infrastructure significantly reduces the cost of enabling communication across all smart grid applications, but also makes the AMI network more attractive

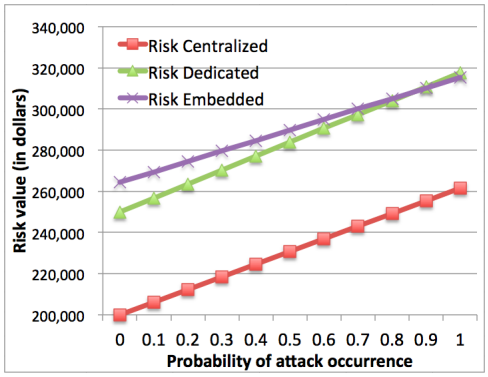

Fig. 5. Risks for the three different architectures when considering a DDoS attack on the DCU.

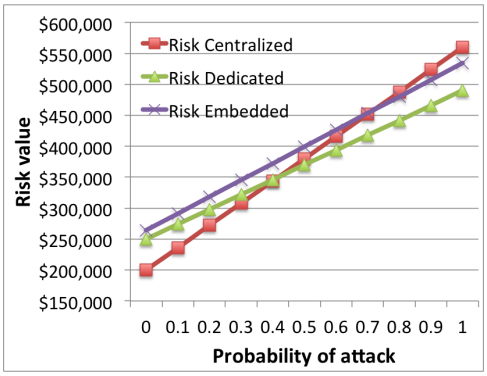

Fig. 6. Risks for the three different architectures when considering an attack against the distribution automation system.

to adversaries interested in disrupting the energy delivery system. In particular, failure to deliver the right commands at the right time to one or more distribution system elements (e.g., voltage regulators, transformers, switches, feeders, reclosers, or capacitors) could affect voltage levels in parts of the grid, causing damage to the elements themselves and also to end user devices, and even leading to outages [21]. These goals could be achieved by tampering with DA packets, e.g., by dropping, delaying, or replaying them. As an example, the cost of damaging a large 33kv transformer is on the order of a hundred thousand dollars [22] and smaller transformers cost several hundred to thousands of dollars. This does not include installation costs and typically distribution networks have a lot of transformers.

Another important function of DA devices is the detection and isolation of component failures in order to prevent or limit the extent of outages. The isolation is accomplished by configuration of switches/feeders around the failure such that electricity would be re-routed around it. Tampering with that functionality can lead to false alarms and unnecessary outages, or to missed failure notifications accompanied by outages, both of which would disrupt service and be expensive to resolve.

We assume an attack scenario with four steps: 1) collection of cryptographic keys via compromise of a set of meters, 2) man-in-the-middle attacks in a few neighborhood area networks (NANs) to re-route AMI traffic through a computer controlled by attackers, 3) a reconnaissance phase during which DA traffic flowing in the NANs is collected and analyzed, and 4) injection of malicious DA traffic and tampering with control commands to cause distribution device failures and, potentially, outages. Following the cost modeling approach presented earlier, we assign cost and detection probability distributions for each attack step and for each monitoring architecture. The cost of attack until detection is set to $100,000 for step 1, $50,000 for steps 2 and 3, and $200,000 for step 4. Probabilities of detection are set to 90% for the embedded sensors, and 80% for both the dedicated and centralized sensors. Finally the total cost of the attack if left undetected is set to $1,000,000. The resulting risk is shown on Figure 6.

This time, we observe that the centralized architecture is the best choice until the probability of attack goes beyond 40%. After this threshold, the dedicated architecture becomes the most cost-efficient solution.

D. Conducting a Sensitivity Analysis

In the previous subsection, historical and evaluation data could be accessed in order to estimate the different parameters required by the risk assessment framework. When such data is not available, utilities can still follow a rigorous approach to gain insights into the type of monitoring architecture that would be best for them. The goal is to separate parameters into two groups, known and unknown, and then to conduct a sensitivity analysis on the unknown parameters in order to understand their impact. In the case of a utility, known parameters are the cost of deploying the different IDS architectures, the number and costs of AMI devices (meters, relays, DCUs), and the cost of interventions (e.g., rolling a truck or replacing a compromised device). Unknown parameters are the probability of occurrence of an attack and the probability of successful detection for a given time period.

To illustrate the benefits of conducting a sensitivity analysis, we revisit the denial-of-attack scenario against a DCU that was presented in the previous subsection, but this time we explore ranges of values for the following three unknown parameters: the probabilities of detection of the DDoS attack by the centralized (D_c), dedicated (D_d), and embedded (D_e) architectures. We vary the values for those 3 parameters from 0 to 1 by increment of 0.1 and analyze the results. Figure 7 shows a 3-dimensional plot that reveals the most cost-efficient architecture (yellow diamond for centralized, green circles for dedicated, and black squares for embedded) for each possible combination of D_c, D_d, and D_e. Given those results, security administrators can make an informed decision according to the region in this 3-dimensional space in which their threat model and environment fit the best. The analysis clearly indicates that a centralized architecture is the most cost-efficient solution, because the embedded and dedicated architectures are only worth the investment when they offer perfect detection accuracies (D_e or D_d equal to 100%). This is due to the small number of meters affected by the attack under consideration, which leads to the most expensive phase of the attack being the denial of service against the DCU (most likely detectable by the centralized IDS) rather than

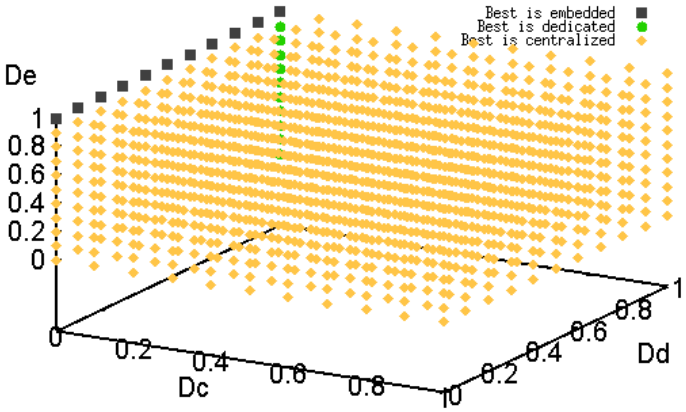

Fig. 7. Selection of the most cost-efficient architecture when exploring different probabilities of detection for the three architectures. The values for D_c, D_d, and D_e vary between 0 and 1 by increment of 0.1. Points in the 3D space represent which architecture is the most cost-efficient for each of the 1,331 combinations explored.

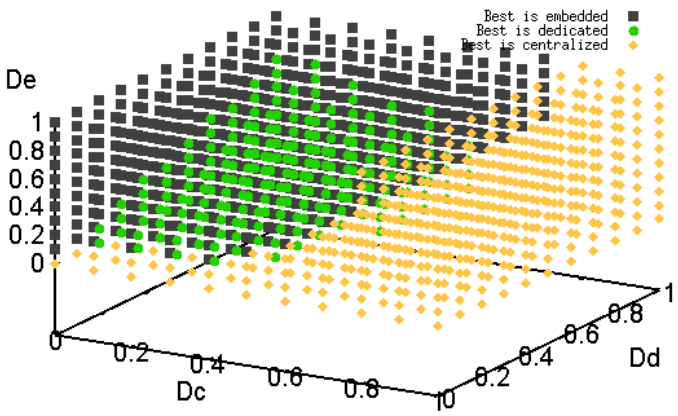

Fig. 8. Selection of the most cost-efficient architecture when exploring different probabilities of detection for the three architecture. The only difference with the previous figure consists in the number of meters affected by the attack (500 instead of 50).

meter compromises (most likely detectable by the embedded or dedicated IDS).

To see when the other architectural option might become more viable we ran the sensitivity analysis a second time by changing the number of affected meters from 50 to 500, which leads to an increase in the costs of the attack if not detected by one order of magnitude. Figure 8 shows that this time, for most combinations of D_c, D_d, and D_e, the optimal architecture is no longer the centralized IDS. Indeed, for 45.3% of the combinations, the embedded architecture is more cost-efficient, compared to 38.4% for the centralized architecture, and 16.3% for the dedicated architecture. Ideally one would like to design the IDS system to deal with the worst attack scenario possible within cost constraints, and ultimately a utility has to choose an architectural option based on their cost constraints and worst perceived threat. The examples presented were meant to illustrate the application of the framework and not to recommend one architecture over the other as choice of the architecture clearly depends on many factors. However, as was evident in the previous section, the framework does indicate that certain IDS architectures are inherently more suitable for certain AMI deployments.

VII. Discussion and Conclusion

In this paper we presented a comprehensive study of a problem faced by many utilities interested in improving their security posture in a cost-effective way: How to choose among multiple deployment options for intrusion detection systems in AMI networks?

We started by describing the possible ways we could deploy monitoring architectures and we then created a risk-assessment framework that considered the attack threats (and consequences of successful attacks) and the possible deployment options, and used it to select the best option for a given network deployment.

The biggest challenge in applying our risk-assessment formulation in practical scenarios is the difficulty of estimating some of the probabilities and costs used as model inputs. However, we were able to show that there are some basic principles inherent to each deployment option that hold true irrespective of the specific values used. In addition, while our model might not be applicable in all possible use-cases or when the risk uncertainty is large enough to be unquantifiable, with our model we are able to analyze decisions based on clear assumptions that can be discussed, critiqued, and improved by future researchers. Our hope in this paper is to provide the first steps towards creating a better, systematic way of evaluating IDS deployment options in AMI networks and improve over the current heuristics.

A thorough and realistic risk assessment needs to account for all possible attacks within the threat model under consideration and also the deployment's life cycle, as attacks are likely to occur multiple times over the lifetime of the infrastructure. In our future work, we plan to conduct this kind of study using advanced stochastic modeling frameworks such as ADVISE [23], [24].

References

[1] "DoE energy delivery systems cybersecurity roadmap." http://energy.gov/sites/prod/files/EnergyDeliverySystemsCybersecurityRoadmap_finalweb.pdf, 2011.

[2] "Cybersecurity risk management process," http://energy.gov/oe/services/cybersecurity/cybersecurity-risk-management-process-rmp, 2012.

[3] H. Wei, D. Frinke, O. Carter, and C. Ritter, "Cost-Benefit Analysis for Network Intrusion Detection Systems," in *28th Annual Computer Security Conference, CSI Proc.*, 2001.

[4] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *4th International Conference on Critical Information Infrastructures Security (CRITIS), Proc.*, 2009, pp. 176–187.

[5] R. Berthier, W. Sanders, and H. Khurana, "Intrusion detection for advanced metering infrastructures: Requirements and architectural directions," in *1st International Conference on Smart Grid Communications (SmartGridComm), IEEE Proc.*, 2010, pp. 350 –355.

[6] P. Jokar, H. Nicanfar, and V. Leung, "Specification-based intrusion detection for home area networks in smart grids," in *International Conference on Smart Grid Communications (SmartGridComm), IEEE Proc.*, Oct. 2011, pp. 208 –213.

[7] S. McLaughlin, D. Podkuiko, S. Miadzvezhanka, A. Delozier, and P. McDaniel, "Multi-vendor penetration testing in the advanced metering infrastructure," in *26th Annual Computer Security Applications Conference (ACSAC), Proc.*, 2010, pp. 107–116.

[8] S. McLaughlin, D. Podkuiko, and P. McDaniel, "Energy theft in the advanced metering infrastructure," in *4th International Conference on Critical Information Infrastructures Security (CRITIS), Proc.*, 2010, pp. 176–187.

[9] D. Grochocki, J. H. Huh, R. Berthier, R. B. Bobba, W. H. Sanders, A. A. Cardenas, and J. G. Jetcheva, "AMI Threats, Intrusion Detection Requirements and Deployment Recommendations," in *3rd IEEE International Conference on Smart Grid Communications (SmartGridComm), IEEE Proc.*, Nov. 2012, pp. 395–400.

[10] X. Wang and P. Yi, "Security framework for wireless communications in smart distribution grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 809 –818, Dec. 2011.

[11] M. A. Faisal, Z. Aung, J. R. Williams, and A. Sanchez, "Securing advanced metering infrastructure using intrusion detection system with data stream mining," in *Intelligence and Security Informatics*. Springer, 2012, pp. 96–111.

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

[13] N. Beigi Mohammadi, J. Mišić, V. B. Mišić, and H. Khazaei, "A framework for intrusion detection system in advanced metering infrastructure," *Security and Communication Networks*, 2012.

[14] A. Fawaz, R. Berthier, and W. H. Sanders, "Cost modeling of response actions for automated response and recovery in AMI," in *3rd International Conference on Smart Grid Communications (SmartGridComm), IEEE Proc.* IEEE, 2012, pp. 348–353.

[15] E. Beroset, "Meter security in the smart grid context," *Western Energy Magazine*, pp. 14–16, Summer 2012.

[16] R. Conaway, "Tales from the towers, chapter 14: How to make money as a wireless ISP," http://www.muniwireless.com/2010/09/21/how-to-make-money-as-a-wireless-isp/, September 2010.

[17] "NISTIR 7761. Guidelines for Assessing Wireless Standards for Smart Grid Applications." http://collaborate.nist.gov/twiki-sggrid/pub/SmartGrid/PAP02Wireless/NISTIR7761.pdf, Feb. 2011.

[18] "2010 Guidelines for Successful Large Scale Outdoor Wi-Fi Networks," http://www.scribd.com/doc/25026923/2010-Guidelines-for-Large-Scale-Outdoor-WiFi, Dec. 2009.

[19] "U.S. Census 2000 Summary File 1 - Population, Housing Units, Area, and Density: 2000," http://www.census.gov/population/www/censusdata/density.html.

[20] "Sensus distribution automation," http://www.sensus.com/web/usca/solution/distribution-automation.

[21] I. Lim, S. Hong, M. Choi, S. Lee, T. Kim, S. Lee, and B. Ha, "Security protocols against cyber attacks in the distribution automation system," *IEEE Transactions on Power Delivery*, vol. 25, no. 1, pp. 448 –455, Jan. 2010.

[22] W. H. Bartley, "Analysis of Transformer Failures." http://www.bplglobal.net/eng/knowledge-center/download.aspx?id=191, 2003.

[23] E. LeMay, M. Ford, K. Keefe, W. Sanders, and C. Muehrcke, "Model-based Security Metrics Using ADversary VIew Security Evaluation (ADVISE)," in *8th International Conference on Quantitative Evaluation of Systems (QEST), Proc.* IEEE, 2011, pp. 191–200.

[24] E. LeMay, W. Unkenholz, D. Parks, C. Muehrcke, K. Keefe, and W. H. Sanders, "Adversary-driven state-based system security evaluation," in *6th International Workshop on Security Measurements and Metrics, Proc.*, 2010, pp. 5:1–5:9.