Optimal Cybersecurity Investments Using SIS Model: Weakly Connected Networks

Van Sy Mai
CTL, NIST
Gaithersburg, MD USA
vansy.mai@nist.gov

Richard J. La
ACMD, NIST & Univ. of MD
College Park, MD USA
hyongla@umd.edu

Abdella Battou
CTL, NIST
Gaithersburg, MD USA
abdella.battou@nist.gov

***Abstract*—We study the problem of minimizing the (time) average security costs in large systems comprising many interdependent subsystems, where the state evolution is captured by a susceptible-infected-susceptible (SIS) model. The security costs reflect security investments, economic losses and recovery costs from infections and failures following successful attacks. However, unlike in existing studies, we assume that the underlying dependence graph is only weakly connected, but not necessarily strongly connected. When the dependence graph is not strongly connected, existing approaches to computing optimal security investments cannot be applied. Instead, we show that it is still possible to find a good solution by *perturbing* the problem and establishing necessary continuity results that then allow us to leverage the existing algorithms.**

I. Introduction

In complex engineering systems, comprising systems work together to deliver their services, e.g., information and communication networks and power systems, introducing interdependence among the systems. This interdependence among systems also allows a local failure and infection of a system by malware to spread to other systems. Therefore, the structure of interdependence among the systems should be taken into consideration when determining their security investments.

There is already a large volume of literature that examines how to optimize the (security) investments in complex systems or the mitigation of disease spread. For example, in [6], [8], [9], [5], researchers adopted a game theoretic formulation to study the problem of security investments with distributed agents. In another line of research, which is more closely related to our study, researchers investigated optimal strategies using vaccines/immunization (prevention) [4], [16], antidotes or curing rates (recovery) [3], [10], [15] or a combination of both preventive and recovery measures [14], [17]. In addition, in recent studies [11], [12], Mai *et al.* investigated the problem of minimizing the (time) average costs of a system operator, where the costs includes both security investments and recovery/repair costs ensuing infections or failures.

All of theses studies assume that the underlying dependence graph is strongly connected. In this study, we adopt the framework used in [11], [12] but allow the dependence graph to be *weakly* connected. When the graph is only weakly connected, some of key properties and results proved for strongly connected networks do not hold. As a result, we cannot directly apply the algorithms from existing studies, including those of [12].

In a related study, Khanafer et al. [7] extended the earlier studies on the stability of the susceptible-infectious-susceptible (SIS) model (e.g., [16]) to weakly connected networks. A weakly connected network comprises a set of strongly connected components (SCCs) $\{S_1, S_2, \ldots, S_n\}$. They assumed that the n SCCs can be ordered $S_1 \prec S_2 \prec \cdots \prec S_n$, where $S_i \prec S_{i+1}$ indicates the presence of a directed path from a node in S_i to another node in S_{i+1} but not vice versa, and proved the following: (i) if every SCC has a reproduction number less than 1, then the disease-free state is the unique globally asymptotically stable (GAS) equilibrium; and (ii) if S_1 has a reproduction number larger than 1 and every other SCC has a reproduction number smaller than 1, then there is a unique endemic GAS equilibrium.

In our model, attacks targeting systems arrive according to some (stochastic) process. Successful infections of systems can spread to other systems via dependence among the systems. The system operator decides suitable security investments to fend off the attacks, which in turn determine the *breach probability* that they fall victim to attacks and become infected. Our goal is to minimize the (time) average costs of the system operator managing a large system comprising many systems, e.g., large enterprise intranets. The overall costs in our model account for both security investments and recovery/repair costs ensuing infections, which we call *infection costs*.

Contributions: This paper presents an important extension of the work reported in [12] to more general and common situations in practice, where the underlying dependence graph is only weakly connected. Our approach based on perturbation of the external attack rates allows us to leverage efficient methods for solving the nonconvex perturbed problem approximately. In particular, we show that the optimal point and optimal value of the problem are continuous and increasing in the perturbation vector. As a result, we can solve the perturbed problem instead, for which suboptimality can be quantified using computable upper and lower bounds on the optimal value. We also provide a sufficient condition under which these bounds coincide, i.e., the perturbed problem can be solved exactly despite its nonconvexity.

Notation and Terminology: Let $\mathbb{R}$ and $\mathbb{R}_+$ denote the set of real numbers and nonnegative real numbers, respectively.

For a matrix $A = [a_{i,j}]$, let $a_{i,j}$ denote its (i,j) element, A^T its transpose, and $\rho(A)$ its spectral radius. For two matrices A and B, we write $A \geq B$ if $A - B$ is a nonnegative matrix. We use boldface letters and numbers to denote vectors, e.g., $\mathbf{x} = [x_1, ..., x_n]^\mathsf{T}$ and $\mathbf{1} = [1, ..., 1]^\mathsf{T}$. For any two vectors $\mathbf{x}$ and $\mathbf{y}$ of the same dimension, $\mathbf{x} \circ \mathbf{y}$ is their element-wise product. For $\mathbf{x} \in \mathbb{R}^n$, $\mathtt{diag}(\mathbf{x}) \in \mathbb{R}^{n \times n}$ denotes the diagonal matrix with diagonal elements $x_1, \ldots, x_n$. For $\mathbf{x} > \mathbf{0}$, $\mathbf{x}^{-1}$ denotes its element-wise inverse, i.e., $\mathbf{x}^{-1} = [x_1^{-1}, ..., x_n^{-1}]^\mathsf{T}$.

A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a vertex set $\mathcal{V}$ and an edge set $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$. A directed path is a sequence of edges in the form $\big((i_1, i_2), (i_2, i_3), ..., (i_{k-1}, i_k)\big)$. The graph $\mathcal{G}$ is strongly connected if there is a directed path from each node to any other node. The directed graph $\mathcal{G}$ is said to be weakly connected if the undirected graph we obtain after replacing its directed edges with undirected edges is connected.

The rest of the paper is organized as follows: Section II describes the setup and the problem formulation. Section III presents the perturbed problem and the main results, followed by our solution approach in Section IV. Section V provides some numerical results.

II. Model and Formulation

Suppose that the overall system consists of N systems, and let $\mathcal{A} = \{1, \ldots, N\}$ denote the set of systems. The set $\mathcal{A}$ can be partitioned into $\{\mathcal{A}_p, \mathcal{A}_n\}$. Each system $i \in \mathcal{A}_p \subsetneq \mathcal{A}$ experiences external attacks from malicious actors in accordance with a Poisson process with rate $\lambda_i > 0$. On the other hand, the systems in $\mathcal{A}_n$ do not experience external attacks and, hence, $\lambda_i = 0$ for all $i \in \mathcal{A}_n$. When a system experiences an attack, it suffers an infection and subsequent economic losses with some probability, called *breach probability*.

In addition to external attacks from malicious actors, systems also experience *secondary* attacks from other infected systems. Thus, even the nodes in $\mathcal{A}_n$ can experience secondary attacks. When a system suffers a secondary attack, it becomes infected with the same breach probability mentioned above. In other words, the breach probability is the same whether the attack is external or secondary.

The breach probability of system i depends on the security investment on the system: let $s_i \in \mathbb{R}_+$ be the security investment on system i (e.g., investments in monitoring and diagnostic tools). The breach probability of system i is determined by some function $q_i : \mathbb{R}_+ \to (0, 1]$. In other words, when the operator invests s_i on system i, its breach probability is equal to $q_i(s_i)$. We assume that q_i is decreasing, strictly convex and continuously differentiable for all $i \in \mathcal{A}$. It was shown [2] that, under some conditions, the breach probability is decreasing and log-convex (hence, strictly convex).

This can model the spread of virus/malware or failures in complex systems. The rate at which the infection of system i causes that of another system j is denoted by $\beta_{i,j} \in \mathbb{R}_+$. When $\beta_{i,j} > 0$, we say that system i supports system j or system j depends on system i. Let $B = [b_{i,j} : i, j \in \mathcal{A}]$ be an $N \times N$ matrix that describes the infection rates among systems, where the element $b_{i,j}$ is equal to $\beta_{j,i}$. We adopt the convention $\beta_{i,i} = 0$ for all $i \in \mathcal{A}$.

When system i falls victim to an attack and becomes infected, the operator incurs costs at a certain rate for recovery (e.g., inspection and repair of servers). Besides recovery costs, the infection of system i may also cause economic losses if, for example, some servers in system i have to be taken offline for inspection and repair and are inaccessible during the period to other systems that depend on the servers. To model the recovery costs and economic losses, we assume that the infection of system i causes total losses of c_i per unit time. Recovery times of system i following its infections are modeled using independent and identically distributed exponential random variables with parameter $\delta_i > 0$. Furthermore, the recovery times of different systems are mutually independent.

A. Model

Define a directed graph $\mathcal{G} = (\mathcal{A}, \mathcal{E})$, where a directed edge from system i to system j, denoted by (i, j), belongs to the edge set $\mathcal{E}$ if and only if $\beta_{i,j} > 0$. Unlike in our earlier studies [11], [12], here we do not assume that matrix B is irreducible; instead, we assume that the graph $\mathcal{G}$ is only weakly connected, but not strongly connected.[1] We say that node i is *exposed* to attackers if there is a directed path from some node $j \in \mathcal{A}_p$ to node i, and denote the set of exposed nodes by $\mathcal{A}_E$. We call the remaining nodes *accessible* to the attackers and denote the set of accessible nodes by $\mathcal{A}_A$.

Let us comment on the distinction between 'exposed' versus 'accessible' systems. In our model, exposed systems may represent systems known to malicious actors, which communicate or exchange information with each other frequently. Thus, the attackers can target them either directly via external attacks or indirectly through secondary attacks. On the other hand, communication or exchange of information between accessible systems and exposed systems is highly asymmetric; most of communication is from accessible systems to exposed systems. Due to very limited communication from exposed systems to accessible systems, their infection rates by exposed systems are difficult to estimate reliably. For this reason, the infection rates are set to zero in our model even though they can still be vulnerable to occasional infections. As we will show, such systems can still serve as reservoir for infection in that the stable equilibrium for some of them may be endemic and they will be at 'infected' state with a strictly positive probability at steady state and, thus, can infect other exposed systems. For this reason, it is important to model such accessible systems.

A weakly connected network can be partitioned into a set of maximally strongly connected components (MSCCs) $\{C_1, C_2, \ldots, C_m\} =: \mathcal{V}^{(C)}$ [7].[2] The set $\mathcal{A}_E$ comprises a subset of MSCCs, and $\mathcal{A}_A$ includes the remaining MSCCs. Based on this observation, we construct another directed graph

[1] If the graph $\mathcal{G}$ is not weakly connected, we can consider each weakly connected component of $\mathcal{G}$ separately.

[2] A subgaph of a directed graph is maximally strongly connected if (a) it is strongly connected and (b) adding another node leads to a subgraph that is no longer strongly connected.

$\mathcal{G}^{(C)} = (\mathcal{V}^{(C)}, \mathcal{E}^{(C)})$, where a directed edge from v to v' in $\mathcal{E}^{(C)}$ indicates $\beta_{i,j} > 0$ for some system i in v and another system j in v'. Since the vertices in $\mathcal{V}^{(C)}$ are MSCCs, there is no directed cycle in $\mathcal{G}^{(C)}$, i.e., $\mathcal{G}^{(C)}$ is a directed acyclic graph (DAG). Using this property, we can show that $\mathcal{V}^{(C)}$ has leaf vertices with no incoming edges; if this were not true, every vertex in $\mathcal{V}^{(C)}$ has an incoming edge. Since $\mathcal{V}^{(C)}$ is a finite set, this implies that there is a directed cycle, which contradicts the assumption that the vertices in $\mathcal{V}^{(C)}$ are MSCCs.

We partition the vertex set $\mathcal{V}^{(C)}$ into $\{\mathcal{V}_0, \ldots, \mathcal{V}_m\}$, where $\mathcal{V}_\ell$, $\ell = 0, 1, \ldots, m$, is the set of MSCCs whose *maximum* distance from leaf vertices is ℓ.[3] Obviously, $\mathcal{V}_0$ is the set of leaf vertices. Note that there is no directed edge coming into any vertex in $\mathcal{V}_k$ from any other vertex in $\mathcal{V}_\ell$, $\ell \geq k$. Hence, the frequency of attacks experienced by a system i that belongs to some MSCC in $\mathcal{V}_k$ depends only on the states of systems in $\cup_{\ell=0}^{k-1}\mathcal{V}_\ell$ along with λ_i.

B. Dynamics and Equilibria

We adopt a similar framework used in [11], [12] and use the SIS model to capture the evolution of the system state. Let $p_i(t)$ be the probability that system i will be at the 'infected' state (I) at time $t \in \mathbb{R}_+$, and define $\mathbf{p}(t) := (p_i(t) : i \in \mathcal{A})$. The dynamics of $\mathbf{p}(t)$ are approximated by the following (Markov) differential equations for $t \in \mathbb{R}_+$:

$$\dot{\mathbf{p}}(t) = (\mathbf{1} - \mathbf{p}(t)) \circ \mathbf{q}(\mathbf{s}) \circ \big(\boldsymbol{\lambda} + B\mathbf{p}(t)\big) - \boldsymbol{\delta} \circ \mathbf{p}(t) \qquad (1)$$

where $\mathbf{p}(0) \in [0,1]^N$, $\mathbf{s} := (s_i : i \in \mathcal{A})$ is the security investment vector, and $\mathbf{q}(\mathbf{s}) = (q_i(s_i) : i \in \mathcal{A})$ is the corresponding breach probability vector.

Suppose that for each security investment vector $\mathbf{s}$, $\mathbf{p}(t)$ converges to a unique stable equilibrium $\bar{\mathbf{p}}(\mathbf{s})$ (the existence and uniqueness of such an equilibrium will be addressed in the subsequent section). Since the unique stable equilibrium of the differential system in (1) specifies the probability that each system will be infected at steady state, the average cost of the system is given by

$$F(\mathbf{s}) := w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\bar{\mathbf{p}}(\mathbf{s}),$$

where $w(\mathbf{s})$ is the cost of investing $\mathbf{s}$ in the security of the systems (e.g., $w(\mathbf{s}) = \sum_{i\in\mathcal{A}} s_i$), and $\mathbf{c} := (c_i : i \in \mathcal{A})$. We are interested in solving the following problem:

$$F^* := \min_{\mathbf{s}\in\mathcal{S}} \; F(\mathbf{s}) = \min_{\mathbf{s}\in\mathcal{S}} \; \big(w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\bar{\mathbf{p}}(\mathbf{s})\big) \qquad (2)$$

where $\mathcal{S} \subset \mathbb{R}_+^N$ is the feasible set for $\mathbf{s}$. Throughout the paper, we assume that the cost function w is a convex function and the feasible set $\mathcal{S}$ is a convex set.

The main challenge in solving this problem is that $\bar{\mathbf{p}}(\mathbf{s})$ does not have a closed-form expression and need not be convex. For this reason, we consider the following alternative formulation with a higher dimension. From (1), for fixed $\mathbf{s}$, $\bar{\mathbf{p}}(\mathbf{s}) \in [0,1]^N$ is a solution to the following equation:

$$(\mathbf{1} - \mathbf{p}) \circ \mathbf{q}(\mathbf{s}) \circ \big(\boldsymbol{\lambda} + B\mathbf{p}\big) = \boldsymbol{\delta} \circ \mathbf{p}$$

Since $\mathbf{q}(\mathbf{s}) > \mathbf{0}$, the above equation is equivalent to

$$(\mathbf{1} - \mathbf{p}) \circ \big(\boldsymbol{\lambda} + B\mathbf{p}\big) - \mathbf{q}(\mathbf{s})^{-1} \circ \boldsymbol{\delta} \circ \mathbf{p} = \mathbf{0}, \qquad (3)$$

where $\mathbf{q}(\mathbf{s})^{-1} := (q_i(s_i)^{-1} : i \in \mathcal{A})$. Now, (2) can be reformulated in a following more explicit form:

$$[\mathbf{P}] \qquad \min_{\mathbf{s}\in\mathcal{S}, \mathbf{p}\in[0,1]^N} f(\mathbf{s},\mathbf{p}) := w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}$$
$$\text{subject to} \quad (3)$$

Unfortunately, depending on $\boldsymbol{\lambda}$ and B, for fixed $\mathbf{s}$, there may be more than one $\mathbf{p}$ that satisfies (3), rendering the problem nonconvex. This problem does not arise when B is irreducible (i.e., $\mathcal{G}$ is strongly connected) and $\boldsymbol{\lambda} \gneq \mathbf{0}$ (as considered in [12]) because the uniqueness of the solution is already ensured. However, this is not the case when $\mathcal{G}$ is only weakly connected.

Our main idea to tackling this issue is as follows. For our problem **[P]**, we are interested in a solution of (3) which is a stable equilibrium of (1). Even when the stable equilibrium is unique, explicitly computing the unique stable equilibrium to carry out the optimization in (2) does not lead to a computationally efficient approach because the optimization problem is not convex. In order to skirt this issue, in the following section, we propose a more practical approach based on perturbed problems, which then allows us to leverage the efficient algorithms proposed in [12].

III. Perturbed Problem and Main Results

In this section, we describe how we can find a good solution to the problem in (2) in a computationally efficient manner. To this end, we first construct a new approximated problem by perturbing the attack arrival rate $\boldsymbol{\lambda}$ by adding a nonnegative vector $\boldsymbol{\varepsilon} \gneq \mathbf{0}$. Although we can work with any nonnegative vector $\boldsymbol{\varepsilon}$ such that a unique stationary vector $\bar{\mathbf{p}}$ is strictly positive, in order to simplify our exposition, we assume that the perturbation vector $\boldsymbol{\varepsilon}$ takes the form $\boldsymbol{\varepsilon} = \epsilon\mathbf{1}$ with $\epsilon > 0$.[4] In other words, we perturb the external attack rate of every system by ϵ. For fixed $\epsilon \geq 0$, we define $\boldsymbol{\lambda}^\epsilon := \boldsymbol{\lambda} + \epsilon\mathbf{1}$.

Suppose that the security investment $\mathbf{s}$ is fixed. Then, for all $\epsilon > 0$, there is unique $\bar{\mathbf{p}}^\epsilon(\mathbf{s}) > \mathbf{0}$ that satisfies the following [12]:

$$(\mathbf{1} - \mathbf{p}) \circ \big(\boldsymbol{\lambda}^\epsilon + B\mathbf{p}\big) - \mathbf{q}(\mathbf{s})^{-1} \circ \boldsymbol{\delta} \circ \mathbf{p} = \mathbf{0} \qquad (4)$$

Define

$$F_\epsilon^* := \min_{\mathbf{s}\in\mathcal{S}} \quad \big(w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\bar{\mathbf{p}}^\epsilon(\mathbf{s})\big). \qquad (5)$$

Our approach is as follows: first, we know that, for fixed $\epsilon > 0$, $\bar{\mathbf{p}}^\epsilon(\mathbf{s})$ is continuous in $\mathbf{s}$ [12]. We will prove that, for fixed $\mathbf{s}$, $\bar{\mathbf{p}}^\epsilon(\mathbf{s})$ is continuous in $\epsilon \geq 0$. From the continuity of the objective function, this implies that F_ϵ^* is continuous in

[3] Here, the maximum distance from leaf vertices is defined to be the maximum among the maximum distances from the leaf vertices, i.e., the length of the longest path from any leaf vertex.

[4] As long as we perturb the attack arrival rate of at least one system in each accessible MSCC (by ϵ), our results continue to hold.

$\epsilon \geq 0$. Finally, we can solve the perturbed problem for some small ϵ, and use the solution to the perturbed problem as an approximated solution to the original problem with $\epsilon = 0$.

Recall that when $\epsilon = 0$, there could be multiple solutions to (3). Therefore, in order to make use of this observation, we need to establish that $\lim_{\epsilon \to 0} \bar{\mathbf{p}}^{\epsilon}(\mathbf{s})$ exists and coincides with the *unique* stable equilibrium of (3). To this end, the following lemma establishes the existence of a unique stable equilibrium of (3). Due to a space constraint, the proofs of main results are omitted here and can be found in [13].

Proposition 1. *For fixed $\mathbf{s} \in \mathcal{S}$, there is a unique stable equilibrium $\bar{\mathbf{p}}(\mathbf{s})$ of (1), which satisfies (3).*

A. Continuity of Stable Equilibria $\bar{\mathbf{p}}$

For $\mathbf{s} \in \mathcal{S}$, define $\mathbf{g}^{\mathbf{s}} : \mathbb{R}^{n+1} \to \mathbb{R}^n$, where

$$\mathbf{g}^{\mathbf{s}}(\epsilon, \mathbf{p}) = (\mathbf{1} - \mathbf{p}) \circ \big(\boldsymbol{\lambda}^{\epsilon} + B\mathbf{p}\big) - \mathbf{q}(\mathbf{s})^{-1} \circ \boldsymbol{\delta} \circ \mathbf{p}. \quad (6)$$

Clearly, $\mathbf{g}^{\mathbf{s}}$ is a continuously differentiable function and its partial derivative w.r.t. $\mathbf{p}$ is given by

$$\partial_{\mathbf{p}} \mathbf{g}^{\mathbf{s}}(\epsilon, \mathbf{p}) = \mathtt{diag}(\mathbf{1} - \mathbf{p})B - \mathtt{diag}(\mathbf{q}(\mathbf{s})^{-1} \circ \boldsymbol{\delta} + \boldsymbol{\lambda}^{\epsilon} + B\mathbf{p}).$$

Proposition 2. *For each $\mathbf{s} \in \mathcal{S}$, $\bar{\mathbf{p}}^{\epsilon}(\mathbf{s})$ is continuous in $\epsilon > 0$.*

For $\mathbf{s} \in \mathcal{S}$, define a mapping $\bar{\boldsymbol{\lambda}}^{\mathbf{s}} : \mathbb{R}_+ \to \mathbb{R}_+^N$, where, for each $v \in \mathcal{V}^{(C)}$,

$$\bar{\boldsymbol{\lambda}}^{\mathbf{s}}_v(\epsilon) = \begin{cases} \boldsymbol{\lambda}^{\epsilon}_v + \big(B_{-v}\bar{\mathbf{p}}^{\epsilon}_{-v}(\mathbf{s})\big)_v & \text{if } \epsilon > 0, \\ \boldsymbol{\lambda}_v + \big(B_{-v}\bar{\mathbf{p}}_{-v}(\mathbf{s})\big)_v & \text{if } \epsilon = 0, \end{cases}$$

where B_{-v} is a submatrix of B without the columns corresponding to the systems that belong to the MSCC v, and $\bar{\mathbf{p}}^{\epsilon}_{-v}(\mathbf{s})$ and $\bar{\mathbf{p}}_{-v}(\mathbf{s})$ are the subvectors of $\bar{\mathbf{p}}^{\epsilon}(\mathbf{s})$ and $\bar{\mathbf{p}}(\mathbf{s})$, respectively, obtained after removing the elements for the systems in v. Obviously, $\bar{\boldsymbol{\lambda}}^{\mathbf{s}}(\epsilon)$ tells us the total attack rates, including both external attacks from malicious actors and secondary attacks coming from other systems that do not belong to the same MSCC, at the unique stable equilibrium as a function of the security investments $\mathbf{s}$ and perturbation ϵ.

For each $v \in \mathcal{V}^{(C)}$, let $n_v := |v|$. For each $\mathbf{s} \in \mathcal{S}$, define $\hat{\mathbf{g}}^{\mathbf{s}}_v : \mathbb{R}^{n_v+1} \to \mathbb{R}^{n_v}$, where

$$\hat{\mathbf{g}}^{\mathbf{s}}_v(\epsilon, \mathbf{p}_v) = (\mathbf{1} - \mathbf{p}_v) \circ \big(\bar{\boldsymbol{\lambda}}^{\mathbf{s}}_v(\epsilon) + B_v\mathbf{p}_v\big) - \mathbf{q}_v(\mathbf{s})^{-1} \circ \boldsymbol{\delta}_v \circ \mathbf{p}_v, \quad (7)$$

and B_v is the $n_v \times n_v$ submatrix of B with columns and rows corresponding to the systems in v. Clearly, $\hat{\mathbf{g}}^{\mathbf{s}}_v$ is the mapping $\mathbf{g}$ defined in (6) restricted to the MSCC v with fixed attack rates $\bar{\boldsymbol{\lambda}}^{\mathbf{s}}_v(\epsilon)$ coming from outside the MSCC v. We point out that, from the viewpoint of a system, there is no distinction between an external attack from a malicious actor or a secondary attack coming from another system in a different MSCC. Clearly, for fixed $\epsilon > 0$, $\bar{\mathbf{p}}^{\epsilon}(\mathbf{s})$ is the unique solution that satisfies

$$\hat{\mathbf{g}}^{\mathbf{s}}_v\big(\epsilon, \bar{\mathbf{p}}^{\epsilon}_v(\mathbf{s})\big) = \mathbf{0} \quad \text{for all } v \in \mathcal{V}^{(C)}. \quad (8)$$

Similarly, $\bar{\mathbf{p}}(\mathbf{s})$ satisfies

$$\hat{\mathbf{g}}^{\mathbf{s}}_v\big(0, \bar{\mathbf{p}}_v(\mathbf{s})\big) = \mathbf{0} \quad \text{for all } v \in \mathcal{V}^{(C)}. \quad (9)$$

Proposition 3. *Fix $\mathbf{s} \in S$. Then, for any decreasing positive sequence $\{\epsilon_l : l \in \mathbb{N}\}$ with $\lim_{l \to \infty} \epsilon_l = 0$, we have $\lim_{l \to \infty} \bar{\mathbf{p}}^{\epsilon_l}(\mathbf{s}) = \bar{\mathbf{p}}(\mathbf{s})$.*

One would expect that the minimum cost we can achieve by solving (5) would not decrease with ϵ. The following proposition tells us that this is indeed the case.

Proposition 4. *For each $\mathbf{s} \in \mathcal{S}$, $\bar{\mathbf{p}}^{\epsilon}(\mathbf{s})$ is increasing in $\epsilon > 0$. As a result, the optimal value F^*_{ϵ} is increasing in $\epsilon \geq 0$.*

Proposition 4, together with the continuity of the objective functions in (2) and (5) and Propositions 2 and 3, tells us that as we reduce ϵ, the optimal value F^*_{ϵ} of the perturbed problem will decrease to the optimal value F^* of the original problem. Therefore, this suggests that if we solve the perturbed problem (5) with sufficiently small ϵ, the optimal point we obtain will likely be a good solution for the original problem, and the optimal value of (5) will serve as an upper bound to F^*.

IV. Solving the Perturbed Problem

In this section, we discuss how we can solve the perturbed problem in (5) by extending the formulation and adapting the algorithms proposed in [12]. First, we rewrite the perturbed problem as follows:

$$[\mathbf{PP}] \qquad \begin{aligned} \min_{\mathbf{s},\mathbf{p}} \quad & f(\mathbf{s}, \mathbf{p}) = w(\mathbf{s}) + \mathbf{c}^{\mathsf{T}}\mathbf{p} \\ \text{s.t.} \quad & (\mathbf{p}^{-1} - \mathbf{1}) \circ \big(\boldsymbol{\lambda}^{\epsilon} + B\mathbf{p}\big) = \mathbf{q}(\mathbf{s})^{-1} \circ \boldsymbol{\delta} \\ & \mathbf{s} \in \mathcal{S}, \quad \mathbf{p} > 0 \end{aligned} \quad (10)$$

Note that constraint (10) is the same as (4) when $\mathbf{p} > 0$. Next, we introduce the following change of variables:

$$d_i = (q_i(s_i))^{-1} \quad \text{or, equivalently,} \quad s_i = q_i^{-1}(d_i^{-1}),$$

where q_i^{-1} is the inverse map of q_i. Let $\mathcal{D} = \{\mathbf{q}(\mathbf{s})^{-1} \mid \mathbf{s} \in \mathcal{S}\}$. As a result, we obtain the following equivalent problem:

$$[\mathbf{EP}] \qquad \begin{aligned} \min_{\mathbf{d},\mathbf{p}} \quad & \tilde{f}(\mathbf{d}, \mathbf{p}) = \tilde{w}(\mathbf{d}) + \mathbf{c}^{\mathsf{T}}\mathbf{p} \\ \text{s.t.} \quad & (\mathbf{p}^{-1} - \mathbf{1}) \circ \big(\boldsymbol{\lambda}^{\epsilon} + B\mathbf{p}\big) = \mathbf{d} \circ \boldsymbol{\delta} \\ & \mathbf{d} \in \mathcal{D}, \quad \mathbf{p} > 0 \end{aligned} \quad (11)$$

where $\tilde{w}(\mathbf{d}) := w(\mathbf{q}^{-1}(\mathbf{d}^{-1}))$. This problem is nonconvex in general because of the cost function, the equality constraint in (11), and possibly the constraint set $\mathcal{D}$. But, as shown in [12], a local minimizer can be found efficiently using a reduced gradient method (RGM), providing an upper bound on F^*_{ϵ}.

Suppose that $\mathcal{D}$ is convex and $\tilde{w}(\mathbf{d})$ is convex in $\mathbf{d} \in \mathcal{D}$, which implies the convexity of $\tilde{f}(\mathbf{d}, \mathbf{p})$. It can be shown that, provided that w is convex and increasing, e.g., $w(\mathbf{s}) = \mathbf{1}^T\mathbf{s}$, the second assumption holds for a family of breach probability functions $q_i(s_i) = (1 + \kappa_i s_i)^{-\beta_i}$ for some $\kappa_i > 0$ and $\beta_i \in (0, 1]$, in which case $s_i = (d_i^{1/\beta_i} - 1)\kappa_i^{-1}$ is convex in d_i. When such assumption does not hold, one might consider a suitable convex lower bound of $\tilde{w}(\mathbf{d})$ instead. In addition, when $\mathcal{S} = \{\mathbf{s} \in \mathbb{R}_+^N \mid \mathbf{1}^{\mathsf{T}}\mathbf{s} \leq s_{\text{budget}}\}$, where s_{budget} is the total budget, the constraint set $\mathcal{D}$ would be convex for the aforementioned family of breach probability functions.

Following an approach analogous to [12], we can obtain a convex relaxation of the perturbed problem to deal with the nonconvex constraint in (10). Define the following variables:

$$\mathbf{p} := e^{-\mathbf{y}},\ \mathbf{t} := \boldsymbol{\lambda}^{\epsilon} \circ e^{\mathbf{y}},\ U := \mathtt{diag}(e^{\mathbf{y}})B\mathtt{diag}(e^{-\mathbf{y}}) \quad (12)$$

Using these new variables, (11) can be rewritten as follows.

$$\mathbf{t} + U\mathbf{1} = \boldsymbol{\lambda}^{\epsilon} + B\mathbf{p} + \mathbf{d} \circ \boldsymbol{\delta} \quad (13)$$

Finally, we relax the equality constraints in (12) using the following convex inequality constraints.

$$\mathbf{1} \geq \mathbf{p} \geq e^{-\mathbf{y}},\ \mathbf{t} \geq \boldsymbol{\lambda}^{\epsilon} \circ e^{\mathbf{y}},\ U \geq \mathtt{diag}(e^{\mathbf{y}})B\mathtt{diag}(e^{-\mathbf{y}}) \quad (14)$$

These yield the following convex relaxation of **[EP]**:

$$[\mathbf{CR}] \qquad \min_{\mathbf{d},\mathbf{p},\mathbf{y},\mathbf{t},U} \quad \tilde{f}(\mathbf{d},\mathbf{p}) = \tilde{w}(\mathbf{d}) + \mathbf{c}^{\mathsf{T}}\mathbf{p} \quad (15)$$
$$\text{s.t.} \quad (13), (14),\ \mathbf{d} \in \mathcal{D},\ \mathbf{y} > \mathbf{0}$$

Theorem 1. *Suppose* $\mathbf{x}_{\mathrm{R}}^{+} := (\mathbf{d}^{+}, \mathbf{p}^{+}, \mathbf{y}^{+}, \mathbf{t}^{+}, U^{+})$ *is an optimal point of* **[CR]**. *Then, we have*

$$\tilde{f}(\mathbf{d}^{+}, \mathbf{p}^{+}) \leq F_{\epsilon}^{*} \leq \tilde{f}(\mathbf{d}', \mathbf{p}'), \quad (16)$$

where the pair $(\mathbf{d}', \mathbf{p}')$ *is a feasible point of* **[EP]** *given by*

$$\mathbf{p}' = e^{-\mathbf{y}^{+}} \quad \textit{and} \quad \mathbf{d}' = \mathbf{d}^{+} + \mathtt{diag}(\boldsymbol{\delta}^{-1})B(\mathbf{p}^{+} - \mathbf{p}').$$

In addition, **[CR]** *is exact, i.e.,* $(\mathbf{d}', \mathbf{p}')$ *solves* **[EP]**, *if*

$$B^{\mathsf{T}}\mathtt{diag}(\boldsymbol{\delta}^{-1})\nabla\tilde{w}(\mathbf{d}) \leq \mathbf{c} \quad \textit{for all } \mathbf{d} \in \mathcal{D}. \quad (17)$$

We end this section with the following remarks. First, note that condition (17) can be checked prior to solving the relaxed problem **[CR]**. This can be done easily when $\nabla\tilde{w}$ (or an upper bound) is known and the constraint set $\mathcal{D}$ is simple enough. Second, our approach in this section is based on the reformulation **[PP]** of the original problem, where we use the condition that $\mathbf{p} > \mathbf{0}$; in the relaxation **[CR]**, this is ensured by imposing the condition $\mathbf{p} \geq e^{\mathbf{y}}$. If $p_i = 0$ for some i at an optimal point of the original problem (which can happen when $\mathcal{G}$ is only weakly connected and $\lambda_i = 0$), this condition would be violated and we cannot use (10); in the relaxation, this would correspond to having $y_i \to \infty$. As a result, our perturbation of the attack rates introduced in Section IV proves to be meaningful in practice as it not only allows us to employ efficient methods to solve the problem approximately (as will be demonstrated numerically in the next section), but also takes into account more stringent scenarios (with varying external attack rates).

V. Numerical Examples

In this section, we provide some numerical results to evaluate the proposed method. Our numerical studies are carried out in MATLAB (version R2018b) on a laptop with 8GB RAM and a 2.4GHz Intel Core i5 processor.[5] We assume that the breach probability can be approximated (in the regime of interest) using a function of the form $q_i(s) = (1+\kappa_i s)^{-1}$ for all $i \in \mathcal{A}$. In practice, the parameter $\kappa_i > 0$ models how quickly the breach probability decreases with security investment for system i. Here, for simplicity, we take $\kappa_i = \delta_i^{-1}$. In the first example, we use an artificial scale-free network, whereas the second example makes use of an Internet peer-to-peer network.

Example: We consider a weakly connected network consisting of two MSCCs, denoted by $\mathcal{G}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ and $\mathcal{G}_2 = (\mathcal{V}_2, \mathcal{E}_2)$ with $|\mathcal{V}_1| = 50$ and $|\mathcal{V}_2| = 150$. Here, each $\mathcal{G}_i$ is a bidirectional scale-free network generated with the power law parameter for node degrees set to 1.5, and the minimum and maximum node degrees equal to 2 and $\lceil 3\log|\mathcal{V}_i| \rceil$, respectively, in order to ensure network connectivity with high probability. We also add 10 directed edges chosen uniformly at random (u.a.r) from $\mathcal{G}_1$ to $\mathcal{G}_2$.

We fix $\delta_i = 0.1$ for all $i \in \mathcal{A}$ and $\mathcal{S} = \mathbb{R}_{+}^{N}$. The infection rates $\beta_{j,i}$ are chosen u.a.r. between $[0.01, 1]$. We choose $w(\mathbf{s}) = \mathbf{1}^{\mathsf{T}}\mathbf{s}$ and $\mathbf{c} = (\nu\mathbf{1} + 0.2\mathbf{c}_{\mathrm{rand}}) \circ B^{\mathsf{T}}\mathbf{1}$, where the elements of $\mathbf{c}_{\mathrm{rand}}$ are chosen u.a.r in $(0, 1)$, and $\nu \geq 0$ is a varying parameter. We select $\mathbf{c}$ above to reflect an observation that systems that support more neighbors should, on the average, have larger economic costs modeled by c_i (Section II-A). We select u.a.r 10 nodes in $\mathcal{G}_2$ to have positive primary attack rates $\lambda_i = 0.1$. Here, $\mathcal{G}_1$ is not exposed and we consider $\lambda_1^{\epsilon} = \epsilon$ in our perturbed problem.

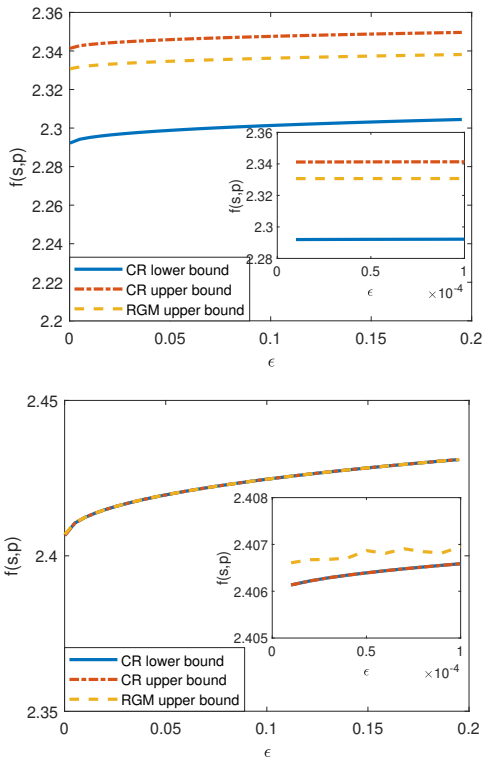

Fig. 1. Bounds on optimal costs (normalized by N) when varying ϵ in two scenarios: $\nu = 0.9$ in *top plot* and $\nu = 1.1$ in *bottom plot*. Insets correspond to a small range of perturbations $\epsilon \in (10^{-5}, 10^{-4})$

Fig. 1 shows the objective function values obtained using (a) the RGM (for finding a local minimizer starting from initial point $(\mathbf{0}, \bar{\mathbf{p}}(\mathbf{0}))$),[6] (b) the optimal point $(\mathbf{d}^{+}, \mathbf{p}^{+})$ of **[CR]** and

[5]Mention of commercial products does not imply NIST's endorsement.

[6]We combine RGM with a backtracking line search algorithm using initial step size $\gamma_0 = 1$ and a shrinking factor of 0.85; see [12] for details.

(c) the feasible point $(\mathbf{d}', \mathbf{p}')$ introduced in Theorem 1, as we vary ϵ. Here we solve **[CR]** using MOSEK package [1]. The top plot corresponds to $\nu = 0.9$ and shows that **[CR]** is not exact because there is a gap between the objective function values achieved by $(\mathbf{d}^+, \mathbf{p}^+)$ and $(\mathbf{d}', \mathbf{p}')$. In fact, in this case the RGM provides a better upper bound on the optimal value than $(\mathbf{d}', \mathbf{p}')$. The bottom plot corresponds to $\nu = 1.1$ and clearly indicates that **[CR]** is exact as both upper and lower bounds overlap. Moreover, we can verify that the condition for **[CR]** to be exact (in Theorem 1) holds for any $\nu \geq 1$. The insets plot the achieved objective function values over a small range of perturbations $\epsilon \in (10^{-5}, 10^{-4})$.

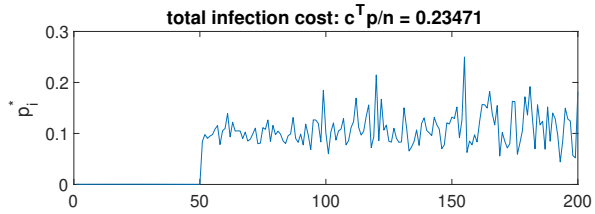

Fig. 2. Optimal solution for the case $\nu = 1.1$ and $\epsilon = 10^{-5}$.

Both MOSEK and RGM take less than 1 second to run for most values of ϵ. However, the RGM can take up to 5 seconds when $\nu = 1.1$ and the value of perturbation ϵ is very small (in the interval $(10^{-5}, 10^{-4})$) because some $\bar{p}_i^{\epsilon}$'s are very close to 0, causing the Jacobian matrix of constraint function (6) to become almost singular. Fig. 2 plots the stable equilibrium $\bar{\mathbf{p}}(\mathbf{s}_{\epsilon}^*)$ for the case with $\nu = 1.1$ and $\epsilon = 10^{-5}$. The figure suggests that the first MSCC comprising systems 1 through 50 will likely be fully protected, i.e., achieves disease-free stable equilibrium, at the optimal point of the original problem in (2). In this case, the condition number of the Jacobian matrix of g in (6) is approximately 3.2×10^4, explaining longer running times of the RGM as discussed above.

Example 2: We consider the Gnutella peer-to-peer network from August 9, 2002 with 8,114 nodes and 26,013 directed edges.[7] We use similar settings as in the first example, except that we set $\lambda_i = 0.01$ for $i = 1, \ldots, \frac{N}{2}$ and 0 otherwise. In the perturbed problem, we select $\lambda_i^{\epsilon} = \epsilon$ for $i = \frac{N}{2} + 1, \ldots, N$.

Fig. 3 shows the lower and upper bounds obtained using the RGM and convex relaxation **[CR]** on the optimal values of the perturbed problem when $\nu = 0.8$ and $\mathbf{c}_{\text{rand}} = \mathbf{0}$. In this example, we first use MOSEK to solve **[CR]**, which takes less than 40 seconds to run for each value of ϵ. Then, we use the RGM with an initial solution $(\mathbf{d}', \mathbf{p}')$ obtained from **[CR]**, which takes less than 10 seconds on average. In this example, **[CR]** is not exact as there is a gap between $(\mathbf{d}^+, \mathbf{p}^+)$ and $(\mathbf{d}', \mathbf{p}')$. But, the gap between that of $(\mathbf{d}^+, \mathbf{p}^+)$ and the local minimizer obtained by the RGM is less than a little over 4 percent. Also, the inset suggests that all three values of the objective function converge as ϵ diminishes, validating our analytical results.

[7]Data is available at https://snap.stanford.edu/data/index.html.

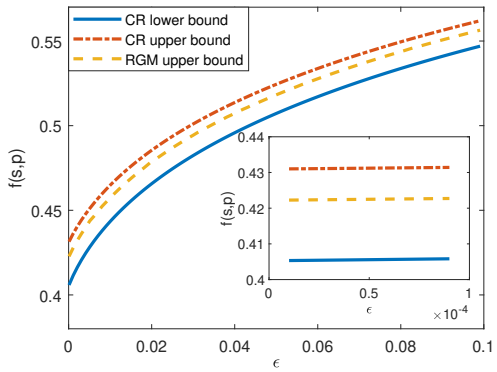

Fig. 3. Bounds on optimal costs (normalized by N) when varying ϵ with $\nu = 0.8$. Insets correspond to a small range of perturbations $\epsilon \in (10^{-5}, 10^{-4})$

REFERENCES

[1] MOSEK ApS. *The MOSEK optim. toolbox for MATLAB manual. Version 9.0.*, 2019. http://docs.mosek.com/9.0/toolbox/index.html.

[2] Yuliy Baryshnikov. IT security investment and Gordon-Loeb's $1/e$ rule. In *Proc. of WEIS*, 2012.

[3] Christian Borgs, Jennifer Chayes, Ayalvadi Ganesh, and Amin Saberi. How to distributed antidote to control epidemics. *Random Structures & Algorithms*, 37(2):204–222, September 2010.

[4] Reuven Cohen, Shlomo Havlin, and Daniel ben Avraham. Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.*, 91(247901), December 2003.

[5] Ashish R. Hota and Shreyas Sundaram. Interdependent security games on networks under behavioral probability weighting. *IEEE Control Netw. Syst.*, 5(1):262–273, March 2018.

[6] Mohammad M. Khalili, Xueru Zhang, and Mingyan Liu. Incentivizing effort in interdependent security games using resource pooling. In *Proc. of NetEcon*, 2019.

[7] Ali Khanafer, Tamer Basar, and Bahman Gahresifard. Stability properties of infection diffusion over directed network. In *Proc. of IEEE Conference on Decision and Control*, 2014.

[8] Richard J. La. Interdependent security with strategic agents and global cascades. *IEEE/ACM Trans. Netw.*, 24(3):1378–1391, June 2016.

[9] Marc Lelarge and Jean Bolot. A local mean field analysis of security investments in networks. In *Proc. of International Workshop on Economics of Networks Systems*, pages 25–30, 2008.

[10] Van Sy Mai, Abdella Battou, and Kevin Mills. Distributed algorithm for suppressing epidemic spread in networks. *IEEE Contr. Syst. Lett.*, 2(3):555–560, 2018.

[11] Van Sy Mai, Richard J. La, and Abdella Battou. Optimal cybersecurity investments for sis model. In *Proc. of IEEE Globecom*, 2020.

[12] Van Sy Mai, Richard J. La, and Abdella Battou. Optimal cybersecurity investments in large networks using SIS model: algorithm design. *IEEE/ACM Trans. Netw.*, 29(6):2453–2466, December 2021.

[13] Van Sy Mai, Richard J. La, and Abdella Battou. Optimal cybersecurity investments using SIS model: weakly connected networks. *Preprint arXiv:2204.06035*, 2022. https://arxiv.org/abs/2204.06035.

[14] Cameron Nowzari, Victor M Preciado, and George J. Pappas. Optimal resource allocation for control of networked epidemic models. *IEEE Control Netw. Syst.*, 4(2):159–169, June 2017.

[15] Stefania Ottaviano, Francesco De Pellegrini, Stefano Bonaccorsi, and Piet Van Mieghem. Optimal curing policy for epidemic spreading over a community network with heterogeneous population. *J. Complex Networks*, 6(6), October 2018.

[16] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George Pappas. Optimal vaccine allocation to control epidemic outbreaks in arbitrary networks. In *Proc. of IEEE Conference on Decision and Control*, pages 7486–7491. IEEE, 2013.

[17] Victor M. Preciado, Michael Zargham, Chinwendu Enyioha, Ali Jadbabaie, and George J. Pappas. Optimal curing policy for epidemic spreading over a community network with heterogeneous population. *IEEE Control Netw. Syst.*, 1(1):99–108, March 2014.