Risk Analysis, Vol. 38, No. 2, 2018 DOI: 10.1111/risa.12844

Cyber Risk Management for Critical Infrastructure: A Risk Analysis Model and Three Case Studies

M.-Elisabeth Paté-Cornell,* Marshall Kuypers, Matthew Smith, and Philip Keller

Managing cyber security in an organization involves allocating the protection budget across a spectrum of possible options. This requires assessing the benefits and the costs of these options. The risk analyses presented here are statistical when relevant data are available, and system-based for high-consequence events that have not happened yet. This article presents, first, a general probabilistic risk analysis framework for cyber security in an organization to be specified. It then describes three examples of forward-looking analyses motivated by recent cyber attacks. The first one is the statistical analysis of an actual database, extended at the upper end of the loss distribution by a Bayesian analysis of possible, high-consequence attack scenarios that may happen in the future. The second is a systems analysis of cyber risks for a smart, connected electric grid, showing that there is an optimal level of connectivity. The third is an analysis of sequential decisions to upgrade the software of an existing cyber security system or to adopt a new one to stay ahead of adversaries trying to find their way in. The results are distributions of losses to cyber attacks, with and without some considered countermeasures in support of risk management decisions based both on past data and anticipated incidents.

KEY WORDS: Cyber risk management; cyber security; infrastructure protection

1. CYBER ATTACKS: PROBLEM AND STAKEHOLDERS

Repeated, frequent cyber attacks in the last few years have shaken the public's confidence in the ability of infrastructure managers to protect their systems. Catastrophism tends to dominate the discourse, with comparisons to the Pearl Harbor attack or even a nuclear disaster. Hand wringing will not make matters better. One needs to address the questions: What should be done, in what order, and at what cost? The objective of this article is to show how quantitative risk analysis can support these decisions given the uncertainties involved, e.g., about the identity of potential attackers and the probability and consequences of a successful attack.

The countermeasures considered in this article address the different phases of a cyber attack on specific organizations. They include two-factor authentication (TFA) to reduce the risks of intrusion, software designed, and updated, to protect information or prevent its exfiltration, and optimization of the number of connections and vulnerabilities in a system. Each of these measures implies costs and benefits, both often uncertain. Some of the costs are tangible, for instance, that of a new software, while others are not, such as the inconvenience to the users. The risk assessments described here yield a set of risk curves for a given system, with and without considered improvements. These results can then become inputs to a decision analysis, allowing

Department of Management Science and Engineering, Stanford University, Stanford, CA, USA.
*Address correspondence to M.-Elisabeth Paté-Cornell, Department of Management Science and Engineering, Stanford University, Stanford, CA 94305, USA; mep@stanford.edu.

a decisionmaker to identify the measure or set of measures that will maximize his or her expected utility.[1]

Cyber attacks can affect—and in some cases already have—various forms of the U.S. critical infrastructure, including the electric power grid, seaport operations, air traffic control, and several other services vital to the nation's economy. A successful full-scale cyber attack on the global positioning system (GPS) would be devastating given its role in numerous sectors such as communication, finance, navigation, and transportation.[2] Stakeholders include not only the target organizations, which could lose their intellectual property and revenues, but also their clients, whose service could be interrupted and personnel information and intellectual property stolen.

Recent, high-profile attacks have targeted several entities of the U.S. government. These include:

- **The Department of Energy (DoE).** The U.S. government reported that the DoE was successfully hacked 159 times over a four-year period.(1) This is of serious concern since the DoE participates in the management of the U.S. nuclear weapon stockpile and of critical energy infrastructure, and oversees sensitive research at national laboratories.
- **The Office of Personnel Management (OPM).** Millions of personnel records, including Social Security numbers and fingerprints, were stolen from the OPM starting in 2014.(2) These incidents are suspected to be state sponsored, and the stolen data could be used for intelligence collection.

In addition, private industries and companies such as Target and Sony have been successfully attacked and experienced high costs, both to the organizations and to their customers. The question of course is: What should one do next—besides lamenting an obvious change in the nature of criminality and warfare—given the assets at risk, the nature of the attackers, and the known or potential vulnerabilities of information systems? As usual, the challenge is to set priorities to avoid misallocating and wasting resources.

[1]In all the applications presented here, the decisionmaker is assumed to be risk neutral. Different risk attitudes can be introduced through other utility functions.

[2]Note that we do not consider here the secondary costs of a cyber attack beyond the limits of the target organization.

2. CURRENT APPROACHES TO CYBER RISK MANAGEMENT AND INSTITUTIONAL RESPONSES

2.1. Countermeasures

Industry and government efforts to manage cyber security risks have focused on developing security software, best practices for system design and operations, and improving investments in the cyber workforce.(3,4) Cyber security countermeasures include setting up firewalls, software encryption, virus detection, and system compartmentalization. When asked what is most sensitive and what measures are most effective, industry managers tend to answer: "everything." Resource limitations, however, require setting priorities based on the cost effectiveness of security measures, instead of blindly adopting a blanket approach. In most cases, the need for security has to be balanced with productivity since various options can be cumbersome and interfere with the functioning of a system and an organization. For example, a randomly generated password of 16 characters is unlikely to be easily remembered, even if shown effective.

So far, quantitative methods in the field of cyber security have been approached in a piecemeal fashion. Scholars have begun developing quantitative risk management models,(5–7) e.g., the U.S. Air Force has applied decision analysis and probabilistic methods to its investments in cyberspace in both defensive and offensive capabilities.(8) Rao *et al.*(9) presented a game-theoretic attacker–defender model for a cyber network of physical components, and analyzed the tradeoffs faced by the defender in a choice of network protection.

2.2. Current Institutional Approaches

Many efforts to manage cyber risks are based on a "top-down" management approach, with the objective to encourage system designers and operators to adopt best practices, but often without specific considerations of the system's structure. For instance, the U.S. Department of Defense directs its agencies and services to consider cyber security when designing and operating industrial control systems (ICS), but without much guidance on how to do it.(10) A number of general documents are available on the topic.(4,11,12)

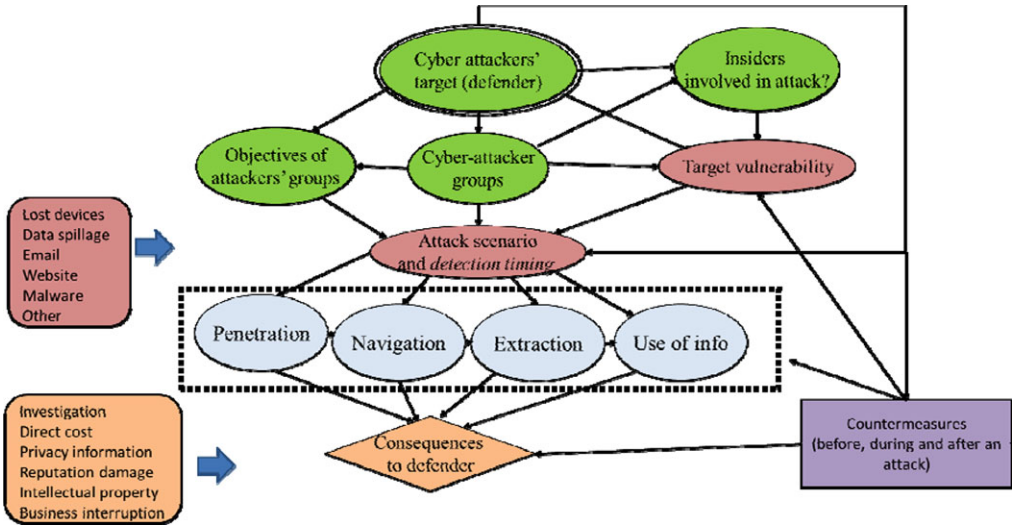


Fig. 1. Decision diagram structure for the choice of options regarding cyber security. Adapted from Ref. 6.

Most cyber risk management practices for infrastructure protection can be divided into three groups:

(1) **Structural** (composition and structure of the ICS and hardware systems)
(2) **Procedural** (system's management and operation)
(3) **Responsive** (responses and damage control after an incident is detected)

Tailored probabilistic risk analysis (PRA) methods are designed to assess the effectiveness of each of these options for specific organizations in more detail than general guidelines. The focus of this article is on the overall structure of the risk analysis problem with three illustrative examples to further address tradeoffs and priorities: an empirical analysis of a database of cyber security incidents, an optimization model designed to balance a system's operational efficiency with its protection from cyber attacks (here applied to a smart electric grid), and a dynamic game model of some interactions between infrastructure operators and their adversaries in cyber space.

3. OVERALL RISK ANALYSIS AND UNCERTAINTIES

3.1. Risks and Decisions

The general cyber risk analysis framework presented here is designed to support decisions from the defenders given the uncertainties that they face. It is an overarching model of the cyber risk and the effectiveness of different countermeasures, to be adapted to the case of a specified organization. This model is represented in Fig. 1 by a schematic influence diagram. It is based on an interaction, reduced here to one move on each side, between a system's manager and potential attackers of various kinds. This influence diagram represents a model structure to be developed further to represent the characteristics of an organization of interest and of its information system.

The decisionmaker generally knows most of the relevant information about the organization, hence a deterministic node at the top of Fig. 1. That deterministic node specifies the type of target and its assets. It can be, for instance, a national lab, an industrial company holding sensitive intellectual property, a commercial enterprise with financial data such as credit card information, or a hospital vulnerable to extortion. Each of these types of organization is more or less likely to be threatened by different types of attackers—a country, individuals, or criminal groups. The vulnerability[3] described in Fig. 1 refers to possible access points and is characterized by a probability of intrusion through these points. As shown further in the case of the smart grid, these intrusion points can be identified by an analysis of the structure of the

[3]The term vulnerability is used here to refer to potential points of entry in the target system, thus to the probability of an intrusion (as opposed to its classic definition as a system's capacity in physical systems).

information system and its connections to operations. These are generally known or discoverable, but it is not always the case. Several questions may have to be addressed first and treated as uncertainties if needed:

- What are the organization's main assets, where are they, and what are they connected to? For instance, credit card companies such as Master Card and Visa were victims in 2012 of attacks by hacking of third-party processors. In that case, the question is to identify the structure and location of customers' information and of their linkage to clients and sellers that use their services.
- What protection measures can be considered, given practical constraints as well as legal ones, against offensive response?
- What are the risks of an insider's attack given who has access to all or some parts of the information network? This major component of the risk depends on the nature of the organization and the management of its personnel and contractors (hiring, vetting, firing procedures, and access to information). One example is the Snowden attack on the National Security Agency (NSA) information systems held by Booz Allen, where insiders were given wide access to these files. Another case was that of the 2013 hacking of Target's credit cards by maintenance contractors.[4]

The remaining uncertainties of the general model can then be divided into three main groups: attacker demographics, attack scenarios, and their consequences.

3.2. Attacker Demographics and Entry Points

Attackers vary in resources and sophistication, from low-end hackers to well-supported, experienced ones. The spectrum involves:

- Hackers and petty criminals, sometimes called "script kiddies" or "ankle biters," who generally do not do much damage, but can, for instance, severely alter or deface websites.
- Criminal organizations that seek monetary gain through intellectual property theft, financial information, or ransom for stolen data. Some, for example, have recently specialized in "ransomware," currently targeting hospitals, or stealing financial records such as credit card information.
- State-sponsored persistent attackers who try to impede the functioning of organizations and countries, and to steal intellectual property by seeking immediate intrusion or means to do it later.
- Other companies or organizations that may seek to benefit from damaging a competitor or stealing its intellectual property through industrial espionage.
- Terrorist organizations that may simply have the objective of causing maximum damage to societies and to gain resources to maintain their financial operations.
- Malicious insiders who can include disgruntled employees, or individuals whose values are not aligned with those of the organization, including people who simply seek monetary gains.
- Contractors such as maintenance operators who may have or find access to the internal information network to disrupt the organization, either as pure malfeasance or to sell the stolen information.

In each case, the questions are: Who might these attackers be, what do they want, what do they know, and what do they have? In the global model, a specified company, based on its experience and the nature of its assets, may be able to put a probability distribution on the kind of attackers that it might face and their attack methods.

The next uncertainty for the defender is that of its vulnerabilities and the entry points through which the attackers can penetrate its system. These include:

- Unintentional defects or "bugs" in the defender's software that may have been introduced during initial programming.
- Intentional defects that may be inserted into computer hardware or software along the supply chain by potential attackers with the intent to use them later.
- Peripheral defects—flaws or "bugs" in systems outside of the control of the decisionmaker—that may be introduced through equipment connected to the main system. A virus on a contractor's laptop may be used to penetrate an otherwise secure system if they are connected to provide some services and, as mentioned earlier, a credit card company's assets may be vulnerable to the hacking of third-party processors.

[4]This category is also considered to include vulnerabilities unintentionally introduced by employees into the system.

3.3. Attack Scenarios

Different organizations can thus be subjected to various types of attack with probability distributions that depend on the nature of their operations. The most frequent attack scenarios include lost or stolen devices, data spillage, malicious email (phishing) attacks, website attacks, and malware introduced via web browsing or USB drives. The Common Vulnerability and Exposures (CVE) database compiled by MITRE Corporation provides a numeric scoring system designed to categorize the probability of different, known software flaws that can be used to cause damage, or provide an adversary with control or access to a system.[13]

A successful attack, for instance, with the objective to steal or damage intellectual property, generally includes four phases (dotted box in Fig. 1):

- Enter the system
- Move laterally inside the system to gain unauthorized access to its different parts
- Cause damage to both software and hardware
- Retrieve, exfiltrate, and use the information of interest

Countermeasures can be implemented to address each or several of these attack phases, including fire walls to protect against entry, a system "maze" to discourage free navigation, barriers to exfiltration, and monitoring of data use to detect the attack as soon as possible and minimize the damage (for instance, the loss of stolen credit card files). The realizations of each of these probabilistic nodes and the corresponding decision nodes from the target organization represent the different ways in which a specified adversary can attack a particular network and its chances of success.

3.4. Attack Impacts

The impact of a successful attack is thus a function of the nature of the organization and the damaged or stolen assets, and of the effectiveness of damage control measures. As mentioned earlier, a key factor is the time elapsed between an incident and its discovery. For instance, the loss of credit card information may be detected by the victims themselves but it may then take weeks for banks and organizations to effectively control the damage. Therefore, the costs of cyber attacks to the business operations and the reputation of the defender are highly variable depending on the nature and the value of the targets. Impacts on business operations can be substantive, with loss of production during down time, investigation of the incident, remediation measures, and loss of assets such as critical technology in a space company. In addition, the organization may experience reputation damage, which, in some cases, can be measured by the loss of value of a stock price; but it seems that this last effect has been somehow limited, as was the case, for instance, of the 2013 attack on Target, where the decrease in the value of the company was immediate but relatively minor and short-lived.

To our knowledge, the assessment of attack damage has so far been mostly qualitative. Yet, quantitative, analytic, and economic techniques exist and can be used to aggregate these attack impacts, based either on actual data or on expert opinions. A single measure of monetary losses can be a simple additive function if one assumes that these costs are independent; but for lack of other measure, the decisionmaker's willingness to pay to avoid these losses, or the insurance and reinsurance costs, can be used as surrogates.

In this context, summary loss statistics such as mean and variance are not sufficient to support a rational decision and can be misleading because the loss distributions may be heavy-tailed. Therefore, investment benefits based on such a single value may be distorted and resources misallocated.

3.5. Illustrations

Different parts of this general model are illustrated below by three examples focused on specific aspects of the method and analytical features (treatment of statistics, analysis of networks, and dynamics of attack and defense).

- The first example is a statistical analysis of an actual database of repeated cyber attacks on a large U.S. organization. These include malicious email attacks, website attacks, data spillage incidents, lost or stolen devices, and malware incidents. The challenge is to extend the risk curves to the domain where attacks have not yet occurred, based on a probabilistic analysis of extreme scenarios.
- The second example is a probabilistic model of attacks on a smart electric grid, with a notional illustration based on a network and system analysis. The question here is: How smart is smart enough and at what point is an additional

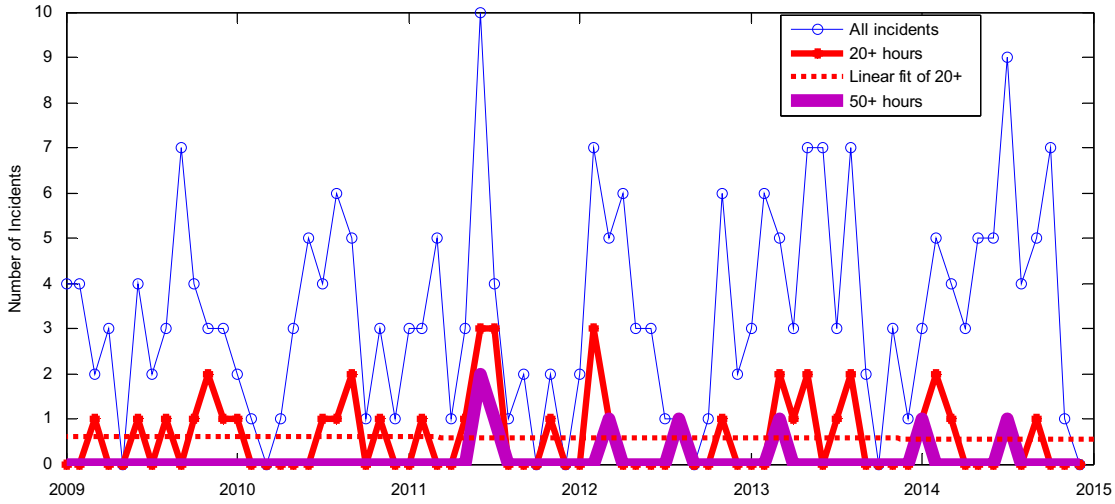


Fig. 2. Data spillage trend in the considered organization over six years.

connection going to cause marginal risk costs that will exceed its benefits?

- The third example is a dynamic, probabilistic model designed to guide the timing of a defender's decisions in the management of an industrial software system that needs to be regularly improved to prevent adversaries from disrupting operations or stealing intellectual property. The notional illustration presented here is the management of the computer that controls a city's water distribution system.

4. EXAMPLE #1: AN EMPIRICAL ANALYSIS OF A DATA SET OF CYBER ATTACKS

4.1. Statistical Analysis

The authors were given confidential access to a database of cyber attacks on a large, U.S.-based organization. This allowed us to illustrate our risk analysis method, including, first, a statistical analysis of a past event, then an extension of the results to the possibility of more serious events in the future.[5] This database includes more than 60,000 cyber attacks over a recent six-year period (2009–2015). The severity of an attack is simply measured in hours of investigation. We analyzed these data to determine whether the attack frequency had increased over time, and to compare some options to reduce the current cyber risk.

These records of attacks are seldom made public but large amounts of data exist within some organizations. They reflect the past, but can be used with some additional information to assess the frequency and impact of similar events in the future. The future risk will depend on the evolution—if any—of attack methods and of the effectiveness of defensive measures. In the database that was analyzed, the attacks were divided into the different classes relevant to that specific organization: lost or stolen devices, data spillage, email attacks ("phishing"), website attack, malware, and other rarer kinds such as shellshock attacks. Focusing on data spillage, Fig. 2 represents the frequency and the costs of all data spillage incidents, including those that required more than 5, 20, and 50 hours of investigation. Based on a linear regression represented by the dotted line, the rate of incidents over six years in that organization appears relatively stable.

Fig. 3 focuses on another type of attack (malicious email incidents) in the same organization. It displays the complementary cumulative distribution function (risk curve) of the investigation time required by these events as a dashed line based on a power-law fit of the data. The statistical analysis shows that even accounting only for the incident investigation time, the losses must be characterized by a heavy-tailed distribution.[6] One outlier that would

[5]For reasons of confidentiality and at their request, the name of that organization and some characteristics of this database are not disclosed here.

[6]In reality, a more complete cost model is needed to characterize the consequences of attacks. In the considered organization, it

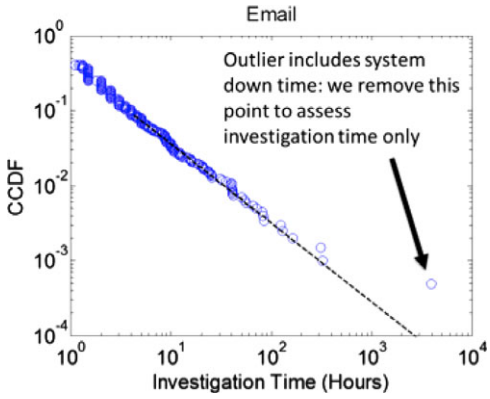
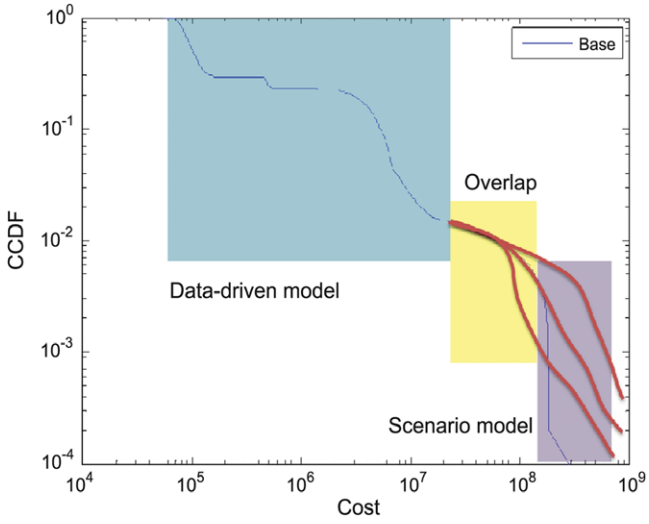


Fig. 3. Modeling the distribution of investigation times for email incidents in our database. Risk curve obtained by fitting a power-law distribution to the statistical data.

have required a more complete cost analysis was not included in the curve fitting. The analysis suggests two things: first, a full probabilistic risk assessment (as opposed to a single descriptive number) is critical to support rational and efficient decisions; second, a statistical analysis of past incidents alone does not provide a full description of the risk of future attacks.

includes not only investigation time, but also business interruption, privacy information losses, intellectual property loss, direct costs, and reputation damage.[5] Large uncertainties remain but relevant information can be found in the academic literature, public reports, and domain experts' opinions (see, for example, Cavusoglu *et al.*[29]).



4.2. The Limits of Statistics and Probabilistic Extensions

Given the nature of observed incidents and the quick evolution of the cyber world, problems remain about the upper bound of potential cyber losses and the tail of the damage distribution. By definition, severe problems that have not happened yet do not appear in the existing data, but these potential attacks need to be included in a complete risk analysis, and special cyber defense measures may have to be implemented to address these cases. To provide a complete picture, one thus needs to identify potential scenarios of more severe attacks than those that have been experienced so far, and to assess the corresponding risks.

These extreme events are included in the results through a probabilistic analysis of potential scenarios given all available information. Some of the elements of these scenarios sometimes already exist in the database (e.g., attempts to exploit a specific vulnerability without reaching the highest possible level of damage). These statistical data are used to partially inform the probability of extreme attack scenarios. They are completed by expert opinions, if needed, to provide a full analysis of these high-consequence scenarios. Other assessments of the risk of large attacks rely mostly on expert opinions, for example, the capability of a determined adversary to access essential functions of some critical U.S. communication systems, with the effect of causing very high levels of damage. In-between, there is thus an "overlap" zone between the data-driven part of the curve and the model of scenarios that have never occurred so far (Fig. 4).

Fig. 4. Combining a data-driven model with a scenario-based model involves overlapping the two risk curves.

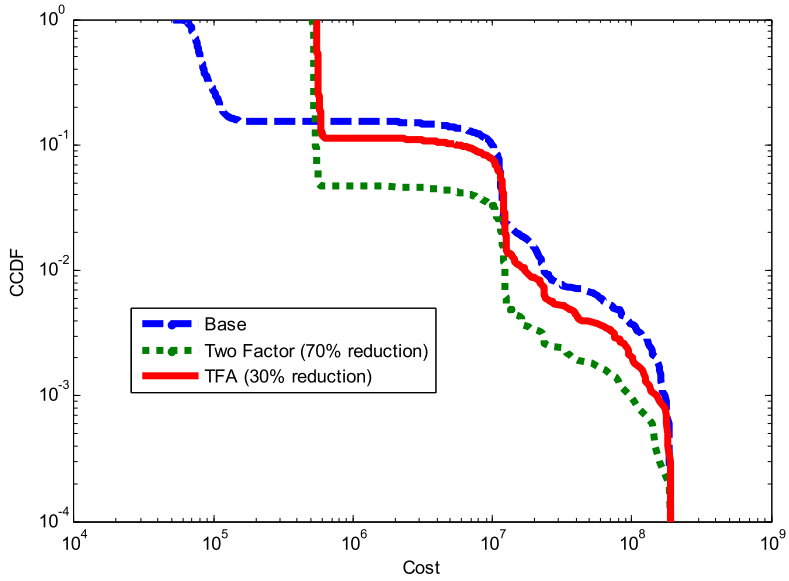


Fig. 5. Risk curves illustrating the benefits of two-factor authentication at two different levels of implementation.

The next question is to estimate the effectiveness of protection measures for the full spectrum of attack severity levels. The benefits' computation is illustrated in Fig. 5 by the case of a TFA requirement at two levels of implementation for access to different services, such as virtual private network, webmail, or workstations. A Monte Carlo simulation is run to obtain risk curves comparing the effectiveness of these two options measured by the risk reduction benefits across the span of incident severity. The base case (blue line) represents the current risk, while the green and red lines show the risk curves that one obtains, assuming that implementation of a TFA eliminates 30% (green) or 70% (red) of successful attacks.

This probabilistic cyber risk analysis based on existing statistics and extended to extreme events can be replicated in other organizations if they have gathered statistical data about cyber attacks and can use the opinions of in-house experts to characterize severe-loss scenarios. One can then compare the effectiveness of security safeguards for a whole spectrum of risk management options.

5. EXAMPLE #2: A NETWORK PROBLEM: THE CASE OF THE ELECTRIC GRID AND THE COSTS AND RISKS OF "SMARTNESS"

5.1. Modeling Cyber Risk in a "Smart" Electric Network

The power outage in Ukraine on December 23, 2015—the first known instance of a cyber attack causing a power outage—demonstrates the value of quantifying the risks and the benefits of a "smart" interconnected electric network. In this incident, three Ukrainian electricity distribution companies were the victims of a coordinated attack that included (1) a spear-phishing campaign to gain a foothold in the corporate networks; (2) keystroke loggers to obtain credentials to access the ICS networks; and (3) access to the supervisory control and data acquisition (SCADA) systems, which allow remote monitoring and control and were used to hijack human–machine interfaces and issue commands to open breakers.[14] An initial analysis suggests that these three distribution companies may have been targeted *because* they were the most connected.[15] Specifically, the level of automation and connectivity within their distribution networks enabled the attackers not only to pivot from the enterprise network to the control network, but also to gain remote control of the SCADA systems to open substation breakers. This connectivity allowed the attackers to disconnect 30 distribution substations, causing approximately 225,000 residents to lose power for several hours. This intrusion happened despite the strong cyber security countermeasures that the distribution companies had put in place, including a control network that was well segmented from the business network behind a robust firewall.

Indeed, the emergence of the smart grid promises to deliver many benefits to the overall operation of the North American electric distribution, including increased efficiency, improved reliability,

better incorporation of renewable energy sources, and more choices for electricity consumers. Yet, because of an increase in the number of interconnections, the same technologies that improve the performance of a smart grid also expose it to digital threats such as denial of service attacks, intellectual property theft, invasion of privacy, and sabotage of a critical national infrastructure. "Smartness" increases the number of potential points of access and vulnerabilities, and given the integrated nature of these cyber-physical systems, cyber-induced failures of the grid can cascade to other critical infrastructure—e.g., transportation networks, water treatment, or financial systems—and cause extensive physical damage and economic disruption.

Therefore, the benefits of increased connectivity must be weighed against the risk of increased exposure to cyber attacks as electricity sector stakeholders consider upgrading a traditional power grid architecture by incorporating smart grid technologies and new intelligent components. At some point, decision-makers must ask: How smart is smart enough? The PRA model described earlier allows quantification of the increase of risk of a cyber attack to the electric grid (in terms of both probability and consequences) as it becomes "smarter." In this case, the probability of an intrusion in the system and of the attackers' ability to navigate through it—both key factors of the success of an attack scenario—are linked to the overall cyber risk as a function of the connectivity level. The benefits of the additional connections are assessed separately.

In order to identify the optimal degree of connectivity, we consider, first, a physical power grid network represented as an undirected graph, where each node is a distinct electrical point in the network (generators, consumers, or substations) and edges are transmission lines between them (see Fig. 6). At any point, the system's operator may choose to connect a subset of nodes to the cloud, creating an overlaid information network on top of the physical network. Each new connection to an information node, however, is a potential vector for cyber intrusion, and each information link provides an opportunity for a cyber attack that can spread to adjacent nodes of the physical network. By considering sequentially each incremental connection, this network model allows discretization of the utility's decision of how "smart" to make its system. The system's operators and the authorities that balance electrical distribution commonly aggregate customers into large loads. Therefore, it makes sense in the model to bundle the total system load into aggregated nodes in the same fashion. Our objective is to illustrate how the PRA approach can then be used to help network operators prioritize protection efforts given the costs and the benefits of marginal increases of connectivity.

Intuitively, as the grid becomes smarter, one expects two things to happen:

(1) **The marginal benefit will decrease**. For many smart grid technologies, most of the benefits can be obtained with a moderate level of connectivity. For example, a recent report found that installing conservation voltage-reduction technology on 40% of distribution feeders achieves 80% of the benefit obtained by upgrading all feeders.[16] A similar trend has been demonstrated for the smart meters that comprise the advanced metering infrastructure.[17]
(2) **The cyber security risk will increase**. While the shape of the dependency is not immediately apparent, increased connectivity can cause an increase in both the probability of cyber attacks (due to an increase in the number of attack paths) and their consequences (due to tighter couplings among control components in the power network).

The PRA proposed here to quantify this tradeoff for a variety of smart grid technologies is depicted in Fig. 7. This approach augments existing statistical data with engineering models of the electric grid, stochastic models, simulation of attack scenarios, and network analysis in order to trace the dynamic evolution of failure modes induced by cyber vulnerabilities. In the first step, we perform a system's analysis to identify classes of failure scenarios that can be induced by exploiting cyber vulnerabilities in a smart grid network, as drawn from industry expert opinion[18] and known incidents. We then use an economic dispatch model to evaluate the financial benefit and risk of increased smartness, based on physics and economics of its effect on the power system. The outputs of the economic analysis are then inputs to a probabilistic model used to determine a strategy for the optimal allocation of cyber security resources. Finally, a decision analytic model (assuming risk neutrality) is used to compare the effectiveness of possible risk mitigation measures, such as improved cyber defenses, segmentation of critical infrastructure networks, installation of backup systems, and cyber

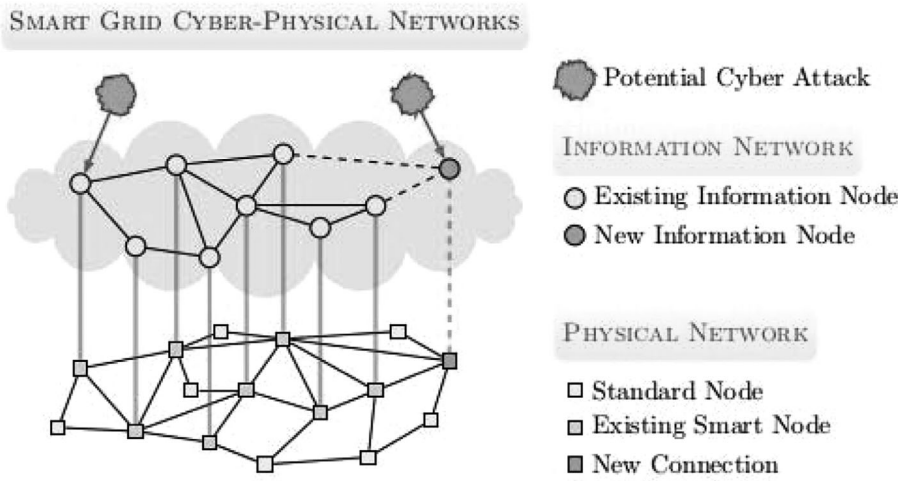


Fig. 6. Conceptual overview of a system where the electric utility considers sequentially the marginal benefits and the risks of connecting one additional node to its smart grid network.

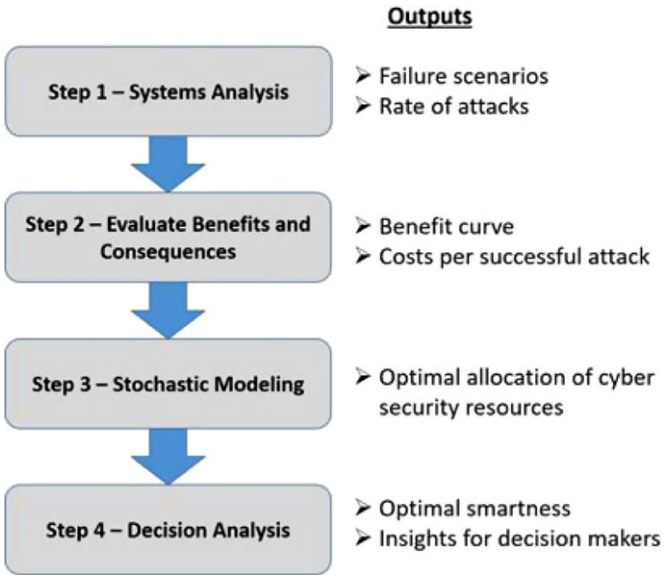


Fig. 7. Overview of our PRA approach to a cyber risk analysis of a smart grid.

insurance. From these results, we can assess the optimal degree of connectivity.

This modeling approach has many advantages: it provides tools to facilitate the combination of statistical data with expert opinion and engineering models; it allows decisionmakers to perform a sensitivity analysis for key parameters, including their own risk tolerance and several other organization-specific security factors; it is well suited to handle the many levels of uncertainty introduced by the complexity of the engineering systems and the influence of human behavior; and it indicates what should be the focus of additional modeling efforts.

5.2. Illustrative Application of the Optimization Model

We apply this model to an illustrative power grid network representative of a three-state portion of the Western Interconnection, as previously studied by Larrosa *et al.*[19][7] We then consider the decision regarding the degree to which to integrate demand–response into any subset of 10 consumption

[7]This is a purely illustrative example of interstate grid connections, meant to show the potential for cascading effects of failures across these grids.

Table I. Cyber Failure Scenarios Introduced by the Incorporation of Demand–Response

Failure Scenario	Modeling Impact
FS1 – Loss of Situational Awareness	The smart node no longer mitigates impact of transmission line outages.
FS2 – Local Outage Triggered Remotely	The cost of $3.76 per kWh of demand is not met, as determined by the highest four-hour period for that node.
FS3 – Denial of Service Blocks DR Messages	The peak load increases by 2% over normal profile (with no demand–response).
FS4 – Price/Meter Manipulation	The effective customer prices are reduced by 1% from nominal values, decreasing the revenue of the utility.
FS5 – Theft of Private Information	There is a fixed cost of $2.3k for any node.

Note: Data breach costs are based on per-capita costs from recent data breach reports.[23]

nodes. Demand–response is an emerging smart grid procedure that allows utilities to use price signals to influence customer behavior, inducing lower electricity demand when operating costs are high (e.g., during peak demand) or when the system's reliability is in jeopardy (e.g., during an unexpected outage or other system contingency). Due to the communication overhead needed to support and enable this process, demand–response is a good example of a smart grid technology where the increased connectivity provides benefits to system operators, while simultaneously introducing cyber risks because it increases the number of cyber attack vectors.[20]

The smartness of the grid is measured on a scale of 0 to 1, as the fraction of nodes that are enabled with demand–response technology, 1 meaning that every node is linked to the control center. The benefits and risks of increased smartness are then calculated using an economic dispatch model of the power market, which computes the cheapest way to generate enough electricity to meet the daily demand, subject to system constraints, based on current electricity prices, and customer demand data for the region.[21] Specifically, the benefits due to reduction in peak demand and quicker reaction to unexpected system contingencies are computed by adapting analytical techniques from Kwag and Kim,[22] who demonstrated a theoretical model incorporating demand–response into the economic dispatch algorithm to find the optimal system operating cost for a utility with quadratic generator cost functions. To model the risks associated with increased smartness, we considered five classes of cyber security failure scenarios related to the incorporation of demand–response technologies, based on those identified in a recent analysis by the National Electric Sector Cybersecurity Organization Resource (NESCOR).[18] The five failure scenarios are listed in Table I, along with a description of how we evaluate their financial impact on the daily operating cost for the utility.

To compute the expected costs to the utility of successful cyber attacks, we combine, for each scenario, its financial impact with its likelihood score as defined by NESCOR. This score, ranging from 0 to 45, represents the ease with which a threat agent could cause that particular failure. To use that metric as a proxy for the probability of success of a specific type of cyber attack attempt, we scale this score to the [0,1] range. Based on physics and economics, we then compute the cyber risk associated with each new connection as the expected increase of risk costs for the five considered failure scenarios. We assume that the connection decisions are made sequentially in decreasing order of marginal benefit/risk–cost ratio. We repeat this analysis for any number of possible connections to obtain the benefit curve, the risk curve, and the resulting tradeoff, as shown in Fig. 8. The optimal level of smartness is the point where the marginal benefit equals the marginal risk cost. In this illustrative case, the optimum is to connect 6 nodes out of 10.

Increased smartness is thus beneficial up to the point where the marginal risk of a new cyber connection—and vulnerability—outweighs its incremental benefits. Again, the model presented here is a generic illustration of the method described earlier. It needs to be adapted to specific cases to support the incremental decisions of a system's managers to increase or not its connectivity, given its internal links and the potential cascading effects of a failure in a global grid.

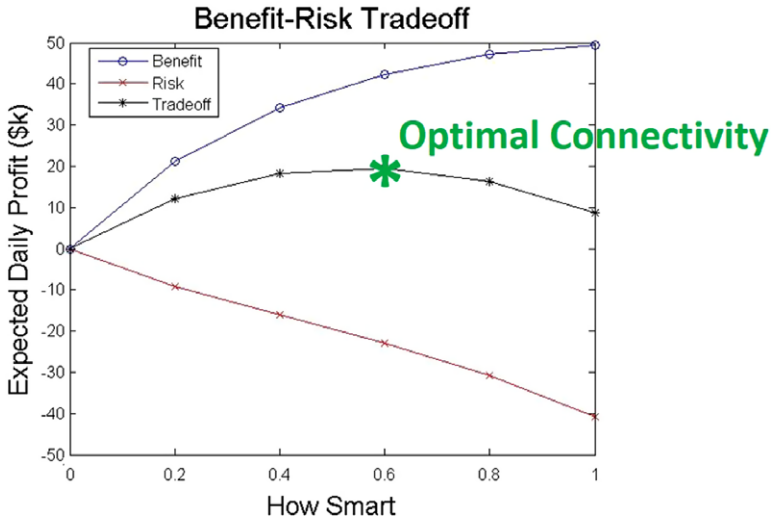


Fig. 8. Benefit–risk tradeoff of increased connectivity for a smarter grid at the margin (how smart is smart enough?).

6. EXAMPLE #3: STAYING AHEAD OF THE ADVERSARIES: DYNAMIC MODELS OF TECHNOLOGY ADOPTION AND RECONFIGURATION

6.1. Thwarting or Delaying a Cyber Attack by Changing the Software

Consider now the case of the manager of a water distribution system who suspects that his operations are threatened by the possibility of a cyber attack and considers the options to update or change his operating software in order to delay or thwart that attack. In this last example, we use a dynamic PRA to model the interaction between such risk management decisions and those of the adversaries. Decisions to adopt, modify, or reconfigure the operating software are treated as a stochastic-control problem based on the link between the system operator's decisions and those of the adversary.[8]

6.2. A Dynamic Model of Software Protection

The age of the software and its influence on cyber decisions are represented by a stochastic-control model. An empirical analysis of software vulnerabilities in the CVE database[13] shows that cyber attacks can be modeled as Poisson processes for common software packages. Additional factors that are critical to decisions of software change or update include the nature of the current software and its known vulnerabilities, and the availability of new software and software patches. These factors are considered here as characteristics of the system's states.

The transitions among these system's states are characterized by a Markov model. We estimate the probability of new vulnerabilities based on data from the National Vulnerability Database (NVD) supported by National Institute of Standards and Technology (NIST).[24] We also assess the software "shelf life" and the occurrence rate of new attacks through vulnerabilities known only to the adversaries ("zero-day vulnerabilities") based on empirical data from Bilge and Dumitras.[25] We estimate the time elapsed between a software release and the publication of the first vulnerabilities based on the pedigree of that software. Clark *et al.*[26] show that an adversary's familiarity with previous "parent" versions of a software can reduce the time that it will take him to discover new vulnerabilities in the defender's system. The delayed availability of patches for new vulnerabilities—the period during which systems cannot be protected—is modeled based on the work of Nappa *et al.*[27] The availability of software patches and discovery of new software vulnerabilities are considered as exogenous state variables, out of the control of the system's operator but parts of the system's states.

The consequences of a cyber attack are addressed using the probabilistic methods presented in the previous sections. The Common Vulnerability Scoring System (CVSS) provides a method to assess the impact and the consequences of different vulnerabilities.[28] The analysis of the CVSS score

[8]In reality, other factors are also key to the success of software management, including its quality in the first place. We do not consider the cases in which a new software may be defective.

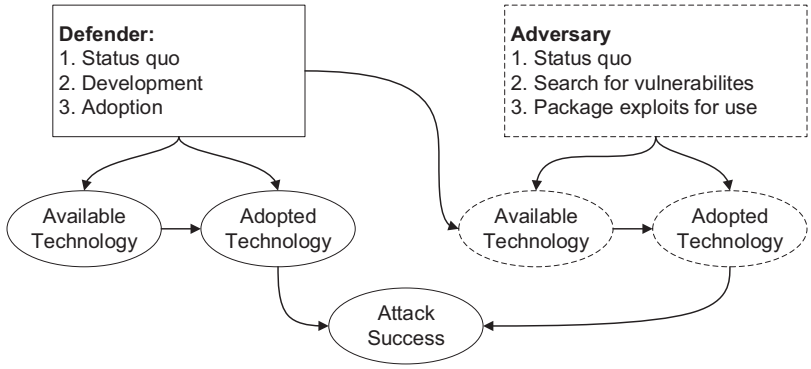


Fig. 9. Simplified model structure of decisions of technology development by a defender, and attack by an adversary through a cyber vulnerability (dashed lines represent uncertainties for the defender).

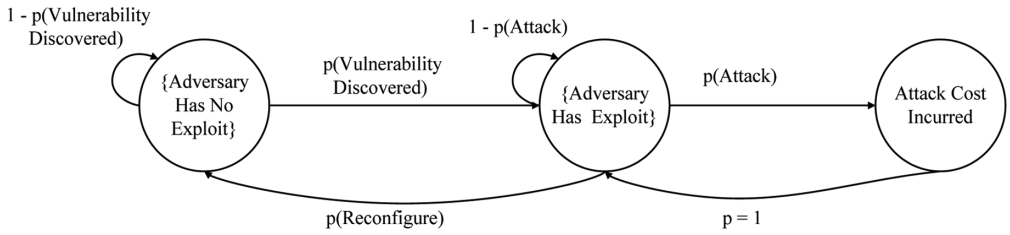


Fig. 10. A simple example of cyber attack Markov model.

for various vulnerabilities in the NVD allows computation of the probability of existence of different vulnerability types and their severity in a given time window.

Reconfiguring software by installing a new version or changing the software settings reduces the probability that adversaries have access to a vulnerability that they know and can exploit. In practice, software reconfiguration can undo the adversaries' work, forcing them to restart their efforts to exploit a system. Waiting too long to update software systems thus increases their vulnerability to an attack, but frequent upgrading can be expensive. Reconfiguration often requires a down time for the physical infrastructure that a software system controls (e.g., electric, gas, water, or traffic). It also implies direct monetary costs for the system's operators if they decide to buy new software licenses and retrain their workforce. We use a decision analysis model to determine when an operator should reconfigure or patch the system, based on the state variables described above, to ensure a desired level of cyber security.

The elements of a simplified cyber attack model involving both an attacker and a defender, and the dependencies between the technology decisions of both sides, are represented in Fig. 9. At each time period, the defenders can decide to maintain their technology's status quo, develop new technology configurations, or adopt new products, but adopting new software can also introduce new vulnerabilities that can be exploited by the attacker. Therefore, both sides may try to improve their technologies, and the problem for the defender is to stay ahead of the attacker.

To illustrate this general method we use a discrete Markov model that captures the evolution of a cyber technology and its management (see Fig. 10). The model involves two basic states (whether or not a known vulnerability is available to the adversary), and a third state in which an attack occurs.

Getting back to the example of a single computer controlling the water distribution system of a city, at each time period, the system's operator has two options: maintain the current system, or reconfigure it to set back an adversary in his attempt to penetrate it. The system's administrator assesses the probability per time unit that the adversary finds a vulnerability (e.g., based on the mean time between two such events) and is able to complete the development of a new attack capability to exploit that vulnerability. The operator then updates the probability that the adversary will use this discovered vulnerability with new capabilities in the same way as one would estimate the reliability of a deteriorating physical system. If the system's administrator (the defender) decides to reconfigure the system, it would be a "fresh start," eliminating any vulnerability previously found by the adversary.

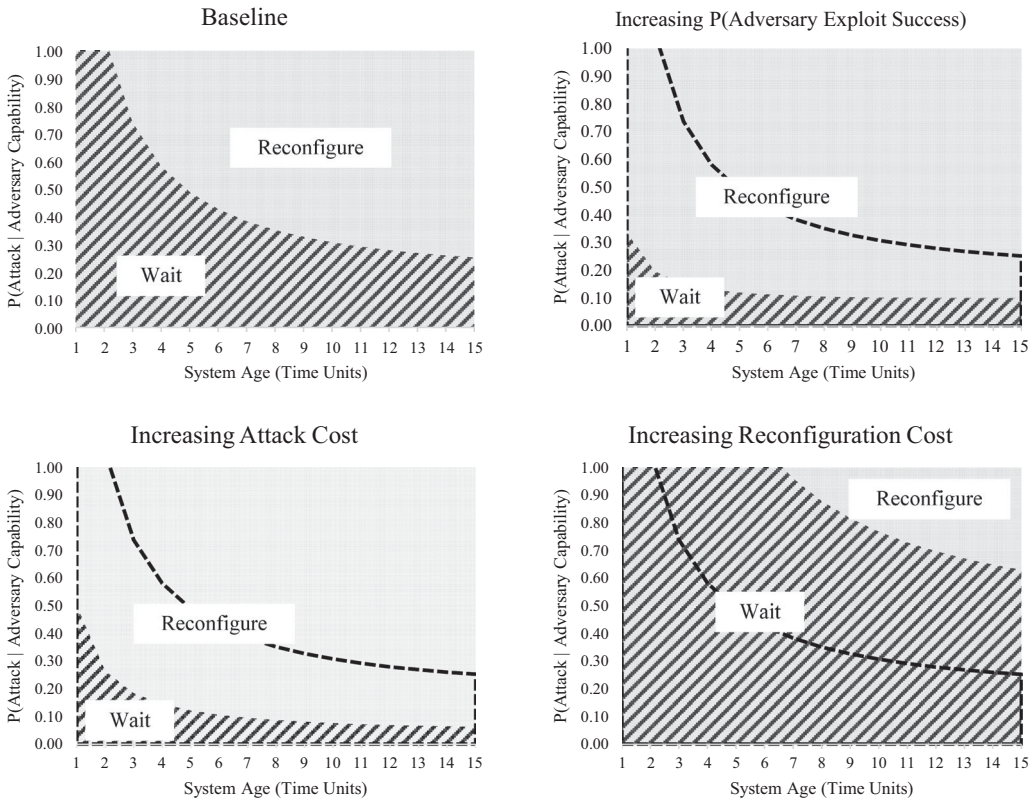


Fig. 11. Evolution of system over time and defender's decisions (the dash line provides a comparison of each decision frontier with the baseline).

To make these decisions, the system's administrator needs to estimate the cost of a successful attack and of reconfiguring the system, including labor and down time. The model is then used to calculate the frontier that divides the two options (adopt a new configuration or wait) in terms of probability of an attack as a function of the system's age. Illustrative results are shown in Fig. 11 based on a notional decision scenario in which we assume the following data. The probability that an adversary develops a new cyber exploit is 0.10 per time unit (e.g., per month), and the probability that he chooses to use it in the next time unit given that it is an option is also 0.10. A successful attack costs the system's operator $1,000 and reconfiguring the software to thwart an attack costs $200. The top left of Fig. 11 shows the baseline results. The remaining three parts provide comparisons between the baseline decision frontier (dashed line) and cases where the input values are changed: increasing the probability of exploit success to 0.60, the attack cost to $4,000, and the reconfiguration cost to $600.

Fig. 11 thus illustrates the dynamic nature of the considered cyber system. As time elapses, one can expect the adversary to gain an advantage. The system's administrator can assess the opponent's strategy based on his assessment of the chances that an adversary will attack once he finds a vulnerability. As time elapses, reconfiguring the system will become preferable to waiting. Institutions that face more technically capable adversaries such as state actors will want to reconfigure sooner (top right of Fig. 11), as will institutions that suffer higher costs if an attacks occurs (bottom left). However, higher reconfiguration costs will delay the optimal timing of a system's reconfiguration (bottom right).

7. CONCLUSION: RISK QUANTIFICATION IN CYBER RISK MANAGEMENT AND THE USE OF PROBABILITY

Large amounts of money are dedicated to cyber risk management, often with little information about the effectiveness of the measures adopted and the

priorities among them. Industries, individuals, and nations see cyber risk as worrisome and, in some cases, terrifying. Insurance companies, when they decide to enter that market, are hard pressed to rationalize premium prices.

We presented here several ways to gather and use the information available to quantify the cyber risk to a potential target, recognizing that the relevant information involves more than past statistics. The three illustrative cases described in this article provided examples of decision support by quantifying the cyber risk and the risk reduction benefits of a spectrum of countermeasures such as encryption, reduction of the number of entry points, or modification of the operation software. The choice is then a function of the risk attitude of the decisionmaker. We assumed here risk neutrality, but other risk attitudes can be introduced through other utility functions and certain equivalents.

Substantial progress in the development of cyber risk analysis has been made in the last few years, from the deep-seated conviction that there was no way to quantify cyber risks to asking how that quantification can be improved. Gathering statistics of incidents and their consequences, then sharing the data in proper forums to protect organizations when needed, is a good place to start. The next task is to analyze and extend these databases to address the risk of attacks that have not happened yet.

In many cases, the first basic step is to understand what needs to be protected, where it is, and what is the structure of the internal information network that could be breached. From the perspective of the defender, that analysis has to include the uncertainties and the dynamics of cyber attacks, considering the evolution of the adversaries' knowledge, means, and intentions.

Our main recommendations to a potential target organization thus include:

- Know your system and decide what needs to be protected in priority.
- Gather data about attacks that have occurred; access and update that information regularly, and extrapolate to consider potential attacks that may not have happened yet but need to be protected against.
- Identify the spectrum of countermeasures that make sense given your structure and your mission.
- Analyze the data and relevant additional information to decide what set of countermeasures should be adopted given budget limits, potential losses, and protection benefits.

In each of the three illustrative cases considered here, we identified an optimum that reflects budgets and tradeoffs. For the organization that gave us access to their attack statistics, it is a choice of basic countermeasures. For smart grids, it is an optimum level of connectivity. For the management of an infrastructure's software, it is a frequency of update or replacement. Again, the details of a cyber risk analysis will depend on the characteristics of specific systems and managing organizations. We showed here that these optimums exist and can be quantified.

ACKNOWLEDGMENTS

M.-Elisabeth Paté-Cornell is the Burt and Deedee McMurtry Professor in the Stanford School of Engineering. This research (including Marshall Kuypers' Research Assistantship) was funded in part by a grant from NASA/JPL and by a Burt and Deedee McMurtry graduate student fellowship. Philip Keller's research was conducted with U.S. federal government support under an NDSEG Fellowship. Matthew Smith is an Operations Research/Systems Analysis Officer in the U.S. Army, thus funded by the DoD. All three co-authors were PhD candidates in the Engineering Risk Research Group of the Stanford Department of Management Science and Engineering at the time when this article was written. The statements in this article do not represent the views of the U.S. government or Stanford. University.

REFERENCES

1. Storm D. Attackers hacked Department of Energy 159 times in 4 years. News Analysis, Computer World, September 14, 2015.
2. Peterson A. OPM says 5.6 million fingerprints stolen in cyberattack, five times as many as previously thought. Washington Post, 2015;
3. North American Electric Reliability Corporation. Cyber Security: BES Cyber System Categorization. Atlanta, GA, 2014.
4. Department of Homeland Security. Recommended Practice: Improving Industrial Control Systems Cybersecurity with Defense-in-Depth Strategies. Washington, DC, 2009.
5. Kuypers MA, Maillart T, Paté-cornell E. An Empirical Analysis of Cyber Security Incidents at a Large Organization. Working Paper, Palo Alto, CA, 2016. Available at: cisac.fsi.stanford.edu.
6. Kuypers MA. Risk in Cyber Systems. PhD thesis, Department of Management Science and Engineering, Stanford University, March 2017.
7. Smith M, Paté-Cornell E. Cyber risk analysis for a smart grid: How smart is smart enough? A multi-armed bandit approach.

Pp. 37–56 in Proceedings of the 2nd Singapore Cyber Security Research Conference, 2017.
8. Parnell GS, Butler RE, Wichmann SJ, Tedeschi M, Merritt D. Air force cyberspace investment analysis. Decision Analysis, 2015; 12(2):81–95.
9. Rao NSV, Poole SW, Ma CYT, He F, Zhuang J, Yau DKY. Defense of cyber infrastructures against cyber-physical attacks using game-theoretic models. Risk Analysis, 2015; 1–17.
10. Department of Defense Chief Information Officer. Department of Defense Instruction: Cyber Security 8500.01. Washington, DC, 2014 March.
11. Stoddard M, Bodeau D, Carlson R, Glantz C, Haimes Y, Lian C, et al. Process Control System Security Metrics—State of Practice. I3P Institute for Information Infrastructure Protection Research Report, 2005, Vol. 1.
12. North American Electric Reliability Corporation. Security Guidelines for the Electricity Sector. Atlanta, GA, 2015.
13. The MITRE Corporation. Common Vulnerabilities and Exposures. National Cybersecurity FFRDC, 2017. Available at: https://cve.mitre.org/.
14. US ICS-CERT. Cyber-Attack Against Ukrainian Critical Infrastructure, 2016. Available at: https://ics-cert.us-cert.gov/alerts/IR-ALERT-H-16-056-01.
15. Lee R, Assante MJ, Conway T. Analysis of the Cyber Attack on the Ukrainian Power Grid. SANS Industrial Control Systems. E-ISAC Report, Washington, DC, 2016.
16. Schneider KP, Fuller JC, Tuffner FK, Singh R. Evaluation of Conservation Voltage Reduction (CVR) on a National Level. Pacific Northwest National Laboratory, 2010; July 2010:114.
17. Morgan M, Apt J, Lave L, Ilic M, Sirbu M, Peha J. The many meanings of "Smart Grid." Department of Engineering and Public Policy, Research Showcase at Carnegie Mellon University, 2009. Available at: http://repository.cmu.edu/cgi/viewcontent.cgi?article=1026&context=epp, Accessed October 29, 2014.
18. Lee A. Electric Sector Failure Scenarios and Impact Analyses. National Electric Sector Cybersecurity Organization Resource (NESCOR) Technical Working Group, Vol. 1, 2013.
19. Larrosa C, Kaneshiro L, Zhao J, Elisabeth Pate-Cornell M. Effect of severe space weather on cascading power grid failure: An illustrative model and policy implications. 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012, PSAM11 ESREL 2012, 2012; 7:5253–5262.
20. Deng R, Yang Z, Chow M-Y, Chen J. A survey on demand response in smart grids: Mathematical models and approaches. IEEE Transactions on Industrial Informatics, 2015; 11(3): 570–582.
21. U.S. Energy Information Administration. Electricity Detailed Survey Data Files, 2015. Available at: https://www.eia.gov/electricity/data/eia861/.
22. Kwag HG, Kim JO. Optimal combined scheduling of generation and demand response with demand resource constraints. Applied Energy, Elsevier Ltd, 2012; 96:161–170.
23. Ponemon Institute. Cost of Data Breach Study: Global Analysis. Ponemon Institute Research Report, 2016.
24. U.S. National Institute of Standards and Technology. National Vulnerability Database. Available at: https://nvd.nist.gov/home.cfm, Accessed April 3, 2017.
25. Bilge L, Dumitras T. Investigating zero-day attacks. Login, 2013; 38(4):6–13.
26. Clark S, Frei S, Blaze M, Smith J. Familiarity breeds contempt: The honeymoon effect and the role of legacy code in zero-day vulnerabilities. Pp. 251–260 in Proceedings of the Annual Computer Security Applications Conference (ACSAC), Orlando, FL, 2010.
27. Nappa A, Johnson R, Bilge L, Caballero J, Dumitraş T. The attack of the clones: A study of the impact of shared code on vulnerability patching. Proceedings of the IEEE Symposium on Security and Privacy, 2015; 2015 July:692–708.
28. Mell P, Scarfone K, Romanosky S. A complete guide to the Common Vulnerability Scoring System Version 2.0. Pp. 1–23 in FIRSTForum of Incident Response and Security Teams, 2007.
29. Cavusoglu H, Mishra B, Raghunathan S. The effect of Internet security breach announcements on market value: Capital market reactions for breached firms and Internet security developers. International Journal of Electronic Commerce, 2004; 9(1):69–104.