In Quest of Benchmarking Security Risks to Cyber-Physical Systems

Saurabh Amin, Massachusetts Institute of Technology
Galina A. Schwartz, University of California at Berkeley
Alefiya Hussain, University of Southern California

Abstract

We present a generic yet practical framework for assessing security risks to cyber-physical systems (CPSs). Our framework can be used to benchmark security risks when information is less than perfect, and interdependencies of physical and computational components may result in correlated failures. Such environments are prone to externalities, and can cause huge societal losses. We focus on the risks that arise from interdependent reliability failures (faults) and security failures (attacks). We advocate that a sound assessment of these risks requires explicit modeling of the effects of both technology-based defenses and institutions necessary for supporting them. Thus, we consider technology-based security defenses grounded in information security tools and fault-tolerant control in conjunction with institutional structures. Our game-theoretic approach to estimating security risks facilitates more effective defenses, especially against correlated failures.

Survivability of critical infrastructures in the presence of security attacks and random faults is of national importance. These infrastructures are spatially distributed across large physical areas, and consist of heterogeneous cyber-physical components interconnected by communication networks with complex peering and hierarchies. Networked control systems (NCSs) and supervisory control and data acquisition (SCADA) systems are widely used to monitor, control, and remotely manage infrastructures over private or shared communication networks. Such cyber-physical systems (CPSs) permit synergistic interactions between physical dynamics and computational processes. Wide deployment of information and communication technologies (ICT) in CPSs results in higher reliability and lower operational costs relative to the traditional proprietary and closed systems. However, as recent incidents indicate, today's CPSs face new security threats driven by their exposure to ICT insecurities.

Security Threats

To develop a classification of security threats to CPSs, we first outline how the operator(s) of modern CPSs typically approach the monitoring, control, and management of infrastructures. As shown in Fig. 1, they use a layered architecture consisting of *regulatory control* (layer 1), *supervisory control* (layer 1), and a *management level* (layer 3). This architecture enables robust composition of multilevel controllers, and permits CPS operators to use *defenses* to limit the effects of failures caused by *faults* and/or *attacks*.

The regulatory control layer directly interacts with the underlying physical infrastructure dynamics through a network of sensors and actuators. These field devices are connected to programmable logic controllers (PLCs) or remote terminal units (RTUs), and implement detection and regulation mechanisms that are primarily reactive in nature. These mechanisms can also respond to localized failures of field devices and communication links. The regulatory controllers (or PLCs) interact with the supervisory controllers via a control network.

At the supervisory control layer, model-based diagnostic tools are combined with optimal control-based tools to ensure on-time response to distributed failures. The supervisory workstations are used for data logging, diagnostic functions such as fault diagnosis, and supervisory control computations such as set-point control and controller reconfigurations.

Lastly, the management (topmost) layer focuses on strategies that maximize the operator's profit while minimizing its losses due to security and reliability failures. The CPS operator and other authorized remote users can access information about the CPS processes and send specifications to the controllers at lower layers via the Internet or a corporate network.

Security threats to hierarchically managed CPSs arise from four channels. First, CPSs inherit vulnerabilities from embedded commercial off-the-shelf ICT devices, and are subject to correlated software bugs and hardware malfunctions. Second, the proprietary protocols and closed networks are being replaced with standard open Internet protocols and shared networks. Malicious attackers capable of exploiting protocol and network insecurities can target CPS operations. Third, numerous parties generate, use, and modify CPS data. This poses new challenges in access control and authorization among the strategic players such as the operators, SCADA and ICT vendors, and end users of the system. Fourth, CPSs employ a large number of remote field devices that can be accessed via short-range communications. Thus, CPSs are vulnerable to adversarial manipulation, both remote and local.

Adversaries can exploit the aforementioned threat channels via denial-of-service (DoS) and deception attacks, which result in losses of availability and integrity of sensor-control data,

0890-8044/13/$25.00 © 2013 IEEE

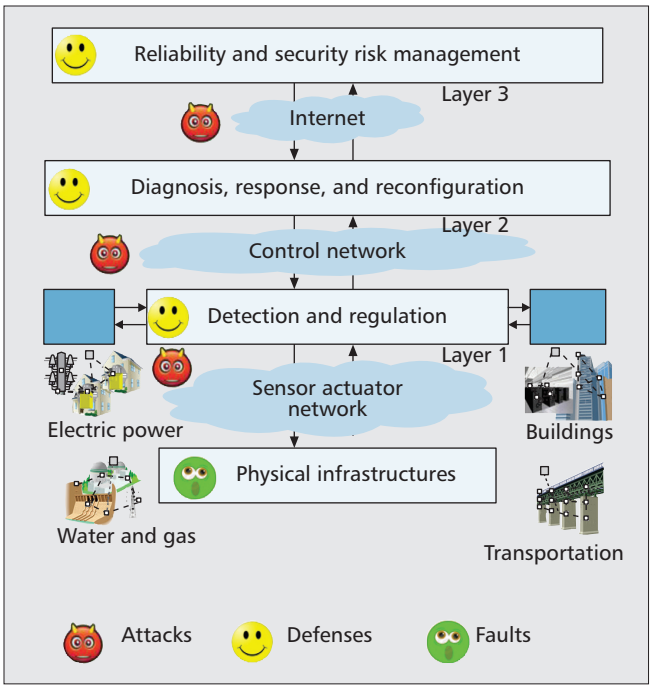

Figure 1. *A layered architecture for management of CPS.*

respectively. In Table 1, we present examples of security attacks on the regulatory and supervisory control layers. Attacks at the management level are similar to attacks on computer networks. We refer the reader to [1, 2] for specific discussions on security attacks to smart grid infrastructures.

Classification of Correlated Failures

The danger of correlated failures becomes especially profound in CPSs due to the tight coupling of typically continuous physical dynamics and discrete dynamics of embedded computing processes. Correlated failures originate from one or more of the following events:

- *Simultaneous attacks*: Targeted cyber attacks (e.g., failures due to Stuxnet); non-targeted cyber attacks (e.g., failures due to Slammer worm, distributed DoS attacks [3], congestion in shared networks); coordinated physical attacks (e.g., failures caused by terrorists)
- *Simultaneous faults*: Common-mode failures (e.g., failure of multiple ICT components in an identical manner [4], programming errors); random failures (e.g., natural events such as earthquakes and tropical cyclones, and operator errors such as an incorrect firmware upgrade)
- *Cascading failures*: Failure of a fraction of nodes (components) in one CPS subnetwork can lead to progressive escalation of failures in other subnetworks (e.g., power network blackouts affecting communication networks, and vice versa) [5].

The above classification is neither fully disjoint nor exhaustive. Still, we envision that it will be useful for CPS risk assessment. We term correlated failures caused by simultaneous attacks as security failures and simultaneous faults as reliability failures. Due to the tight cyber-physical interactions, it is extremely difficult (and often prohibitively time-consuming) to isolate the cause of any specific failure using the diagnostic information, which, in general, is imperfect and incomplete. Thus, reliability and security failures in CPSs are inherently intertwined. We believe that the quest to find a mutually exclusive and jointly exhaustive partition of failure events must be abandoned. Instead, the research emphasis should shift to the analysis of *interdependent reliability and security failures*, and risk assessment.

Information and CPS Risks

The Interplay of Technological Defenses and Institutions

There are two types of technological means to reduce CPS risks: ICT security tools and control-theoretic tools. The *ICT security tools* include authentication and access control mechanisms, network intrusion detection systems, patch management, and security certification. In practice, the effectiveness of these security tools is limited by CPS reliability and cost considerations. For example, the frequency of security patch updates is limited by the real-time constraints on the availability of CPS data; common criteria certification is limited by the resources for CPS security and so on. The *control-theoretic tools* include model-based attack/fault detection and isolation, robust control strategies that maintain closed-loop stability and performance guarantees under a class of DoS/deception attacks, and reconfigurable (switching) control strategies to limit the effect of correlated failures. Recently, several organizations (e.g., NIST, NERC, DHS) have proposed security standards and recommendations that combine the ICT-specific security defenses with control theoretic tools.

While technology-based defenses for CPS are the main channel to improve their survivability against correlated failures, the mere existence of these defenses is not sufficient. It is well established that the lack of private parties' incentives for security improvements is a severe impediment to achieving socially desirable improvements of CPS security [6]. Indeed, large-scale critical infrastructures are typically managed by profit-driven private entities. Proper implementation of technological defenses and resilient operation requires compliance of relevant entities. Below we highlight the informational deficiencies that negatively affect the incentives for security.

Informational Deficiencies

Due to the prohibitively high costs of information acquisition, it is often too costly to determine the following:

- Which hardware malfunctions and software bugs have caused a system failure
- Whether the system failure was caused by a reliability failure or security failure or both

In many cases, this information varies significantly across different entities (players), such as CPS operators, SCADA and ICT vendors, network service providers, users, and local/ federal regulatory agencies (or government). Information deficiencies arise from the conflicting interests of individual players whose choices affect the CPS risks. One may say that interdependent failures cause *externalities* that result in misaligned player incentives (i.e., the individually optimal CPS security defenses diverge from the socially optimal ones).

Moreover, in environments with incomplete and also asymmetric (and private) information, the societal costs of a correlated CPS failure typically exceed the losses of the individual players whose products and services affect CPS operations, and on whose actions the CPS risks depend. Specifically, interdependencies between security and reliability failures in CPS are likely to cause negative externalities. In such environments, the individual players tend to underinvest in security relative to a socially optimal benchmark. This requires design of institutional means to realign the individual players' incentives to make adequate investments in security. Examples of institutional means include *regulations* that require players to certify that they possess certain security capabilities, and *legal rules* which mandate that players share information about security incidents with government agencies and/or the public through established channels.

	Control layer	
	Regulatory control	**Supervisory control**
Deception attacks	Spoofing, replay	Set-point change
	Measurement substitution	Controller substitution
DoS attacks	Physical jamming	Network flooding
	Increase in latency	Operational disruption

Table 1. *Cyber-attacks to CPS control layers.*

Clearly, these individual players cannot completely eliminate the risk of CPS failures even in the presence of advanced technological defenses and institutional measures, which aim to reduce (or even eliminate) incentive misalignment between individual and socially optimal security choices. For example, consider a benchmark case when security defenses are optimally chosen by the social planner for a given technological and institutional environment. There still remains a residual risk driven by fundamental physical limits. Indeed, when security defenses are chosen by individual players, the risk is only higher. Thus, non-negligible (public) residual risks are characteristic for CPSs that are subjected to correlated failures.

So far, the occurrence of extreme correlated failures have been statistically rare. However, with the emergence of organized cyber-crime groups capable of conducting intrusions into NCS/SCADA systems, the risks of such rare failure events cannot be ignored. Unsurprisingly, cyber-warfare is projected to become the future of armed conflict, and managing CPS risks must be at the core of any proactive defense program.

Benchmarking CPS Risks

Due to the aforementioned challenges, benchmarking CPS risks is a hard problem, and several questions remain unanswered [7–9]. Our goal in this article is twofold:

- We suggest a game-theoretic framework that assesses security risks by quantifying the misalignment between individually and socially optimal security investment decisions when the CPS comprises interdependent NCS.
- We advocate that better information about these risks is a prerequisite to improvement of CPS security via a combination of more sophisticated technology-based defenses and the advancement of their supporting institutions.

Improved assessment of the CPS risks will lead to several beneficial developments, such as improved risk management at both the individual and societal levels. Thus, a standardized framework should be established that can assess and compare different technological and institutional means for risk management. At the very least, better knowledge of CPS risks will permit the players to make more informed (and therefore better and cheaper) choices of security defenses, thus improving the societal welfare.

Framework to Benchmark CPS Risks

We now present a risk assessment framework from the perspective of CPS operators. Our setup can readily be adapted to assess risks from the perspective of other players.

CPS with a Centralized Control System

Consider a CPS with m independent components managed by a single operator (i.e., centralized control system). For the ith component, let Ω^i denote the set of all hardware flaws, software bugs, and vulnerability points that can be compromised during any reliability and/or security failure event. The failure events form a collection of subsets of Ω^i, which we denote by $\mathcal{F}^i$. Let the random variables $X_R^i: \Omega^i \to \mathbb{R}$ and $X_S^i: \Omega^i \to \mathbb{R}$ represent the reliability and security levels of the i-th component, respectively, with joint (cumulative) distribution function:

$$F^i_{X_R^i, X_S^i}(x_R^i, x_S^i) = P\{\omega \in \Omega^i \mid X_R^i(\omega) \le x_R^i, X_S^i(\omega) \le x_S^i\},$$

where the measure P assigns probabilities to failure events. Notice that the reliability level X_R^i and security level X_S^i are defined on the same measure space $(\Omega^i, \mathcal{F}^i)$, and they are not mutually independent, that is,

$$F_{X_R, X_S}(x_R, x_S) \uparrow F_{X_R}(x_R) \cdot F_{X_S}(x_S).$$

Unfortunately, the CPS operator does not have perfect knowledge of these distributions. Reasonable estimates of $F_{X_R}(x_R)$ may be obtained from historical failure data. However, estimating the joint distribution $F_{X_R, X_S}(x_R, x_S)$ is difficult as attackers continue to find new ways to compromise security vulnerabilities.

In general, the random vector (X_R^i, X_S^i) is influenced by:

- Action set of the CPS operator $\mathcal{A} = \mathcal{U} \cup \mathcal{V}$, where $\mathcal{U} := \{\mathcal{U}^1, \ldots \mathcal{U}^m\}$ and $\mathcal{V} := \{\mathcal{V}^1, \ldots, \mathcal{V}^m\}$ denote the set of control and security choices, respectively
- Action set of other players $\mathcal{B}$, such as vendors, attackers, service providers, users, and regulatory agencies
- Environment $\mathcal{E}$, including the technological, organizational, and institutional factors

For given reliability and security levels x_R^i, x_S^i, let the function $L^i(x_R^i, x_S^i)$ denote the losses faced by the CPS operator when the ith component fails (e.g., the cost of service disruptions, maintenance/recovery costs, and penalties for users' suffering). Then, for CPS with m *independent* components, the aggregate risk can be expressed as:[1]

$$\mathcal{R} = \sum_{i=1}^{m} \mathcal{R}^i\left(L^i\left(X_R^i, X_S^i\right)\right), \tag{1}$$

where the functional $\mathcal{R}^i$ assigns a numerical value to each random variable L^i with distribution function $F^i_{L^i}$. Henceforth, we use the expected (mean) value of loss, $\mu(L^i) = E[L^i(X_R^i, X_S^i)]$, as a metric of $\mathcal{R}^i$, but caution that it is inadequate to capture risk of extreme failure events.[2] From Eq. 1, we observe that the aggregate risk is also influenced by actions $\mathcal{A}$, $\mathcal{B}$, and environment $\mathcal{E}$. To emphasize this dependence, we will use $\mathcal{R}(\mathcal{A}, \mathcal{B}, \mathcal{E})$ to denote the aggregate CPS risk.

For a given environment $\mathcal{E}$ and fixed choices $\mathcal{B}$ of other players, the CPS operator's objective is to choose security actions $\mathcal{V}$ and control actions $\mathcal{U}$ to minimize the total expected cost $\mathcal{J}(\mathcal{U}, \mathcal{V})$ of operating the system:

$$\mathcal{J}(\mathcal{U}, \mathcal{V}) = \mathcal{J}_I(\mathcal{V}) + \mathcal{J}_{II}(\mathcal{U}, \mathcal{V}), \tag{2}$$

where $\mathcal{J}_I(\mathcal{V}) := \Sigma_{i=1}^m \ell^i(\mathcal{V}^i)$ denotes the operator's cost of employing security choices $\mathcal{V}$, and $\mathcal{J}_{II}(\mathcal{U}, \mathcal{V})$ is the expected

[1] *The assumption of independent components can easily be relaxed to include parallel, series, and interlinked components.*

[2] *Other commonly used choices of risk $\mathcal{R}^i$ include the mean-variance model: $\mu(L^i) + \lambda^i\sigma(L^i)$, where $\lambda^i > 0$ and $\sigma(L^i)$ is the standard deviation of L^i; and the value-of-risk model: $VaR_{\alpha^i}(L^i) = min\{z \mid F^i_{L^i}(z) \ge \alpha^i\}$, which is the same as α^i-quantile in distribution of L^i.*

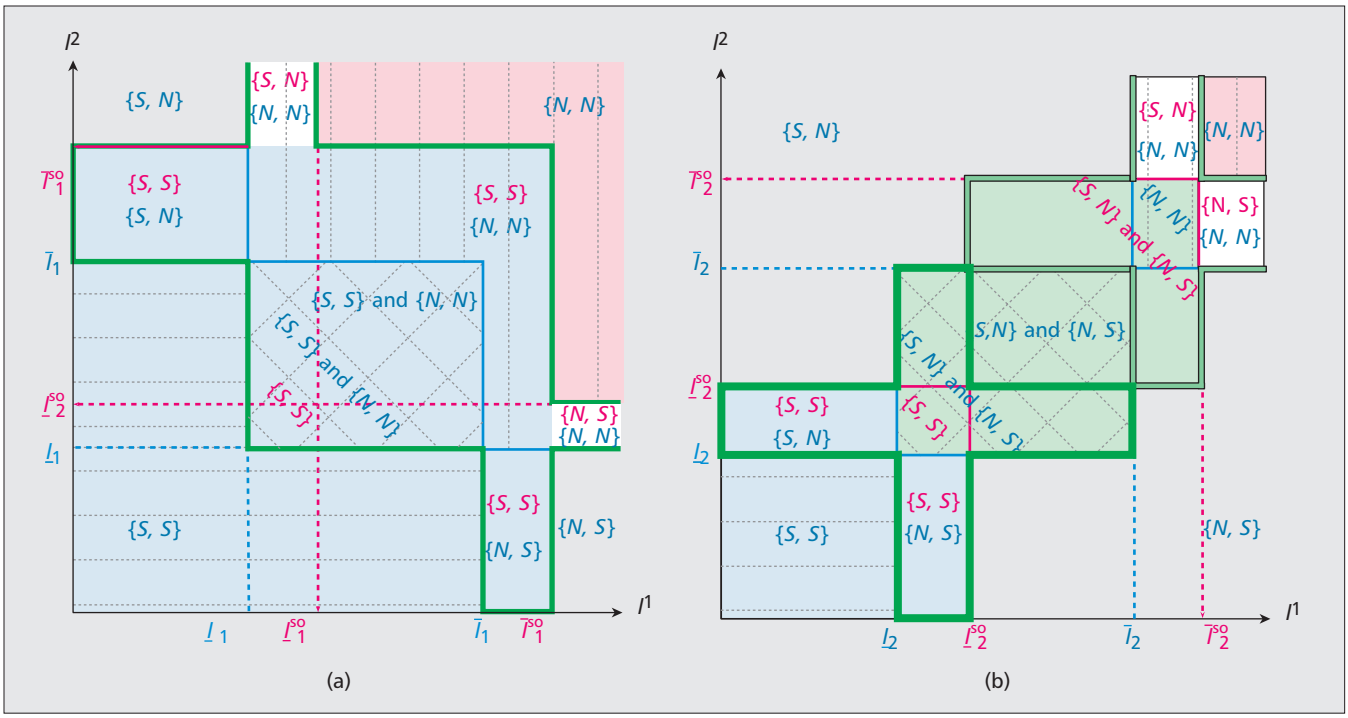

Figure 2. *Individual optima (Nash equilibria) and social optima.*

operational cost. From Eq. 2, when the CPS operator's security choices are $\mathcal{V}$, s/he chooses control actions $\mathcal{U} = \mu^*(\mathcal{V})$ to minimize total expected cost, where $\mu^*(\mathcal{V})$ is an optimal control policy. Let the CPS operator's minimum cost for the case when security choices are $\mathcal{V}$ and $\{\varnothing\}$ (i.e., no security defenses) be defined as $\bar{\mathcal{J}}(\mathcal{V}) := \mathcal{J}(\mu^*(\mathcal{V}),\mathcal{V})$ and $\mathcal{J}^0 := \mathcal{J}(\mu^*(\{\varnothing\}), \{\varnothing\})$, respectively. To evaluate the effectiveness of $\mathcal{V}$, we use the difference of corresponding expected costs:

$$\Delta(\mathcal{V}) := \mathcal{J}^0 - \bar{\mathcal{J}}(\mathcal{V}). \tag{3}$$

Thus, $\Delta(\mathcal{V})$ denotes the CPS operator's gain from employing security choices $\mathcal{V}$. It can be viewed as the reduction of operator's risk when s/he chooses $\mathcal{V}$ over no defenses, that is,

$$\mathcal{R}(\mathcal{A}^0, \mathcal{B}, \mathcal{E}) - \mathcal{R}(\mathcal{A}(\mathcal{V}), \mathcal{B}, \mathcal{E}) = \Delta(\mathcal{V}), \tag{4}$$

where $\mathcal{A}(\mathcal{V})$ and $\mathcal{A}^0$ denote the action set corresponding to security choices $\mathcal{V}$ and $\{\varnothing\}$, respectively. The problem of choosing optimal security choices $\mathcal{V}^*$ can now be viewed as an optimization problem over the set of security defenses:

$$\max_{\mathcal{V}} \Delta(\mathcal{V}), \text{ subject to the constraint } \mathcal{J}(\mathcal{V}) \le K,$$

where K is the available budget for security investments.

The *residual risk* after the implementation of optimal security choices $\mathcal{V}^*$ can be obtained as $\mathcal{R}(\mathcal{A}^0, \mathcal{B}, \mathcal{E}) - \Delta(\mathcal{V}^*)$. Risks from failure events (those resulting from security attacks, random faults, cascading failures, etc.) can thus be estimated and compared, and the best security defenses $\mathcal{V}$ corresponding to anticipated failure types can be selected by the CPS operator.

The above analysis assumes that the choices $\mathcal{B}$ of other players do not change in response to the CPS operator's choices $\mathcal{A}$. When players are strategic, the optimal security choices must be computed as *best responses* to the other players' (Nash) strategies. Finally, government or regulatory agencies can also influence the environment $\mathcal{E}$.

CPS with Interdependent Networked Control Systems

Let us focus on the issue of misalignment between individual and socially optimal actions in the case when a CPS comprises multiple NCSs communicating over a shared network. In contrast to the above, we now assume that each NCS is managed by a separate operator. The NCS operators choose their security levels to safeguard against network-induced risks (e.g., due to distributed DoS attacks). Each NCS is modeled by a discrete-time stochastic linear system, which is controlled over a lossy communication network:

$$\begin{aligned} x_{t+1}^i &= Ax_t^i + v_t^i Bu_t^i + w_t^i \\ y_t^i &= \gamma_t^i Cx_t^i + v_t^i \end{aligned} \qquad t \in \mathbb{N}_0, \quad i \in M, \tag{5}$$

where M denotes the number of players, $x_t^i \in \mathbb{R}^d$ the state, $u_t^i \in \mathbb{R}^m$ the input, $w_t^i \in \mathbb{R}^d$ the process noise, $y_t^i \in \mathbb{R}^p$ the measured output, and $v_t^i \in \mathbb{R}^p$ the measurement noise, for player **P***i* at the tth time step. Let the standard assumptions of linear quadratic Gaussian (LQG) theory hold. The random variables γ_t^i (resp. v_t^i) are i.i.d. Bernoulli with the failure probability $\tilde{\gamma}^i$ (resp. $\tilde{v}^i$), and model a lossy sensor (resp. control) channel.

We formulate the problem of security choices of the individual players as a non-cooperative two-stage game [10]. In the first stage, each **P***i* chooses to make a security investment ($\mathcal{S}$) or not ($\mathcal{N}$). The set of player security choices is denoted $\mathcal{V} := \{\mathcal{V}^1, \ldots, \mathcal{V}^m\}$, where $\mathcal{V}^i = \mathcal{S}$ if **P***i* invests in security and $\mathcal{N}$ if not. Once player security choices are made, they are irreversible and observable by all the players. In the second stage, each **P***i* chooses a control input sequence $\mathcal{U}^i := \{u_t^i, t \in \mathbb{N}_0\}$ to maintain optimal closed-loop performance. The objective of each **P***i* is to minimize his/her total cost:

$$\mathcal{J}^i(\mathcal{V}, \mathcal{U}) = \mathcal{J}_{\mathrm{I}}^i(\mathcal{V}) + \mathcal{J}_{\mathrm{II}}^i(\mathcal{V}, \mathcal{U}), i \in M, \tag{6}$$

where the first stage cost is denoted $\mathcal{J}_{\mathrm{I}}^i(\mathcal{V}) := (1 - \mathcal{I}^i)\ell^i$, and $\mathcal{J}_{\mathrm{II}}^i(\mathcal{V}, \mathcal{U})$ denotes second stage cost (the average LQG cost). Here $\ell^i > 0$ is the security investment incurred by **P***i* only if

s/he has chosen $\mathcal{S}$, and the indicator function $\mathcal{I}^i = 0$ when $\mathcal{V}^i = \mathcal{S}$, and $\mathcal{I}^i = 1$ otherwise.

In order to reflect security interdependencies, in our model, the failure probabilities $\tilde{\gamma}^i$ and $\tilde{\nu}^i$ depend on the **P***i*'s own security choice $\mathcal{V}^i$ *and* on the other players' security choices $\{\mathcal{V}^j, j \uparrow i\}$. Following [10], we assume

$$\mathrm{P}[\gamma_t^i = 0 \mid \mathcal{V}] = \tilde{\gamma}^i(\mathcal{V}) := \mathcal{I}^i\bar{\gamma} + (1 - \mathcal{I}^i\underline{\gamma})\alpha(\eta^{-i}).$$

In Eq. 7, the first term reflects the probability of a *direct failure*, and the second term reflects the probability of an *indirect failure*. The interdependence term $\alpha(\eta^{-i})$ increases as the number of players, excluding **P***i*, who have chosen $\mathcal{N}$ increase, where $\eta^{-i} := \sum_{j\uparrow i}\mathcal{I}^j$; similarly for ν_t^i. The social planner objective is to minimize the aggregate cost:

$$\mathcal{J}^{\mathrm{SO}}(\mathcal{V},\mathcal{U}) = \sum_{i=1}^{\mathrm{m}} \mathcal{J}^i(\mathcal{V},\mathcal{U}). \tag{8}$$

Consider a two-player game, where the interdependent failure probabilities are given by Eq. 8. To derive optimal player actions (security choices $\mathcal{V}^i$), we distinguish the following two cases: *increasing incentives* and *decreasing incentives*. For the case of increasing incentives, if a player secures, other player's gain from securing increases, that is, $\mathcal{J}_{\mathrm{II}}^*(\{\mathcal{N}, \mathcal{N}\}) - \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{S}, \mathcal{N}\}) \le \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{N}, \mathcal{S}\}) - \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{S}, \mathcal{S}\})$, where $\mathcal{J}_{\mathrm{II}}^*(\cdot)$ denotes the optimal second stage cost. Similarly, for the case of decreasing incentives, a player's gain from investing in security decreases when the other player invests in security, that is, $\mathcal{J}_{\mathrm{II}}^*(\{\mathcal{N}, \mathcal{N}\}) - \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{S}, \mathcal{N}\}) \ge \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{N}, \mathcal{S}\}) - \mathcal{J}_{\mathrm{II}}^*(\{\mathcal{S}, \mathcal{S}\})$.

Figure 2a (resp. Fig. 2b) characterizes the Nash equilibria (individually optimal choices) and socially optimal choices of the game for the case of increasing (resp. decreasing) incentives, where we assume $\underline{\ell}_1^{\mathrm{SO}} < \bar{\ell}_1$ (resp. $\bar{\ell}_2 > \underline{\ell}_2^{\mathrm{SO}}$). For $i \in \{1, 2\}$, the thresholds $\underline{\ell}_i$, $\bar{\ell}_i$, $\underline{\ell}_i^{\mathrm{SO}}$, and $\bar{\ell}_i^{\mathrm{SO}}$ are given in [10].

Consider the case of increasing incentives (Fig. 2a). If $\ell^i < \underline{\ell}_1$ (resp. $\ell^i > \bar{\ell}_1$), the symmetric Nash equilibrium $\{\mathcal{S}, \mathcal{S}\}$ (resp. $\{\mathcal{N}, \mathcal{N}\}$) is unique. Thus, $\underline{\ell}_1$ (resp. $\bar{\ell}_1$) is the cutoff cost below (resp. above) which both players invest (resp. neither player invests) in security. If $\underline{\ell}_1 \le \ell^i \le \bar{\ell}_1$, both $\{\mathcal{S}, \mathcal{S}\}$ and $\{\mathcal{N}, \mathcal{N}\}$ are individually optimal. However, if $\ell^1 < \underline{\ell}_1$ & $\ell^2 > \bar{\ell}_1$ (resp. $\ell^1 > \underline{\ell}_1$ & $\ell^2 < \bar{\ell}_1$), the asymmetric strategy $\{\mathcal{S}, \mathcal{N}\}$ (resp. $\{\mathcal{N}, \mathcal{S}\}$) is an equilibrium. Now, if $\ell^i < \bar{\ell}_1^{\mathrm{SO}}$ (resp. $\ell^i > \bar{\ell}_1^{\mathrm{SO}}$), the socially optimal choices are $\{\mathcal{S}, \mathcal{S}\}$ (resp. $\{\mathcal{N}, \mathcal{N}\}$). If $\ell^1 \le \underline{\ell}_1^{\mathrm{SO}}$ & $\ell^2 \le \bar{\ell}^{\mathrm{SO}}$ (resp. $\ell^2 \le \underline{\ell}_1^{\mathrm{SO}}$ & $\ell^1 \le \bar{\ell}_1^{\mathrm{SO}}$), socially optimal choices are $\{\mathcal{S}, \mathcal{N}\}$ (resp. $\{\mathcal{N}, \mathcal{S}\}$). Similarly, we can describe individually and socially optimal choices for the case of decreasing incentives (Fig. 2b).

For both cases, we observe that the presence of interdependent security causes a negative externality. The individual players are subject to network-induced risks and tend to under-invest in security relative to the social optimum. From our results, for a wide parameter range, regulatory impositions to incentivize higher security investments are desirable (discussed later). The effectiveness of such impositions on the respective risks faced by individual players (NCS operators) can be evaluated in a manner similar to Eqs. 3–4.

Challenges in CPS Risk Assessment

Technological Challenges

A significant challenge for the practical implementation of our CPS risk assessment framework is to develop data-driven, stochastic CPS models, which account for dynamics of CPS with interdependent reliability and security failures. Each of these singular/basic models should account for CPS dynamics and focus on a specific failure scenario. The basic models can be composed into a *composite* model to represent various correlated failure scenarios, including simultaneous attacks, common-mode failures, and cascading failures. By using of quantitative techniques from statistical estimation, model-based diagnosis, stochastic simulation, and predictive control, we can automatically generate new failure scenarios from real-time sensor-control data. These techniques enable the synthesis of operational security strategies and provide estimates of residual risks in environments with highly correlated failures and less than perfect information. Thus, theoretical guarantees and computational tools are needed for the following:

- Compositions of stochastic fault and attack models
- Inference and learning of new failure scenarios
- Fast and accurate simulation of CPS dynamics
- Detection and identification of failure events
- Operational ICT and control based strategies

The DETERLab testbed [11] provides the capability to conduct experiments with a diverse set of CPS failure scenarios, where the controllable variables range from IP-level dynamics to introduction of malicious entities such as distributed DoS attacks. The cyber-physical aspects of large-scale infrastructures can be integrated together on DETERLab to provide an experimental environment for assessing CPS risks. Specifically, the DETERLab provides a programmable network emulation environment, and a suite of tools that allow a user to describe the experimentation "apparatus," and monitor and control the experimentation "procedure." Multiple experimentations can be executed at the same time by different users if computational resources are available.

The main challenge for CPS experimentation on the DETERLab testbed is to compose physical system dynamics (real/simulated/emulated) with communication system emulation. The experimentation "apparatus" should model the communication network, the physical network, and their dynamic interactions. The experimentation "procedure" should describe the sensing and actuation policies that are the best responses to strategic actions of other players.

Institutional Challenges

The design of institutional means is a chicken-and-egg problem. On one hand, institutional means such as imposition of legal liability, mandatory incident disclosure, and insurance instruments improve the information about CPS risks. On the other hand, substantial knowledge of CPS risks is required for their design and successful deployment.

Given the limitations of currently available risk assessment tools, the CPS operators find it hard (and, as a result, costly) to manage their risks. This problem is especially acute for risk management via financial means, such as diversification, reallocation to other parties, and insurance. For example, insurance instruments of CPS risks management are meager: the premiums of cyber-security contracts are not conditioned on the security parameters. It would be no exaggeration to say that so far, the cyber-insurance market has failed to develop. For example, the volume of underwritten contracts is essentially unchanged in a decade, despite multiple predictions of its growth by independent researchers and industry analysts. In fact, even the existing superficial "market" is largely sustained by non-market (regulatory) forces.

Indeed, the leading reason for CPS operators to acquire insurance policies at the prevailing exuberant prices is their need to comply with federal requirements for government contractors. Citizens (i.e., federal and state taxpayers) are the final bearers of these costs. We expect that this situation will remain "as is" unless information on CPS risks drastically improves.

Another related problem is that of suboptimal provider incentives (as seen in Fig. 2). A CPS operator's estimates of his/her own risk tend to be understated (relative to societal ones), even when failure probabilities are known to him/her. In such cases, the gap between individually and socially optimal incentives could be reduced via adjustments of legal and regulatory institutions. For example, it would be socially desirable to introduce limited liability (i.e., a due care standard) for individual entities whose products and services are employed in CPSs. This would improve providers' incentives to invest in their products' security and reliability. However, due to information incompleteness, currently there is no liability regime for providers of CPS components and services, for neither security nor reliability driven failures. Indeed, any liability regime is based on knowing (the estimate[s] of) failure probabilities and the induced losses. This again requires benchmarking of CPS risks.

Concluding Remarks

Benchmarking of CPS risks is a hard problem. It is harder than the traditional risk assessment problems for infrastructure reliability or ICT security, which so far have been considered in isolation. Estimation of CPS risks by naively aggregating risks due to reliability and security failures does not capture the externalities, and can lead to grossly suboptimal responses to CPS risks. Such misspecified CPS risks lead to biased security choices and reduce the effectiveness of security defenses.

Modern, and especially upcoming, CPSs are subjected to complex risks, of which very little is known despite the realization of their significance. In this article we are calling on our colleagues to embark on the hard task of assessing interdependent CPS risks. The effectiveness of security defenses can be increased only when our knowledge of CPS risks improves.

Acknowledgments

We are grateful to the anonymous reviewers for their feedback, and thank Professors S. Shankar Sastry (UC Berkeley) and Joseph M. Sussman (MIT) for useful discussions.

References

[1] Y. Mo *et al.*, "Cyber-Physical Security of A Smart Grid Infrastructure," *Proc. IEEE*, vol. 100, no. 1, Jan. 2012, pp. 195–209.

[2] S. Sridhar, A. Hahn, and M. Govindarasu, "Cyber-Physical System Security for the Electric Power Grid," *Proc. IEEE*, vol. 100, no. 1, Jan. 2012, pp. 210–24.

[3] A. Hussain, J. Heidemann, and C. Papadopoulos, "A Framework for Classifying Denial of Service Attacks," *Proc. 2003 ACM Conf. Applications, Technologies, Architectures, and Protocols for Computer Communications*, 2003, pp. 99–110.

[4] S. Amin *et al.*, "Cyber Security of Water SCADA Systems – Part II: Attack Detection Using Enhanced Hydrodynamic Models," *IEEE Trans. Control Systems Technology*, 2012.

[5] S. Buldyrev *et al.*, "Catastrophic Cascade of Failures in Interdependent Networks," *Nature*, vol. 464, no. 7291, Apr. 2010, pp. 1025–28.

[6] C. Hall *et al.*, "Resilience of the Internet Interconnection Ecosystem," *Proc. 10th Wksp. Economics of Information Security*, June 2011.

[7] T. Alpcan and T. Basar, *Network Security: A Decision and Game Theoretic Approach*, Cambridge Univ. Press, 2011.

[8] P. Grossi and H. Kunreuther, *Catastrophe Modeling: A New Approach to Managing Risk*, Springer, 2005, vol. 25.

[9] Y. Y. Haimes, *Risk Modeling, Assessment, and Management*, 3rd ed., Wiley, 2009.

[10] S. Amin, G. A. Schwartz, and S. S. Sastry, "On the Interdependence of Reliability and Security in Networked Control Systems," *CDC-ECE*, IEEE, 2011, pp. 4078–83.

[11] T. Benzel, "The Science of Cyber Security Experimentation: The Deter Project," *Proc. 27th ACM Annual Computer Security Applications Conf.*, 2011, pp. 137–48.

Biographies

SAURABH AMIN (amins@mit.edu) is an assistant professor in the Department of Civil and Environmental Engineering, Massachusetts Institute of Technology (MIT). His research focuses on the design and implementation of high-confidence network control algorithms for critical infrastructures, including transportation, water, and energy distribution systems. He received his B.Tech. in civil engineering from the Indian Institute of Technology Roorkee in 2002, M.S. in transportation engineering from the University of Texas at Austin in 2004, and Ph.D. in systems engineering from the University of California at Berkeley in 2011.

GALINA A. SCHWARTZ is a research economist in the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley. Her primary expertise is game theory and microeconomics. She has published on the subjects of network neutrality, cyber risk management and modeling of cyber-insurance markets, and security and privacy of cyber-physical systems. In her earlier research, she has applied contract theory to study the interplay between information, transaction costs, institutions and regulations. She has been on the faculty in the Ross School of Business at the University of Michigan, Ann-Arbor, and has taught in the Economics Departments at the University of California, Davis and Berkeley. She received her M.S. in mathematical physics from Moscow Institute of Engineering Physics, Russia, and Ph.D. in economics from Princeton University in 2000.

ALEFIYA HUSSAIN is a computer scientist at the University of Southern California's Information Sciences Institute (USC/ISI). Her research interests include statistical signal processing, protocol design, cyber security, and network measurement systems. She received her B.E. in computer engineering from the University of Pune, India, in 1997 and Ph.D. in computer science from University of Southern California in 2005. Prior to joining USC/ISI, she was a senior principal scientist at Sparta Inc.