A Comparison of Different Intrusion Detection Approaches in an Advanced Metering Infrastructure Network using ADVISE

Michael Rausch[1], Brett Feddersen[1], Ken Keefe[1], and William H. Sanders[2]

[1] Information Trust Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
{mjrausc2, bfeddrsn, kjkeefe}@illinois.edu
[2] Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA
whs@illinois.edu

Abstract. Utilities responsible for Advanced Metering Infrastructure (AMI) networks must be able to defend themselves from a variety of potential attacks so they may achieve the goals of delivering power to consumers and maintaining the integrity of their equipment and data. Intrusion detection systems (IDSes) can play an important part in the defense of such networks. Utilities should carefully consider the strengths and weaknesses of different IDS deployment strategies to choose the most cost-effective solution. Models of adversary behavior in the presence of different IDS deployments can help with making this decision as we demonstrate through a case study that uses a model created in the ADversary VIew Security Evaluation (ADVISE) formalism (which calculates metrics used to compare different IDSes). We show how these metrics give valuable insight into the selection of the appropriate IDS architecture for an AMI network.

Keywords: Advanced Metering Infrastructure (AMI) · Smart Grid · ADversary VIew Security Evaluation (ADVISE) · Security Modeling · Intrusion Detection Systems (IDS)

1 Introduction

Many utility companies are creating Advanced Metering Infrastructure (AMI) networks, which incorporate smart meters and other intelligent components into the power grid. The added functionality allows utilities to monitor and control their smart grid with more precision than was previously possible. As an example, a utility company can use an AMI infrastructure to remotely collect more frequent meter readings, which allows them to respond more accurately to fluctuations in power demand.

Unfortunately, AMI networks increase the attack surface of a power grid. For example, an unscrupulous customer may compromise a single smart meter so that it sends false data to under-report electricity consumption, resulting in

a lower bill. Distributed denial of service attacks, traffic injection attacks, and Byzantine attacks are examples of new threats to these cyber-enhanced power grids. As utility companies build and maintain AMI infrastructures they should be aware of the possibility of these attacks, and work to create a cost-efficient architecture that minimizes the expected damage.

One obvious way of limiting the potential damage of an attack is to detect and respond to the attack before it can cause much harm. An intrusion detection system (IDS) can help a utility company detect an attack. There are several different IDS architectures that can be deployed by a utility company as a defensive precaution. Each architecture has a different cost and degree of effectiveness. A utility company must decide whether its application warrants an IDS, and if so, which would give the best protection for the best price.

One approach for informing this critical design decision is to build a sound, state-based stochastic model of the system and the possible IDS architectures that can be applied to it. Quantitative metrics can be calculated on the models to determine which configuration provides the best cost/security balance.

Our approach is to study a multi-layered power grid example and the potential IDS implementations that can be applied to this grid. We used the ADversary VIew Security Evaluation (ADVISE) [6] modeling formalism in the Möbius modeling tool [10] for this work. We consider several different adversaries interested in attacking such a system and calculate useful and relevant security metrics. Using our approach, a utility company can make a more informed decision about how to implement an IDS on its grid.

To make an informed decision regarding the selection and implementation of various IDS approaches, it is necessary to know the probability that an adversary would successfully attack a system, given its type of IDS architecture. Given a particular adversary and IDS approach, it would be useful to know the estimated probability of detecting the adversary, the estimated damage to the utility due to activity of the adversary, and the type of attack chosen by the adversary. We create a model of the adversary behavior that is detailed enough to give insight into these metrics. We do not claim that the quantitative metrics generated by the model are accurate in any absolute sense. However, we do believe that they may be very useful when comparing the relative strengths and weaknesses of modeled systems. A model that gives quantitative security metrics will give a system designer another approach to supplement the advice and intuition of security experts.

The remainder of this paper is organized as follows. Section 2 provides a concise overview of AMI networks, IDS systems on AMI networks, and the ADVISE formalism. Section 3 offers a description of system we modeled. Section 4 offers a detailed explanation of the ADVISE model that was constructed, including the adversary profiles that were considered and the metrics that were defined on the system. Section 5 shows our quantitative results and our interpretation of them. Section 6 discusses previous work that seeks to examine power grid security using a variety of methods. Finally, Section 7 concludes the paper.

2 Background

2.1 AMI Overview

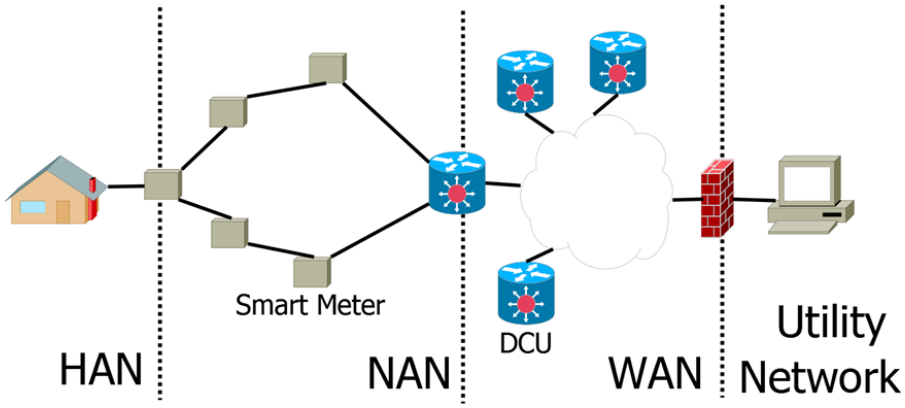

Fig. 1. Example system diagram.

An AMI gives a utility company the ability to remotely communicate with the electric meters in its grid. There are many possible network hierarchies. For example, some smart meters connect to the utility company through the consumer's Internet connection. However, our example system uses a hierarchy of communication gateways that rely on the utility's own network infrastructure, as depicted in Figure 1.

At the bottom level of our hierarchy, a smart meter forms the core of a *home area network* (HAN). The HAN may include other components, such as smart appliances designed to draw less power during times of high demand and more power during times of low demand. If there are multiple devices in the HAN, the smart meter itself may act as a network gateway for the other devices. Multiple HANs, and data collection unit (DCU) gateways serving them, are connected together to form a *neighborhood area network* (NAN). The DCU gateways in NANs are connected to the utility via a *wide area network* (WAN).

Many different communication technologies can be used in an AMI. The WAN usually utilizes higher bandwidth, long-range communication technologies such as long-range wireless, satellite, or power line communication. The NANs don't have the same bandwidth or range requirements, and can use shorter-range wireless. We assume in our case study that the NAN uses a wireless mesh network.

2.2 IDS Overview

Intrusion detection systems are intended to monitor a system for suspicious activity, to raise an alert if a security event occurs, and to log information to

determine how an attack proceeded through the system. A number of different IDS deployment strategies are possible in an AMI. In this paper, we consider centralized IDSes, dedicated distributed IDSes, and embedded distributed IDSes.

A centralized IDS deployment scheme would place an IDS at the top of the network hierarchy, in the utility company's network. The IDS would monitor all traffic flowing into and out of the utility company's command and control center LAN network, and raise an alert if it detected anything suspicious. However, it would be completely unaware of inter-meter traffic, since that would not pass through the top level of the hierarchy. Alternatively, a utility could deploy a distributed set of IDSes to monitor inter-meter communication. This approach would still require a central node to coordinate the monitoring, so it would have all of the benefits of a centralized IDS approach, with some additional installation and maintenance costs associated with the additional IDSes.

We studied two main varieties of distributed IDSes: dedicated and embedded. A dedicated device IDS deployment would have the same components as a centralized deployment, and in addition it would have a number of geographically distributed dedicated IDS devices in wireless communication with the smart meters. These IDSes would monitor all AMI traffic within wireless range. An embedded IDS deployment incorporates intrusion detection directly into the smart meter. Like the dedicated IDS architecture, the embedded IDSes communicate and cooperate with the central IDS device.

There are a number of trade-offs to consider when evaluating these IDS designs. A centralized IDS would potentially miss large families of attacks because it is unaware of inter-meter communication. It is, however, the cheapest IDS option. An embedded or dedicated IDS scheme would be able to observe inter-meter communication, possibly allowing it to detect a larger set of attacks than a centralized scheme, but would cost more. A dedicated architecture would cost more because many additional devices would have to be purchased and maintained, and separate permits and location sites would have to be acquired to install these devices. However, the device would be able to monitor inter-meter communication in the NAN. One dedicated device could serve multiple smart meters. An embedded system would not require separate building sites or permits, but every single meter would cost slightly more because of the added IDS capability. Given the large number of meters involved, even a small increase in price for an individual meter would potentially be very costly for a utility company. In addition to monitoring inter-meter communication, an embedded IDS architecture would be able to detect attacks on the meter itself. This means the embedded IDS option provides the greatest coverage against possible attacks.

Clearly, a utility company should seek the most cost-effective solution. The choice can be made and justified with metrics derived from the analysis of mathematical models, such as the one we developed using the ADVISE modeling formalism.

2.3 ADVISE Overview

The ADversary VIew Security Evaluation (ADVISE) method [6] is used to calculate quantitative security metrics via executable models of adversary behavior in a system [5]. At a high level, a modeler creates an Attack Execution Graph (AEG), which is similar to a standard attack tree, but incorporates additional details about each attack's properties, such as its cost, time to completion, and probability of success. The AEG also contains nodes that track the state of the model, such as the prerequisites and goals held by the adversary at a particular discrete point in time. Different adversaries may exhibit very different behaviors when attacking the same system, since their initial foothold in the system, knowledge, skills, and goals of interest may differ dramatically. An adversary's preference for avoiding cost, avoiding detection, and earning reward also plays a pivotal role in the approach taken when attacking a system. A modeler is given the ability to create different adversary profiles before executing the model in ADVISE to reflect this reality.

An AEG is defined by the tuple

$$< A, R, K, S, G, C >$$

where A is the set of attack steps, R is the set of access domains available to the adversary, K is the set of information that can be known by the adversary, S is the set of skills possessed by the adversary, and G is the set of goals that the adversary attempts to achieve. The relation C defines the set of directed connecting arcs from $e \in R \cup K \cup S \cup G$ to $a \in A$, where e is a prerequisite element needed in order to attempt a. This relation also defines the set of directed connecting arcs from $a \in A$ to $e \in R \cup K \cup S \cup G$, where e is an affected element that may be changed by the performance of a. The elements R, K, S, and G are state variables that hold an integer value that usually represents whether the element represented by the state variable has been obtained (1 or 0).

An attack step is defined by the tuple

$$< B, T, C, O >$$

where B is a Boolean precondition that indicates whether or not the attack step is currently enabled, T is the timing distribution that is sampled to determine the time it takes to complete the attack step, C is the cost to the adversary for attempting the attack, and O is the set of outcomes of the attack (such as success or failure). Each outcome contains a Pr, D, and E, which are the probability the outcome will be selected from an attack step's O, the probability of being detected for that outcome, and the effect of that outcome on the state of the model, respectively. An adversary uses the solution of a Competitive Markov Decision Process [1] as described in [6] to select the best attack step given the adversary's characteristics, limitations and preferences.

System metrics are then defined using rate- and event-based performance variables [8]. Reliability of a device, preferred paths of attack for an adversary,

and expected costs for the adversary and defender are all examples of possible metrics. Discrete event simulation is used to generate a statistically sound estimate of the defined metrics.

3 Power Grid Description

In this case study, we consider a hypothetical utility company with an urban deployment of an AMI network, as shown in Figure 1 and described in Section 2.1. We have based our system on the system described in [2], following it in detail whenever possible. In this network, zero or more smart appliances connect to a smart meter at each home and together form a HAN. Multiple HANs are connected to one another and one or more gateways via a wireless mesh network to form a NAN. Multiple NANs are connected to one another and to the utility command and control center network via a WAN.

The utility wishes to supply power to consumers, protect their equipment, ensure the integrity of communication in the AMI network, and ensure the confidentiality of communication in the AMI network.

The utility company in this scenario is primarily concerned about attacks from three classes of adversaries: unscrupulous customers who wish to under-report their electricity consumption to unfairly lower their bill, disgruntled insider employees who wish to cause as much monetary damage as possible in retribution for a perceived wrong, and sophisticated, well-funded terrorist organizations or nation-states who wish to interrupt the delivery of power and cause as much damage as possible. The utility company estimates that over a 20-year period, there will be 1,000 attempts to under-report electricity consumption, a 0.1% chance that a disgruntled employee will attempt a massively damaging attack, and a 0.01% chance of being attacked by a terrorist organization. An adversary may choose from a variety of attacks to achieve a goal. We utilized the literature search conducted in [3] to compile a list of attacks for inclusion in our model.

The utility company wishes to compare the cost-effectiveness of various proposed IDS architectures. In particular it wishes to compare the centralized IDS solution with the two distributed IDS solutions: embedded and dedicated. The utility can easily obtain the estimated installation and maintenance costs of an IDS from vendors. However, estimating the expected benefit of implementing the IDS is much more difficult. We attempt to make such an estimate with an ADVISE model.

4 ADVISE Model

We used the ADVISE formalism as implemented in Möbius to construct an Attack Execution Graph to gain insight into the adversary behavior. We created a model that was detailed enough to calculate the quantitative security metrics of interest, while minimizing the number of assumptions that a more detailed model would have forced us to make. We were primarily interested in three

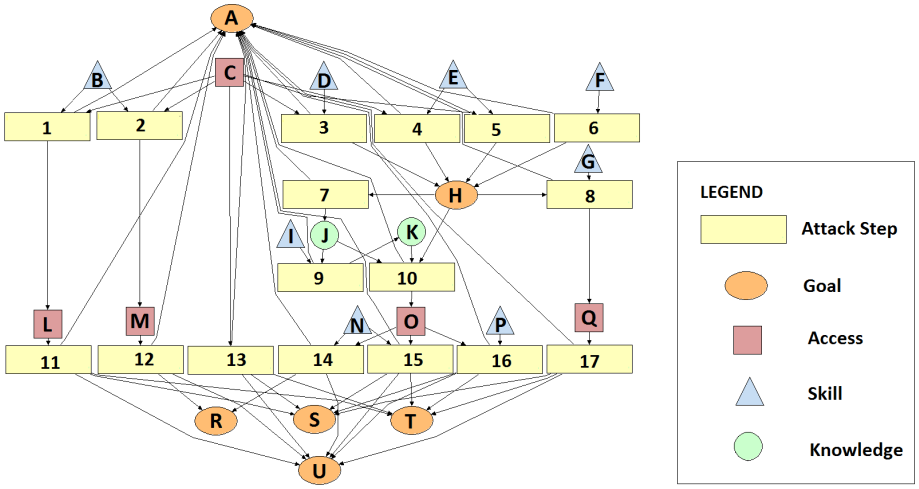

Fig. 2. Attack Execution Graph of ADVISE Model.

key metrics: the estimated probability of detecting an adversary, the estimated damage to the utility company due to adversary behavior, and the attack used to damage the utility company. To calculate these metrics, we developed a model of the attacks against the system and a model of the adversaries that could execute the attacks.

4.1 Attack Execution Graph Model

The Attack Execution Graph, which is shown in Figure 2, contains three main adversary goals; three auxiliary goals; seventeen attack steps that an adversary may attempt when trying to reach a goal; a number of supporting access, knowledge, and skill state variables that may help an adversary satisfy the preconditions for attempting a particular attack; and arcs that connect the attack steps to particular accesses, knowledge, skills, and goals and signify the relationships between them. The access, skill, knowledge, and goal state variables in our model hold a value of zero if they are not held by the adversary, and a positive integer otherwise.

The set of goals desired by the adversary drives his or her behavior and are therefore one of the most important components of the model. Cheating the company by under-reporting electricity consumption, interrupting the delivery of electric power, and damaging the utility's equipment are the three most important goals for the adversary in this study. Those goals are represented in Figure 2 by Goals R, S, and T, respectively. In addition, the adversary wishes to remain undetected; this goal is represented by Goal A, the *Undetected* goal. Goal H is the supporting goal of acquiring compromised smart meters. This is an important prerequisite for several attack steps, but is also a goal in its own right for some adversaries. Finally, Goal U represents the goal of achieving at least

one of the three primary goals described previously. Usually goal state variables initially hold a value of zero; and the value is incremented on the successful conclusion of an attack. The notable exception is the *Undetected* goal, which initially holds a value of 1 and is decremented to 0 if an attack fails and the adversary is detected.

We assume that the adversary may not continue to attack after being detected. For this reason, every attack step in the graph is connected by an arc to the *Undetected* goal, which is a prerequisite for attempting every attack step. If the *Undetected* goal holds a value of zero, the adversary may not attempt any attack, with the exception of the unique *DoNothing* attack step [6]. Attack steps can be attempted by an adversary either to achieve a goal directly or to change the model state to make it easier to achieve a goal later. An attack step must have at least one outcome. In this particular ADVISE model every attack step outcome results either in the certain detection of an adversary, or the adversary remaining undetected. In other words, if p_o is the probability of detection associated with an outcome o, then $p_o = 0$ or $p_o = 1$, but $p_o \notin (0, 1)$.

Attack Step 1 in the AEG diagram is an *Install Long Range Jammer Attack*. This attack step requires the adversary to be undetected, to be in reasonably close proximity to the smart meters, and to have skill in installing wireless jammers. It would result in the adversary's having access to a long-range wireless jammer that incapacitates the wireless mesh network in the NAN. The adversary must hold several prerequisites to attempt this attack, including the *NodeInstallationSkill*, represented by Skill B, the *PhysicalSmartMeterAccess*, represented by Access C, and the *Undetected* goal. At the successful conclusion of the attack, the adversary gains the *LongRangeJammerAccess*, Access L, whose value is incremented from 0 to 1. Attack Step 2, *Install Short Range Jammer*, is very similar, but its purpose is to gain access to a short-range wireless jammer that blocks communication in a HAN rather than a NAN, so Attack 2 is connected to *ShortRangeJammerAccess*, Access M, rather than Attack 1's *LongRangeJammerAccess*. Attack 2's prerequisites are identical to Attack 1's prerequisites.

Any one of Attack Steps 3, 4, 5, and 6 may be attempted by an adversary in an effort to obtain the *NumCompromisedSmartMeters* goal (Goal H), which would give the adversary control of smart meters in the AMI network. Attack Step 3, *InstallMaliciousSmartMeter*, aims to accomplish this goal by installing a new meter (controlled by the adversary) that tricks the AMI network into accepting it as one of its own smart meters. Attack Step 3 requires physical access to the AMI network and skill in installation as prerequisites, and so is connected to the *PhysicalAccess* access and the *SmartMeterInstallationSkill* skill, shown as Access C and Skill D, respectively. Attack Step 4, *PhysicalSmartMeterExploit*, represents an adversary attempt to physically tamper with the smart meter to gain control of it. An adversary must have physical access to the smart meters and skill in this exploit to attempt the attack step, so Attack Step 4 is connected to Access C and Skill E, which are the *PhysicalAccess* access and the *PhysicalSmartMeterExploitSkill* skill, respectively. Attack Step 5, *MassMeterCompromise*, is very similar to Attack Step 4, with the major difference being that 50 smart

meters are compromised if this attack step is achieved instead of just one. Finally, Attack Step 6 also compromises 50 smart meters, but it requires the adversary to have the appropriate skill (*RemoteSmartMeterExploitSkill*, shown as Skill F) and does not require the adversary to have physical access.

Attack Steps 7, 9, and 10 are related because their sequence leads to Access O, the *RoutingCapability* access, which is a prerequisite for Attack Steps 14, 15, and 16. Attack Step 7, *CollectCryptoKeys*, represents the adversary's attempt to collect cryptographic keys from the compromised smart meters. The adversary must have access to compromised smart meters to attempt the attack, and if the attack step is successful, it leads to the acquisition of knowledge of the cryptographic keys, which is represented by Knowledge Item J. The *AnalyzeTraffic* attack step (Attack Step 9) requires the adversary to hold Knowledge Item J and Skill I (I being the *TrafficAnalysis* skill) in order to attempt the attack step. If successful, the adversary gains sufficient knowledge of the traffic in the network to launch sophisticated routing and Byzantine attacks. This knowledge is represented by the knowledge item *TrafficKnowledge*, which is Knowledge Item K. Finally, the adversary may attempt Attack Step 10, *GainRoutingCapability*, if he or she has knowledge of the keys and traffic and at least one compromised smart meter. If the prerequisites have been satisfied, the adversary will successfully execute the attack step and gain the *RoutingCapability* access.

Attack Step 8, *CreateBotnet*, gives the adversary the *BotnetAccess* access, depicted as Access Q, which is a prerequisite for launching resource exhaustion attacks such as DDoS attacks. To attempt the attack step, the adversary must hold Skill G, the *BotnetShepherd* skill, as well as at least 50 smart meters, represented by a value greater than or equal to 50 in Goal H.

There are seven attack steps that directly achieve at least one of the three most significant goals (Attacks 11-17). First, Attack Step 11, the *Major Jamming Attack*, requires the adversary to have access to a long range jammer; it results in a significant interruption of service in the NAN, and also damages equipment, since important commands for coordinating the network are not delivered. Attack Step 12, which is the *Minor Jamming Attack*, requires access to a short-range jammer: it does not result in loss of power or damage to equipment but may be utilized to help an unscrupulous customer give a false power reading. Attack Step 13, *PhysicalAttack*, represents a major physical, non-cyber attack on the equipment of the utility company, e.g. shooting one or more transformers. This attack requires only physical access to the equipment, and causes a significant blackout and major damage to the equipment. It has a relatively high probability of detection, but requires only minimal prerequisites to attempt. Attack Steps 14 and 15, *MinorRoutingAttack* and *MajorRoutingAttack*, respectively, are similar in that they have the same prerequisites, *RoutingAttack* skill and *RoutingCapability* access (Skill N and Access O, respectively), but have different intended goals. The *MinorRoutingAttack* under-reports the electricity consumption of one customer. The *MajorRoutingAttack*, in contrast, leads to interrupted service and damage to the AMI network equipment. Attack Step 16, *ByzantineAttack*, requires that the adversary hold the *RoutingCapability* ac-

cess and the *ByzantineAttack* skill (Access O and Skill P, respectively), and a successful outcome for the adversary leads to damaged equipment and interrupted service. Finally, Attack Step 17, the *ResourceExhaustion* attack, requires *BotnetAccess* and results in damaged equipment and interrupted service.

In addition, there is one implied attack step not shown in the diagram, the *DoNothing* attack step, which an adversary may attempt at any time and has no effect on the model state, costs nothing, and will never lead to the detection of the adversary. This attack step may be attempted by an adversary when the payoff for attempting any other attack step does not justify the risk of detection and the cost of attempting the attack step.

Each attack step contains detailed information about the probability of success, probability of detection, cost to attempt, effects on the system, duration, and other information. Space considerations prevent us from explaining the details of every attack step in this model, but we discuss one attack step as an example. The *CreateBotnet* attack, Attack Step 8 in the diagram, is assumed to cost the adversary $1,000 to attempt, and to take 8 hours to complete. If the attack is to be attempted, the *Undetected* goal state variable and the *BotnetShepherdSkill* skill state variable must both contain a positive value, and the *NumCompromisedSmartMeters* goal state variable must hold a value greater than or equal to 50. If these conditions are not met, the attack step cannot be attempted. If the attack step is attempted, one of three outcomes, $FailureUndetected$, $FailureDetected$, or $Success$, is randomly chosen according to their probabilities of occurrence. The $FailureUndetected$ outcome represents the event in which the adversary attempts the attack and fails, but remains undetected. It has no effect on the state of the model, and has a probability of 0.05. The $FailureDetected$ outcome represents the event in which the adversary attempts the attack, fails, and is detected. If this outcome is randomly selected by the simulation, it modifies the model state by changing the value of the $Undetected$ goal from 1 to 0, disabling any future attack. This outcome is also assumed to have a probability of 0.05. Finally, the $Success$ outcome represents the successful completion of the attack. It has the effect of giving the adversary access to a botnet of smart meters, which is represented by changing the value of the *BotnetAccess*, State Variable Q, from 0 to 1. This outcome has a 0.9 probability of being selected if the attack step is attempted. All the other attack steps in the model have a similar level of detail.

The probability that an attack step will lead to a successful outcome for an adversary, as well as the effect an outcome will have on the system, may be adjusted based on the IDS approach being modeled.

4.2 Attacker Model

In addition to a model of attacks against the system, we need a model of the adversary, since different adversaries have different goals, and different initial access, skills, and knowledge related to the system. Even adversaries with identical goals may weigh these goals differently. These differences can lead to very different behaviors when the attackers are confronted by the same system. Table

Table 1. Initial state values and parameters for adversaries.

Initial State	Customer	Insider	Terrorist
BotnetAccess		X	
RoutingCapability		X	
PhysicalAccess	X	X	X
CryptoKeys		X	
TrafficKnowledge		X	
RoutingAttackSkill	X	X	X
NodeInstallationSkill	X	X	X
SmartMeterInstallationSkill		X	X
TrafficAnalysisSkill	X	X	X
BotnetShepherdSkill		X	X
ByzantineAttackSkill	X	X	X
PhysicalSmartMeterExploit	X	X	X
NumCompromisedMeters		51	
Undetected	X	X	X

1 shows the state variables initially held by each adversary, and corresponds to an initial configuration of state variables in the AEG (Figure 1). As can be seen from the table, the customer adversary is assumed to have access to a physical smart meter (his or her own) and some skill in various attacks, perhaps obtained via compromises published on the Internet. The customer wants to achieve the goal of cheating the power company by under-reporting electricity consumption. The insider adversary is in some ways the most powerful adversary, because the insider starts with the most access, knowledge, and skills of any adversary considered, and in addition is the only adversary assumed to start with a number of compromised smart meters. However, this adversary is constrained by a relatively high desire to avoid detection, which is expressed in the model by placing a high payoff on maintaining the *Undetected* goal. The insider wishes to cause as much monetary damage as possible to the utility company without being detected. Finally, the terrorist adversary has fewer initial forms of access, knowledge, and skill than the insider, but wants to achieve the same goal of causing the utility company as much monetary damage as possible by interrupting the delivery of power and damaging equipment. The terrorist is assumed to be less concerned than the insider with the possibility of being detected and apprehended (expressed in the model by a relatively low payoff on maintaining the *Undetected* goal), which means the terrorist is much more likely to try risky attacks.

4.3 Metrics

We use the ADVISE model described above to calculate three metrics. All three are determined through the creation of performance variables [8] calculated by

simulation in Möbius. We took the cross-product of the adversaries {Insider, Customer, Terrorist} and the IDS approaches {None, Central, Dedicated, Embedded}, and ran a simulation for every element of this set. We estimated the mean of every performance variable with a 0.95 confidence level and a 0.1 confidence interval.

The first metric is qualitative, it is the attack that the adversary attempts that leads to one of the three major goals (stealing electricity, disrupting the delivery of electricity, and damaging the equipment). To find this metric, we created a set of interval-of-time impulse-reward variables, one for each attack step that achieves one of the three main goals. If any one of the outcomes of an attack step is selected during the course of the simulation the performance variable associated with that attack step accumulates a reward. After the simulation, we determine which attack step the adversary chose by observing which element of this set of performance variables accumulated a reward.

The second metric, the probability that the adversary will remain undetected through the end of the attack, was constructed as an instant-of-time rate-reward variable that returned the value of the *Undetected* goal variable at the end of the simulation. At the beginning of the simulation, the *Undetected* goal variable would hold a value of one. Most attack steps in the AEG had an outcome that represented the event in which an adversary was detected if the attack step was executed. If that outcome occurs at some point during the course of the simulation, one of its effects is to set the value of the *Undetected* goal variable to zero. If no outcome representing the detection of the adversary is chosen during the course of the simulation, the value of the *Undetected* goal variable remains 1. In that way we determine whether the adversary was detected during one run of the simulation. Multiple runs of the simulation show the probability that the adversary will remain undetected through the duration of the attack.

The final and perhaps most important metric, the expected monetary damage to the system in the event of an attack by an adversary, was also calculated by an instant-of-time rate-reward variable. The integer values held in the *StealElectricity*, *InterruptService*, and *DamageEquipment* goal state variables represent units of damage. We let one unit of *StealElectricity* equal $600 of damage, one unit of *InterruptService* equal $10,000 of damage, and one unit of *DamageEquipment* equal $100,000 of damage. Initially these goal state variables hold a value of 0, but the value can be increased at the successful conclusion of certain attacks.

5 Results and Discussion

The attack each adversary would attempt when faced with each possible IDS and the total monetary damage in dollars the system would sustain as result of each attack, according to our simulations, are given in Table 2. The probability that the adversary will manage to evade detection to the end of the attack is given in Figure 3.

When we examine the results, we see that an insider adversary will attempt a major routing attack if there is no IDS or if there is a centralized IDS, but will

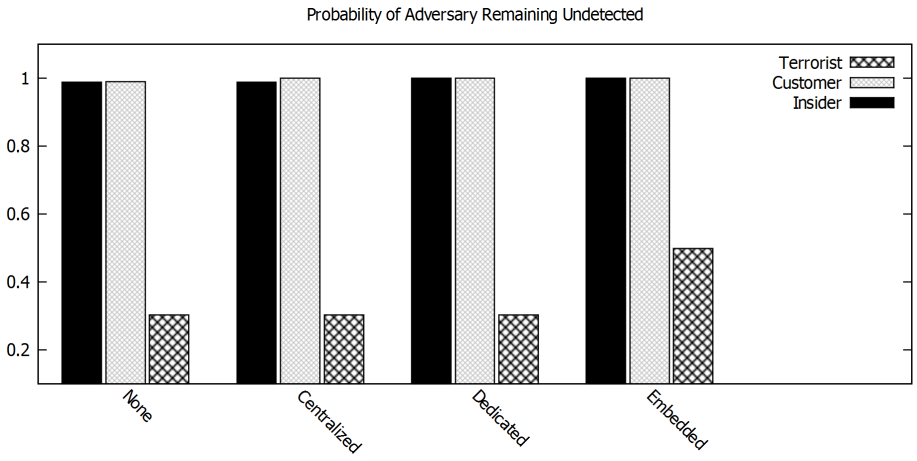

Fig. 3. Probability of remaining undetected.

not attempt any attack at all if the dedicated or embedded IDS is present in the system. When an insider attempts to attack the system and there is no IDS, the expected damage to the system is about one million dollars, but if a centralized IDS is present, the expected damage is halved, since the centralized IDS should be able to detect and limit the effectiveness of the routing attack. Since the insider will attempt no damaging attack when the dedicated or embedded IDS approach is used, the monetary damage to the system in this case is 0. This adversary is strongly incentivized to avoid detection, which can be seen in Fig. 3.

If there is no IDS present, the customer will attempt to jam the wireless communication between the smart meter and the rest of the network to under-report electricity consumption, causing about $600 of damage, and will successfully complete the attack without being detected in the majority of cases. However, if any of the IDS options are enabled, the customer will not attempt any attack, because the probability of obtaining the payoff is too small compared to the probability of being detected and having to pay a fine or penalty. Since no attack is attempted in these cases, the probability of remaining undetected is 1.

The terrorist is not highly incentivized to avoid detection and does not start out with many types of access, knowledge, or skills, so during our simulations the *PhysicalAttack* (which requires minimal prerequisites and causes massive damage with a high risk of detection) was chosen no matter what IDS architecture confronted the terrorist. When an adversary attempts this attack, the expected monetary damage to the system is about $5,000,000. However, there is a greater than 50% chance that the attack will end unsuccessfully with the detection of the adversary, which we see in Fig. 3.

A utility company can use these metrics to compare intrusion detection approaches. The expected monetary loss sustained by a utility company, M, for an

Table 2. Simulation results.

IDS	Adversary	Attack	Monetary Damage	Error
None	Insider	Major Routing	$1.07M	+/- $7.29K
	Customer	Minor Jamming	$594	+/- $0.583
	Terrorist	Physical	$4.98M	+/- $50K
Centralized	Insider	Major Routing	435K	+/- $2.97K
	Customer	Do Nothing	$0	+/- $0
	Terrorist	Physical	$4.98M	+/- $50K
Dedicated	Insider	Do Nothing	$0	+/- $0
	Customer	Do Nothing	$0	+/- $0
	Terrorist	Physical	$4.98M	+/- $50K
Embedded	Insider	Do Nothing	$0	+/- $0
	Customer	Do Nothing	$0	+/- $0
	Terrorist	Physical	$5.02M	+/- $52.1K

IDS configuration, $i \in IDS$, can be calculated with Equation 1

$$M_i = \sum_a N_a * D_a \tag{1}$$

where N is the expected number of attack attempts and D_a is the expected monetary damage to the system, D, per adversary, $a \in Adversaries$.

Consider a hypothetical utility that estimates 1,000 attempts by unscrupulous customers, 0.001 attempts by an insider, and 0.0001 attempts by a terrorist over a 20-year period.

Using Equation 1 and the numbers in Table 2, we calculate the results shown in Table 3. The utility can use Table 3, along with information about installation and maintenance costs provided by vendors, to help determine the most cost-effective architecture for its system.

Table 3. Estimated monetary loss by IDS approach over a 20-year period.

IDS	Monetary Damage
None	$595,568
Centralized	$933
Dedicated	$498
Embedded	$502

Space limitations force us to examine a small subset of the possible system configurations, adversaries, and attacks, but we find that the ADVISE formalism is flexible and scales well. More detail could be added by a utility company as needed. In addition, we chose to use synthetic data in our analysis as input parameters for the model. Utility companies would not have allowed us to publish unsanitized data, and it is uncertain whether any hypothetical sanitized data

would have been more accurate than our educated guesses. This is not a weakness of our approach, since a utility company would already have the data needed for the input parameters for its own ADVISE model. Our synthetic data were based on the existing literature regarding security of AMI, especially [2] and [3].

6 Related Work

The academic community, recognizing the importance of the topic, has done prior work comparing different security approaches in AMI. The analysis in [4] uses a cost-benefit study to determine whether the added cost of RFID technology is justified given its ability to prevent electricity theft. That paper considers considers only energy theft, while our analysis considers attacks on the availability and integrity of the system in addition to energy theft. In [9] the authors propose an IDS for AMI and compare its security and performance with other IDSes for AMI. However, in contrast to our analysis, they explicitly do not include attacks on the meter's availability. The techniques proposed in [7] seek to compare and evaluate the security of different AMI IDSes through penetration testing and the use of archetypal and concrete attack trees. These attack trees could help a modeler create an Attack Execution Graph for an ADVISE model. In contrast to our approach, [7] does not explicitly model the attacker's attributes or motivations in detail. The authors of [2] provide a framework for evaluating the cost-effectiveness of different IDS architectures in an AMI network. However, their approach does not explicitly account for the differences in the behaviors of adversaries when attacking the system. By incorporating the adversary behavior into the model, we hope to achieve more realistic results.

7 Conclusion

In this work we showed how to use the ADVISE state-based stochastic modeling approach to calculate security metrics that are relevant in comparing different IDS architectures in an AMI network.

Unfortunately, it is often not possible to estimate many characteristics of attack steps precisely with a high degree of confidence, including the probabilities of success and the magnitude of damage given a successful attack, as well as the exact amount of protection provided by an IDS against an attack. In addition, adversary characteristics and motivations cannot usually be definitively known. Therefore, the quantitative metrics produced by the ADVISE model should not be thought of as producing exact, accurate predictions of the future. We believe, however, that these metrics can contribute to a development of a relative ranking of IDS approaches in an AMI network and provide insight into general trends of adversary behavior.

We argue that the scientific approach ADVISE offers for security evaluation is a useful complement to a common method of estimating the relative effectiveness of different security approaches: consultation of one or more security experts, who rely on intuition and experience. In contrast, the metrics calculated by ADVISE

are easily auditable by other parties and assumptions are explicitly stated, which allows multiple security experts with different backgrounds to use the ADVISE formalism as a modeling language to collaboratively analyze different system designs.

Acknowledgments

The work described here was performed, in part, with funding from the Department of Homeland Security under contract HSHQDC-13-C-B0014, "Practical Metrics for Enterprise Security Engineering." The authors would also like to thank Robin Berthier, Corky Parks, Carol Muehrcke, and the anonymous reviewers of this paper for their valuable advice, as well as Jenny Applequist for her editorial assistance.

References

1. Bellman, R.: Dynamic Programming. Princeton University Press, Princeton, NJ, USA, 1 edn. (1957)
2. Cardenas, A.A., Berthier, R., Bobba, R.B., Huh, J.H., Jetcheva, J.G., Grochocki, D., Sanders, W.H.: A framework for evaluating intrusion detection architectures in advanced metering infrastructures. Smart Grid, IEEE Transactions on 5(2), 906–915 (March 2014)
3. Grochocki, D., Huh, J.H., Berthier, R., Bobba, R., Sanders, W.H., Cardenas, A.A., Jetcheva, J.G.: AMI threats, intrusion detection requirements and deployment recommendations. In: Smart Grid Communications (SmartGridComm), 2012 IEEE Third International Conference on. pp. 395–400 (Nov 2012)
4. Khoo, B., Cheng, Y.: Using RFID for anti-theft in a Chinese electrical supply company: A cost-benefit analysis. In: Wireless Telecommunications Symposium (WTS), 2011. pp. 1–6 (April 2011)
5. LeMay, E., Ford, M.D., Keefe, K., Sanders, W.H., Muehrcke, C.: Model-based security metrics using ADversary VIew Security Evaluation (ADVISE). In: Proceedings of the 8th International Conference on Quantitative Evaluation of SysTems (QEST 2011). pp. 191–200. Aachen, Germany (Sept 5–8, 2011)
6. LeMay, E.: Adversary-Driven State-Based System Security Evaluation. Ph.D. thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois (2011)
7. McLaughlin, S., Podkuiko, D., Miadzvezhanka, S., Delozier, A., McDaniel, P.: Multi-vendor penetration testing in the advanced metering infrastructure. In: Proceedings of the 26th Annual Computer Security Applications Conference. pp. 107–116. ACSAC '10, ACM, New York, NY, USA (2010)
8. Sanders, W.H., Meyer, J.F.: A unified approach for specifying measures of performance, dependability, and performability. In: Avizienis, A., Kopetz, H., Laprie, J. (eds.) Dependable Computing for Critical Applications, Vol. 4 of Dependable Computing and Fault-Tolerant Systems. pp. 215–237. Springer-Verlag (1991)
9. Tabrizi, F.M., Pattabiraman, K.: A model-based intrusion detection system for smart meters. In: High-Assurance Systems Engineering (HASE), 2014 IEEE 15th International Symposium on. pp. 17–24 (Jan 2014)
10. Möbius team: Möbius Documentation. University of Illinois at Urbana-Champaign, Urbana, IL (2014), https://www.mobius.illinois.edu/wiki/