Optimal Cybersecurity Investments in Large Networks Using SIS Model: Algorithm Design

Van Sy Mai, Richard J. La, Abdella Battou

Abstract—We study the problem of minimizing the (time) average security costs in large networks/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. We show that the resulting optimization problem is nonconvex and propose a suite of algorithms – two based on convex relaxations, and the other two for finding a local minimizer, based on a reduced gradient method and sequential convex programming. Also, we provide a sufficient condition under which the convex relaxations are exact and, hence, an optimal solution of the original problem can be recovered. Numerical results are provided to validate our analytical results and to demonstrate the effectiveness of the proposed algorithms.

Index Terms—Cybersecurity investments; Optimization; SIS model

1 Introduction

Today, many modern engineered systems, including information and communication networks and power systems, comprise many interdependent systems. For uninterrupted delivery of their services, the comprising systems must work together and oftentimes support each other. Unfortunately, this interdependence among comprising systems also introduces a source of vulnerability in that it is possible for a local failure or infection of a system by malware to spread to other systems, potentially compromising the integrity of the overall system. Analogously, in social networks, contagious diseases often spread from infected individuals to other vulnerable individuals through contacts or physical proximity.

From this viewpoint, it is clear that the underlying networks that govern the interdependence among systems have a large impact on dynamics of the spread of failures or malware infections. Similarly, the topology and contact frequencies among individuals in social networks significantly influence the manner in which diseases spread in societies. Thus, any sound investments in the security of large systems or the control of epidemics should take into account the interdependence in the systems and social contacts in order to maximize potential benefits from the investments.

In our model, attacks targeting the systems arrive according to some (stochastic) process. Successful attacks on the systems can also spread from infected systems to other systems via aforementioned dependence among the systems. The system operator decides appropriate security investments to fend off the attacks, which in turn determine their breach probability, i.e., the probability that they fall victim to attacks and become infected.

Our goal is to minimize the (time) average costs of a system operator managing a large system comprising many systems, such as large enterprise intranets. The overall costs in our model account for both security investments and recovery/repair costs ensuing infections or failures, which we call *infection costs* in the paper. To this end, we first consider a scenario where malicious actors launch *external* or *primary* attacks. When a primary attack on a system is successful, the infected system can spread it to other systems, which we call *secondary* attacks by the infected systems, to distinguish them from primary attacks. When primary attacks do not stop, it is in general not possible to achieve an infection-free state at steady state.

In the second case, we assume that there are no primary attacks and examine the steady state, starting with an initial state where some systems are infected. The goal of studying this scenario is to get additional insights into scenarios where primary attacks occur infrequently. It turns out that, even in the absence of primary attacks, infections may persist due to secondary attacks and the system may not be able to attain the infection-free steady state.

We formulate the problem of determining the optimal security investments that minimize the average costs as an optimization problem. Unfortunately, this optimization problem is nonconvex and cannot be solved easily. In order to gauge the quality of a feasible solution, we obtain both a lower bound and an upper bound on the optimal value of our problem. A lower bound can be acquired using one of two different convex relaxations of the original problem we propose. For the convex relaxations, we also derive a sufficient condition under which a solution of either convex relaxation also provides a solution to the original nonconvex optimization problem (Lemma 3). An upper bound on the optimal value can be obtained using an algorithm that finds a local minimizer. Here, we propose two methods – a reduced gradient method (RGM) and sequential convex programming (SCP), both of which produce a local minimizer. Together, our approach offers a bound on the optimality gap.

V.-S. Mai and A. Battou are with the National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA. Email: {vansy.mai, abdella.battou}@nist.gov. R.J. La is with the University of Maryland, College Park, MD 20742, USA.

Any mention of commercial products in this paper is for information only; it does not imply recommendation or endorsement by NIST.

Numerical studies show that the computational requirements for the proposed methods are light to modest even for large systems, except for one method that requires the calculation of an inverse matrix. They suggest that, in almost all cases that we considered, the gap between the lower bound on the optimal value and the cost achieved by our solutions is small; in fact, in most cases, the gap is less than 2-3 percent with the gap being less than 0.3 percent in many cases. In addition, when the infection costs are large, which may be true in many practical scenarios, the sufficient condition for the convex relaxations to be exact holds, and we obtain optimal points by solving the convex relaxations. Finally, the RGM is computationally most efficient (with the computational time being less than two seconds in all considered cases and less than 0.1 seconds in most cases) and the quality of solutions is on par with that of other methods. This suggests that the RGM may offer a good practical solution for our problem.

1.1 Related Literature

Given the importance of cybersecurity, robustness of complex systems, and control of epidemics, there is already a large body of literature that examines how to optimize the (security) investments in complex systems [12], [18], the mitigation of disease or infection spread [6], [9], [27], feasibility and case studies of cyber insurance using pre-screening or differentiated pricing based on the security investments of the insured [13], [30], and designing good attack models and effective mitigating defense against attacks [24], [26], [35]. In view of the volume of existing literature, here we summarize only a small set of studies most closely related to our study.

In [11], [14], [16], [18], the authors adopted a game theoretic formulation to study the problem of security investments with distributed agents or autonomous systems that do not coordinate their efforts. The problem we study in this paper is complementary, but is very different from those studied in the aforementioned studies: in our setting, we assume that the system is managed by a *single* operator and are interested in minimizing the average (security) costs over time by determining (nearly) optimal security investments. Our study is applicable to, for example, the problem of finding suitable security investments in large enterprise intranets supporting common business processes or supervisory control and data acquisition systems comprising many subsystems.

In another line of research, which is most closely related to our study, researchers investigated optimal strategies using vaccines/immunization (prevention) [6], [33], antidotes or curing rates (recovery) [4], [21], [29] or a combination of both preventive and recovery measures [27], [34]. For example, [33] studies the problem of partial vaccination via investments at each individual to reduce the infection rates, with the aim of maximizing the exponential decay rate to control the spread of an epidemic. Similarly, [21] examines the problem of determining the optimal curing rates for distributed agents under different formulations. In particular, the last formulation of the problem [21], for which only partial result is obtained, is closely related to a special case of our formulation studied in Section 6.

Key differences between existing studies, including those listed above, and ours can be summarized as follows: first, unlike previous studies that focus on either the expected costs from single or cascading failures/infections [15], [16], [18] or the exponential decay rate to the disease-free state as a key performance metric, we aim to minimize the (time) average costs of a system operator, while accounting for both security investments and infection costs, with both primary and secondary attacks, by modeling time-varying states of systems due to the transmissions of failures/infections. In the presence of primary attacks, it is in general not possible to achieve the infection-free steady state and, thus, the exponential decay rate is no longer a suitable performance metric for our study. Second, unlike some studies that assume that the expected costs/risks seen by systems are convex functions of security investments (e.g., [12]), the expected risks are derived from the steady state equilibrium of a differential system that describes system states and depends on security investments. As it will be clear, the lack of a closed-form expression for the steady state equilibrium complicates considerably the analysis and algorithm design.

Preliminary results of this paper were reported in [22]. In this paper, we extend the findings of [22] in several significant directions. First, we offer an alternative convex relaxation of the original problem. As we demonstrate, this new relaxation technique based on exponential cones, is more efficient and avoids the key issues of the approach based on M-matrix theory presented in [22]. Second, we present another computationally efficient approach to finding a suboptimal solution using SCP together with our M-matrix theory-based approach, which provides an upper bound on the optimal value of the original nonconvex optimization problem. We compare this approach to one based on the RGM and show that the quality of solutions from these two methods is comparable, but the RGM holds a slight computational edge. Finally, we study a special case with no primary attacks, which is related to the epidemic control problem studied in [21], [25], [29]. Although this can be viewed as a limit case of our problem formulation as the rates of primary attacks go to zero, our approaches to obtaining upper and lower bounds of the optimal value require significant modifications for the reason explained in Section 6. Moreover, we derive sufficient conditions for optimality, which can be verified relatively easily.

The rest of the paper is organized as follows: Section 2 explains the notation and terminology we adopt. Section 3 describes the setup and the problem formulation, including the optimization problem. Section 4 discusses two different convex relaxations of the original problem, followed by two methods for finding local minimizers in Section 5. We discuss a special case with no primary attacks in Section 6. Numerical results are provided in Section 7, followed by a discussion on how our formulation and results can be extended in Section 8. We conclude in Section 9.

2 Preliminaries

2.1 Notation and Terminology

Let $\mathbb{R}$ and $\mathbb{R}_+$ denote the set of real numbers and nonnegative real numbers, respectively. Given a set $\mathbb{A}$, we denote the

closure, interior, and boundary of $\mathbb{A}$ by $\text{cl}(\mathbb{A})$, $\text{int}(\mathbb{A})$, and $\partial\mathbb{A}$, respectively.

For a matrix $A = [a_{i,j}]$, let $a_{i,j}$ denote its (i,j) element, A^T its transpose, $\rho(A)$ its spectral radius, and $\underline{\sigma}(A)$ and $\bar{\sigma}(A)$ the smallest and largest real parts of its eigenvalues. For two matrices A and B, we write $A \geq B$ if $A - B$ is a nonnegative matrix. We use boldface letters 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, $\mathbf{x} \circ \mathbf{y}$ and $\frac{\mathbf{x}}{\mathbf{y}}$ are their element-wise product and division, respectively. For $\mathbf{x} \in \mathbb{R}^n$, $\texttt{diag}(\mathbf{x}) \in \mathbb{R}^{n\times n}$ denotes the diagonal matrix with diagonal elements $x_1, \ldots, x_n$.

A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ consists of a set of nodes $\mathcal{V}$, and a set of directed edges $\mathcal{E} \subseteq \mathcal{V}\times\mathcal{V}$. 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.

2.2 M-Matrix Theory

A matrix $A \in \mathbb{R}^{n\times n}$ is an M-matrix if it can be expressed in the form $A = sI - B$, where $B \in \mathbb{R}_+^{n\times n}$ and $s \geq \rho(B)$. The set of (nonsingular) $n \times n$ M-matrices is denoted by ($\mathbb{M}_+^{n\times n}$) $\mathbb{M}^{n\times n}$. Note that this definition implies that the off-diagonal elements of A are nonpositive and the diagonal elements are nonnegative; any matrix satisfying these conditions is called a Z-matrix. We shall make use of the following results on the properties of a nonsingular M-matrix [32].

Lemma 1. Let $A \in \mathbb{R}^{n\times n}$ be a Z-matrix. Then, $A \in \mathbb{M}_+^{n\times n}$ if and only if one of the following conditions holds:

(a) $A + D$ is nonsingular for every diagonal $D \in \mathbb{R}_+^{n\times n}$.
(b) A is inverse-positive, i.e., $\exists A^{-1} \in \mathbb{R}_+^{n\times n}$.
(c) A is monotone, i.e., $A\mathbf{x} \geq 0 \Rightarrow \mathbf{x} \geq 0, \forall \mathbf{x} \in \mathbb{R}^n$.
(d) Every regular splitting of A is convergent, i.e., if $A = M - N$ with $M^{-1}, N \in \mathbb{R}_+^{n\times n}$, then $\rho(M^{-1}N) < 1$.
(e) A is positive stable, i.e., $\underline{\sigma}(A) > 0$.
(f) $\exists \mathbf{x} > 0$ with $A\mathbf{x} \geq 0$ such that if $[A\mathbf{x}]_{i_0} = 0$, then $\exists i_1, ..., i_r$ with $[A\mathbf{x}]_{i_r} > 0$ and $a_{i_k, i_{k+1}} \neq 0, \forall k \in [0, r-1]$.
(g) $\exists \mathbf{x} > \mathbf{0}$ with $A\mathbf{x} > \mathbf{0}$.

The next result is a direct consequence of [19, Thm. 2].

Lemma 2. Let $A \in \mathbb{M}^{n\times n}$ be irreducible. Then

(i) $\texttt{diag}(\mathbf{z}) + A \in \mathbb{M}_+^{n\times n}$ for every $\mathbf{z} \in \mathbb{R}_+^n \setminus \{\mathbf{0}\}$.
(ii) $\big[\big(\texttt{diag}(\mathbf{z}) + A\big)^{-1}\big]_{i,j}$ is a convex and decreasing function in $\mathbf{z} \in \mathbb{R}_+^n$ for all $1 \leq i, j \leq n$.

3 Model and Formulation

Consider a large system consisting of N systems that depend on each other for their function, and denote the set of comprising systems by $\mathcal{A} := \{1, 2, \ldots, N\}$. The security of the systems is interdependent in that the failure or infection of a system can cause that of other systems.[1] As stated before, we study the problem of determining security investments for hardening each system in order to defend the systems against attacks. The goal of the system operator is to minimize the average aggregate costs for all systems (per unit time), which account for both security investments and any economic losses from failures/infections of systems.

1. Throughout the paper, we use the words 'failure' and 'infection' interchangeably, in order to indicate that a system fell victim to an attack.

3.1 Setup

We assume that each system experiences primary attacks from malicious actors. Primary attacks on system $i \in \mathcal{A}$ occur in accordance with a Poisson process with rate $\lambda_i \in \mathbb{R}_+$. When a system experiences an attack, it suffers an infection and subsequent economic losses with some probability, called *breach probability*.

This breach probability 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 $q_i : \mathbb{R}_+ \to (0, 1]$: 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 has been shown [2] that, under some conditions, the breach probability is decreasing and log-convex.

When system i falls victim to an attack and becomes infected, the operator incurs costs c_i^r per unit time for recovery (e.g., inspection and repair of servers). Recovery times are modeled using independent and identically distributed (i.i.d.) exponential random variables with parameter $\delta_i > 0$. Besides recovery costs, the infection of system i may 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 this, we assume that the infection of system i introduces economic losses of c_i^e per unit time.

Besides primary attacks, systems also experience secondary attacks from other infected systems. For example, 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 $i \in \mathcal{A}$.

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 $\beta_{i,j} > 0$. We assume that matrix B is irreducible. Note that this is equivalent to assuming that the graph $\mathcal{G}$ is strongly connected.

3.2 Model

We adopt the well-known susceptible-infected-susceptible (SIS) model to capture the evolution of system state. Let $p_i(t)$ be the probability that system i is at the 'infected' state (I) at time $t \in \mathbb{R}_+$. We approximate the dynamics of $\mathbf{p}(t) := (p_i(t) : i \in \mathcal{A})$, $t \in \mathbb{R}_+$, using the following (Markov) differential equations, which are derived in [25] and are based on mean field approximation. This model is also similar to those employed in [9], [21], [27], [29], [33], [34]: for fixed security investments, $\mathbf{s} = (s_i : i \in \mathcal{A}) \in \mathbb{R}_+^N$,

$$\dot{p}_i(t) = (1 - p_i(t))q_i(s_i)\Big(\lambda_i + \sum_{j\in\mathcal{A}} \beta_{j,i}p_j(t)\Big) - \delta_i p_i(t). \quad (1)$$

In practice, the breach probability q_i can be a complicated function of the security investment. Here, in order to make progress, we assume that the breach probability functions can be approximated (in the regime of interest) using a function of the form $q_i(s) = (1+\kappa_i s)^{-1}$ for all $i \in \mathcal{A}$. The parameter $\kappa_i > 0$ models how quickly the breach probability decreases with security investment for system i. The assumed function satisfies log-convexity shown in [2].

Define $\alpha_i := \kappa_i \delta_i$, $i \in \mathcal{A}$, and $\boldsymbol{\alpha} := (\alpha_i : i \in \mathcal{A})$. The following theorem tells us that, for a fixed security investment vector $\mathbf{s} := (s_i : i \in \mathcal{A}) \in \mathbb{R}_+^N$, there is a unique equilibrium of the differential system described by (1). Due to a space constraint, the proofs of some of our main results are omitted here and can be found in [23].

Theorem 1. *Suppose $\boldsymbol{\lambda} \gneq \mathbf{0}$, $\boldsymbol{\delta} > \mathbf{0}$ and $\mathbf{s} \geq \mathbf{0}$ are fixed. If the network is strongly connected, i.e., B is irreducible, there exists a unique equilibrium $\mathbf{p}^* \in (0,1)^N$ of (1). Moreover, starting with any $\mathbf{p}_0$ satisfying $\mathbf{p}^* \leq \mathbf{p}_0 \leq \mathbf{1}$, the iteration*

$$\mathbf{p}_{k+1} = \frac{\boldsymbol{\lambda} + B\mathbf{p}_k}{\boldsymbol{\lambda} + B\mathbf{p}_k + \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta}}, \; k \in \mathbb{N}, \tag{2}$$

converges linearly to $\mathbf{p}^*$ with some rate $\rho_0 < 1 - \min_{i\in\mathcal{A}} p_i^*$.

Proof. Please see Appendix A of [23] for a proof. □

Note that the unique equilibrium of the differential system given by (1) specifies the probability that each system will be infected at steady state. For this reason, we take the average cost of the system, denoted by $C_{\text{avg}}(\mathbf{s})$, to be

$$C_{\text{avg}}(\mathbf{s}) := w(\mathbf{s}) + \textstyle\sum_{i\in\mathcal{A}} c_i p_i^*(\mathbf{s}) = w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}^*(\mathbf{s}), \tag{3}$$

where $c_i := c_i^r + c_i^e$, $\mathbf{c} = (c_i : i \in \mathcal{A})$, and $w(\mathbf{s})$ quantifies the security investment costs (per unit time), e.g., $w(\mathbf{s}) = \sum_{i\in\mathcal{A}} s_i$. We assume that w is continuous, (weakly) convex and strictly increasing, and refer to $\mathbf{c}$ simply as the infection costs (instead of infection costs per unit time).

A major difficulty in minimizing the average cost in (3) as an objective function is that the equilibrium $\mathbf{p}^*(\mathbf{s})$ does not have a closed-form expression. As a result, we cannot simply substitute a closed-form expression for the equilibrium $\mathbf{p}^*(\mathbf{s})$ in (3) and minimize the average cost with $\mathbf{s}$ as the optimization variables. For this reason, we formulate the problem of determining optimal security investments that minimize the average cost $C_{\text{avg}}(\mathbf{s})$ as follows:

$$\begin{aligned} (\text{P}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{p}\geq\mathbf{0}} \quad & f(\mathbf{s},\mathbf{p}) := w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p} && \text{(4a)} \\ \text{s.t.} \quad & \mathbf{g}(\mathbf{s},\mathbf{p}) = \mathbf{0} && \text{(4b)} \end{aligned}$$

where $\mathbf{g}(\mathbf{s},\mathbf{p}) = (g_i(\mathbf{s},\mathbf{p}) : i \in \mathcal{A})$, and

$$g_i(\mathbf{s},\mathbf{p}) = (1-p_i)\Big(\lambda_i + \sum_{j\in\mathcal{A}} \beta_{j,i} p_j\Big) - (\alpha_i s_i + \delta_i) p_i, \; i \in \mathcal{A}.$$

Recall that, for given $\mathbf{s} \in \mathbb{R}_+^N$, only the unique equilibrium $\mathbf{p}^* \in (0,1)^N$ in Theorem 1 satisfies the constraint in (4b). Clearly, the solution to problem (P) will also shed light on which systems are more critical from the security perspective and, hence, should be protected. In the problem (P), we do not explicitly model any total budget constraint on security investments for simplicity of exposition. However, we will revisit the issue of constraints on security investments, such as a total budget constraint, and discuss how it affects our main results in Section 8.1

This problem (P) is nonconvex due to the nonconvexity of the equality constraint functions in (4b). In particular, g_i contains both quadratic or bilinear terms $p_i p_j$ and $p_i s_i$. In the following sections, we develop four complementary algorithms for finding good-quality solutions to the nonconvex problem: the first two approaches are based on convex relaxations using different techniques, and provide a lower bound on the optimal value of the problem (P). The last two are designed to find a local minimizer of the problem (P), hence provide an upper bound on the optimal value, and are based on the RGM and SCP.

4 Lower Bounds via Convex Relaxations

In this section, we discuss how we can relax the original problem (P) and construct two different convex formulations, which can be used to obtain (a) a lower bound on the optimal value and (b) a feasible solution to (P) using optimal points of the relaxed problems. Furthermore, we provide a sufficient condition for the relaxed problems to be exact, i.e., their optimal point is also an optimal point of the nonconvex problem (P). The first approach is based on M-matrix theory and the preliminary results were reported in [22]. The second approach is designed to deal with some of computational issues of the first approach. Moreover, as we will show, the optimal point of the first approach can be computed from that of the second approach.

4.1 Convex Relaxation: M-Matrix Theory

Given $\boldsymbol{\lambda} \gneq \mathbf{0}$ and irreducible B, Theorem 1 states that the unique equilibrium of (1) which satisfies (4b) is strictly positive. Hence, we can rewrite the constraints in (4b) as

$$(\mathbf{p}^{-1} - \mathbf{1}) \circ (\boldsymbol{\lambda} + B\mathbf{p}) = \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta}, \tag{5}$$

where $\mathbf{p}^{-1} = (p_i^{-1} : i \in \mathcal{A})$. By introducing a new variable

$$\mathbf{z} := \mathbf{p}^{-1} \circ (\boldsymbol{\lambda} + B\mathbf{p}), \tag{6}$$

the constraint in (5) can be rewritten as

$$\mathbf{z} = \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}. \tag{7}$$

Note that (7) is affine in $\mathbf{z}$, $\mathbf{s}$ and $\mathbf{p}$, and the nonconvexity in the equality constraint functions (mentioned at the end of the previous section) is now captured by $\mathbf{z}$, which from (6) can be expressed as

$$(\texttt{diag}(\mathbf{z}) - B)\mathbf{p} = \boldsymbol{\lambda} \gneq \mathbf{0}. \tag{8}$$

We can show that the matrix $(\texttt{diag}(\mathbf{z}) - B)$ is a nonsingular M-matrix and, hence, $\mathbf{p} = (\texttt{diag}(\mathbf{z}) - B)^{-1}\boldsymbol{\lambda}$ as follows: from (8), since $\boldsymbol{\lambda} \gneq \mathbf{0}$, we have $\lambda_{i^*} > 0$ for some i^*. Since matrix B is assumed irreducible, for any j such that $\lambda_j = 0$, we can find a finite sequence $(i_0 = j, i_1, i_2, \ldots, i_r = i^*)$ such that $[(\texttt{diag}(\mathbf{z}) - B)\mathbf{p}]_{i^*} = \lambda_{i^*} > 0$ and $(\texttt{diag}(\mathbf{z}) - B)_{i_k,i_{k+1}} = -B_{i_k,i_{k+1}} \neq 0$ for all $k \in \{0,1,\ldots,r-1\}$. Because $\mathbf{p} > 0$, Lemma 1-(f) tells us that this is equivalent

to matrix $(\mathtt{diag}(\mathbf{z}) - B)$ being a nonsingular M-matrix. As a result, the original problem (P) can be reformulated as

$$\begin{aligned}(\mathrm{P2}) \qquad \min_{\mathbf{s},\mathbf{p},\mathbf{z}} \quad & f(\mathbf{s},\mathbf{p}) \\ \text{s.t.} \quad & \mathbf{p} = (\mathtt{diag}(\mathbf{z}) - B)^{-1}\boldsymbol{\lambda} \\ & \mathbf{z} = \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p} \\ & \mathbf{s} \in \mathbb{R}_+^N, \quad \mathbf{p} \in \mathbb{R}_+^N, \quad \mathbf{z} \in \Omega,\end{aligned}$$

where

$$\Omega := \{\mathbf{z} \in \mathbb{R}_+^N \mid \mathtt{diag}(\mathbf{z}) - B \in \mathbb{M}_+^{N\times N}\}. \tag{9}$$

We can show that the set Ω in (9) is convex. This is proved in [23, Appendix B]. Also, it follows from Lemma 2 that for any $1 \le i, j \le N$, the element $\left[\left(\mathtt{diag}(\mathbf{z}) - B\right)^{-1}\right]_{i,j}$ is convex and (element-wise) decreasing in $\mathbf{z} \in \Omega$. For these reasons, we obtain the following convex relaxation of (P2).

$$\begin{aligned}(\mathrm{P_{R1}}) \qquad \min_{\mathbf{s},\mathbf{p},\mathbf{z}} \quad & f(\mathbf{s},\mathbf{p}) \\ \text{s.t.} \quad & \mathbf{p} \ge (\mathtt{diag}(\mathbf{z}) - B)^{-1}\boldsymbol{\lambda} && \text{(10a)} \\ & \mathbf{z} = \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p} && \text{(10b)} \\ & \mathbf{s} \in \mathbb{R}_+^N, \quad \mathbf{p} \le \mathbf{1}, \quad \mathbf{z} \in \Omega.\end{aligned}$$

This convex relaxation can be solved by numerical convex solvers to provide a lower bound on the optimal value of (P). Also, as shown in the following theorem, its optimal point also leads to a feasible point for problem (P).

Theorem 2. *Let* $\mathbf{x}_{\mathrm{R}}^* := (\mathbf{s}_{\mathrm{R}}^*, \mathbf{p}_{\mathrm{R}}^*, \mathbf{z}_{\mathrm{R}}^*)$ *denote an optimal point of* $(\mathrm{P_{R1}})$ *and* f^* *the optimal value of* (P)*. Then, we have*

$$f(\mathbf{s}_{\mathrm{R}}^*, \mathbf{p}_{\mathrm{R}}^*) \le f^* \le f(\tilde{\mathbf{s}}(\mathbf{x}_{\mathrm{R}}^*), \tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*)),$$

where $(\tilde{\mathbf{s}}(\mathbf{x}_{\mathrm{R}}^*), \tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*))$ *is a feasible point of problem* (P) *given by*

$$\begin{aligned}&\tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*) = (\mathtt{diag}(\mathbf{z}_{\mathrm{R}}^*) - B)^{-1}\boldsymbol{\lambda} \text{ and} \\ &\tilde{\mathbf{s}}(\mathbf{x}_{\mathrm{R}}^*) = \mathbf{s}_{\mathrm{R}}^* + \mathtt{diag}(\boldsymbol{\alpha}^{-1})B(\mathbf{p}_{\mathrm{R}}^* - \tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*)).\end{aligned}$$

Proof. The first inequality is obvious because $(\mathrm{P_{R1}})$ is a convex relaxation of (P). For the second inequality, note that $(\tilde{\mathbf{s}}(\mathbf{x}_{\mathrm{R}}^*), \tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*), \mathbf{z}_{\mathrm{R}}^*)$ is a feasible point for $(\mathrm{P_{R1}})$. Also, it satisfies (10a) with equality. Thus, it is a feasible point for problem (P), proving the second inequality. □

Clearly, $\mathbf{x}_{\mathrm{R}}^*$ solves (P) if the inequality constraints in (10a) are all active at $\mathbf{x}_{\mathrm{R}}^*$, which means $f(\mathbf{s}_{\mathrm{R}}^*, \mathbf{p}_{\mathrm{R}}^*) = f(\tilde{\mathbf{s}}(\mathbf{x}_{\mathrm{R}}^*), \tilde{\mathbf{p}}(\mathbf{x}_{\mathrm{R}}^*))$. Based on this, we can provide a following sufficient condition for convex relaxation $(\mathrm{P_{R1}})$ to be exact.[2]

Lemma 3. The above convex relaxation $(\mathrm{P_{R1}})$ is exact if

$$B^{\mathsf{T}}\mathtt{diag}(\boldsymbol{\alpha}^{-1})\nabla w(\mathbf{s}) \le \mathbf{c} \text{ for all } \mathbf{s} \ge \mathbf{0}. \tag{11}$$

Proof. Suppose $(\tilde{\mathbf{s}}, \tilde{\mathbf{p}})$ is the feasible point of (P) given in Theorem 2. Since w is convex, we have $w(\tilde{\mathbf{s}}) - w(\mathbf{s}_{\mathrm{R}}^*) \le \nabla w(\tilde{\mathbf{s}})^{\mathsf{T}}(\tilde{\mathbf{s}} - \mathbf{s}_{\mathrm{R}}^*) = \nabla w(\tilde{\mathbf{s}})^{\mathsf{T}}\mathtt{diag}(\boldsymbol{\alpha}^{-1})B(\mathbf{p}_{\mathrm{R}}^* - \tilde{\mathbf{p}})$. From this inequality, the gap $f(\tilde{\mathbf{s}}, \tilde{\mathbf{p}}) - f(\mathbf{s}_{\mathrm{R}}^*, \mathbf{p}_{\mathrm{R}}^*) \le (\nabla w(\tilde{\mathbf{s}})^{\mathsf{T}}\mathtt{diag}(\boldsymbol{\alpha}^{-1})B - \mathbf{c}^{\mathsf{T}})(\mathbf{p}_{\mathrm{R}}^* - \tilde{\mathbf{p}})$. Under condition (11), together with $\mathbf{p}_{\mathrm{R}}^* \ge \tilde{\mathbf{p}}$, this gap is nonpositive. By Theorem 2, this gap must be zero. Thus, $(\mathrm{P_{R1}})$ is exact. □

2. Here, the convex relaxation $(\mathrm{P_{R1}})$ is exact if (P) and $(\mathrm{P_{R1}})$ have the same optimal value and a solution of one problem can be obtained from that of the other problem.

Remark 1. (*Sufficient condition for exact relaxation*) First, roughly speaking, the condition in (11) means that when the infection costs $\mathbf{c}$ are sufficiently high, the convex relaxation $(\mathrm{P_{R1}})$ is exact and we can find optimal security investments, i.e., a solution to (P), by solving $(\mathrm{P_{R1}})$ instead. The intuition behind this observation is the following: as $\mathbf{c}$ becomes larger, the second term in the objective function, namely $\mathbf{c}^{\mathsf{T}}\mathbf{p}$, becomes more important and an optimal point tries to suppress it by reducing $\mathbf{p}$. However, since $\mathbf{p}$ must satisfy the inequality in (10a), it can only be reduced till the equality holds, which satisfies the constraint in problem (P2). Second, condition (11) can be verified prior to solving the relaxed problem. This can be done easily if w is a linear function or an upper bound on the gradient ∇w is known. Finally, even when the convex relaxation is not exact, $(\tilde{\mathbf{s}}, \tilde{\mathbf{p}})$ can still be used as a good initial point for a local search algorithm, such as the RGM developed in Section 5 below.

Remark 2. (*Numerical issues of* $(\mathrm{P_{R1}})$) Although $(\mathrm{P_{R1}})$ is a convex problem, there are a few numerical challenges. First, the Jacobian of constraint functions in (10a), which involves the derivative of inverse matrix $(\mathtt{diag}(\mathbf{z}) - B)^{-1}$, tends to be dense even when B is sparse. Thus, off-the-shelf convex solvers may not be suitable for large systems.

Second, although the constraint set Ω for $\mathbf{z}$ (defined in (9)) is convex, it is not numerically easy to handle, especially for large networks. This is because Ω is not closed and $(\mathrm{P_{R1}})$ becomes invalid outside Ω. Thus, a numerical algorithm ought to stay inside Ω and, for this reason, the nonsingularity of the M-matrix, $\mathrm{diag}(\mathbf{z}) - B$, should be ensured at every step. In general, it takes $O(N^3)$ to check if the matrix satisfies this condition [31]. The following approach can, however, alleviate the computational burden.

- s1 Starting at some $\mathbf{z}_0 \in \Omega$, solve $(\mathrm{P_{R1}})$ only with the constraint $\mathbf{z} \in \mathbb{R}_+^N$. Then, check if the obtained solution $\mathbf{x}_{\mathrm{R}}^*$ satisfies $\mathbf{z}_{\mathrm{R}}^* \in \Omega$, If so, $\mathbf{x}_{\mathrm{R}}^*$ solves $(\mathrm{P_{R1}})$. Otherwise, go to step s2.
- s2 Choose a simpler subset $\tilde{\Omega} \subset \Omega$ and solve $(\mathrm{P_{R1}})$ subject to a stricter constraint $\mathbf{z} \in \tilde{\Omega}$. If $\mathbf{z}_{\mathrm{R}}^*$ in $\mathbf{x}_{\mathrm{R}}^*$ lies in $\mathrm{int}(\tilde{\Omega})$, the solution is optimal for $(\mathrm{P_{R1}})$; otherwise, construct a new $\tilde{\Omega}$ so that $\mathbf{z}_{\mathrm{R}}^*$ belongs to the interior of new $\tilde{\Omega}$ and repeat. Below, we propose an efficient way to choose the subset $\tilde{\Omega}$ that is more suitable for numerical algorithms.

4.1.1 Construction of Convex Subsets of Ω

A key observation to constructing suitable subsets of Ω is that, in view of Lemmas 1 and 2, Ω can be expressed as

$$\Omega = \textstyle\bigcup_{\underline{\mathbf{z}} \in \partial\Omega} \{\mathbf{z} \in \mathbb{R}_+^N \mid \mathbf{z} \gtrsim \underline{\mathbf{z}}\}.$$

Thus, for every $\underline{\mathbf{z}} \in \partial\Omega$, $\tilde{\Omega}(\underline{\mathbf{z}}) := \{\mathbf{z} \in \mathbb{R}_+^N \mid \mathbf{z} \gtrsim \underline{\mathbf{z}}\} \subset \Omega$. Our goal is to find some $\check{\mathbf{z}} \in \partial\Omega$ such that an optimal point $\mathbf{x}_{\mathrm{R}}^*$ that solves the relaxed problem with Ω replaced by $\tilde{\Omega}(\check{\mathbf{z}})$, satisfies $\mathbf{z}_{\mathrm{R}}^* \in \mathrm{int}\,\tilde{\Omega}(\check{\mathbf{z}})$. Below, we provide several possible choices for $\underline{\mathbf{z}}$ with increasing computational complexity.

4.1.1.1 Diagonal dominance: Note that the matrix $\mathtt{diag}(\mathbf{z}) - B$ is nonsingular if it is strictly diagonally dominant. This can be guaranteed by choosing $\underline{\mathbf{z}} > B\mathbf{1}$, where the lower bound $B\mathbf{1}$ represents the total rate of infection from immediate neighbors in the graph $\mathcal{G}$. From (10b), a

trivial sufficient condition is $\boldsymbol{\delta} + \boldsymbol{\lambda} \geq B\mathbf{1}$. But, we observe empirically that this often leads to suboptimal solutions.

4.1.1.2 Dominant eigenvalue: Another straightforward lower bound is given by $\underline{\mathbf{z}} > \rho(B)\mathbf{1}$. Recall that the spectral radius $\rho(B)$ is also an eigenvalue of B and equal to $\bar{\sigma}(B)$, which can be computed efficiently using, for example, the power method.

4.1.1.3 Iterative dominant eigenvalue selection via matrix balancing: Unfortunately, we observe empirically that a static selection of the subset $\tilde{\Omega}$ does not always lead to a good solution and a following iterative algorithm yields better performance: let $\mathbf{h} > \mathbf{0}$ be a normal vector of the plane tangent to the closure of Ω at some $\underline{\mathbf{z}} \in \partial\Omega$ such that

$$\begin{aligned}\underline{\mathbf{z}} &= \arg\min\nolimits_{\mathbf{z}\in\mathbb{R}^N_+} \quad \{\mathbf{h}^\mathsf{T}\mathbf{z} \mid \mathbf{z} \in \mathrm{cl}(\Omega)\} \\ &= \arg\min\nolimits_{\mathbf{z}\in\mathbb{R}^N_+} \quad \{\mathbf{h}^\mathsf{T}\mathbf{z} \mid \underline{\sigma}\big(\mathtt{diag}(\mathbf{z}) - B\big) = 0\}, \quad (12)\end{aligned}$$

where the second equality follows from the fact that we are minimizing a linear function over a closed convex set. The minimization in (12) amounts to finding the smallest diagonal perturbation $\mathbf{z}$ (in 1-norm weighted by $\mathbf{h}$) so that B becomes (negative) stable. In [23, Appendix C], we show that this is in fact a *matrix balancing* problem, for which efficient algorithms exist (see [5], [28] for nearly-linear time centralized algorithms and [20], [21] for distributed algorithms with geometric convergence).

Algorithm 1: Algorithm for Convex Relaxation $(\mathrm{P_{R1}})$

init: $t = 0$, $\bar{h} > 1$, $\underline{\mathbf{z}}^{(0)}$ from (12)
while *stopping cond. not met* **do**
  $(\tilde{\mathbf{s}}_\mathrm{R}^{(t+1)}, \tilde{\mathbf{p}}_\mathrm{R}^{(t+1)}, \tilde{\mathbf{z}}_\mathrm{R}^{(t+1)}) \leftarrow$ solve $(\mathrm{P_R}) : \mathbf{z} \in \tilde{\Omega}(\underline{\mathbf{z}}^{(t)})$
  $\mathcal{I}_{ac} \leftarrow \{i \in \mathcal{A} \mid [\tilde{\mathbf{z}}_\mathrm{R}^{(t+1)}]_i = [\underline{\mathbf{z}}^{(t)}]_i\}$
  if $\mathcal{I}_{ac} = \varnothing$ **then**
    break
  $\mathbf{h}^+ \leftarrow (h_i^+ = 1, i \notin \mathcal{I}_{ac}; h_i^+ = \bar{h}, i \in \mathcal{I}_{ac})$
  $d \leftarrow \underline{\sigma}\big(\mathtt{diag}(\mathbf{h}^+)^{-1}(\mathtt{diag}(\tilde{\mathbf{z}}_\mathrm{R}^{(t+1)}) - B)\big)$
  $\underline{\mathbf{z}}^{(t+1)} \leftarrow \tilde{\mathbf{z}}_\mathrm{R}^{(t+1)} - d\mathbf{h}^+$
  $t \leftarrow t + 1$

Our first proposed algorithm (Algorithm 1) is based on the discussion in this subsection. Initially, we choose some $\bar{h} > 1$ and $\mathbf{h} = \boldsymbol{\alpha}^{-1} \circ \nabla w(\mathbf{s}_0)$, where $\mathbf{s}_0$ is the initial choice of security investments. This heuristic is based on the relaxed problem by weighting only the investment cost $w(\mathbf{s})$ without considering $\mathbf{p}$. Subsequent iterations are based on dominant eigenvalues with varying weights determined by $\mathbf{h}^+$, which reflects active constraints of $\tilde{\mathbf{z}}_\mathrm{R}$ (of the current solution). Since $\tilde{\mathbf{z}}_\mathrm{R} \in \Omega$, we have $\underline{\sigma}\big(\mathtt{diag}(\tilde{\mathbf{z}}_\mathrm{R}) - B\big) > 0$. Thus, we can construct a new subset $\tilde{\Omega}(\underline{\mathbf{z}})$ by translating the set $\{\mathbf{z} \geq \tilde{\mathbf{z}}_\mathrm{R}\}$ towards the boundary $\partial\Omega$ in the direction of $\mathbf{h}^+$, so that $\tilde{\mathbf{z}}_\mathrm{R}$ lies in the interior of new $\tilde{\Omega}(\underline{\mathbf{z}})$. In our numerical studies (Section 7), we use $\bar{h} = 10$.

Note that Algorithm 1 is guaranteed to converge because the problem is convex and the objective function value decreases after each iteration. Although we cannot provide a convergence rate, numerical studies in Section 7 show that only a few iterations are needed in most cases.

4.2 Convex Relaxation Based on Exponential Cones

As explained in the previous subsection, a possible difficulty in solving the convex relaxation in $(\mathrm{P_{R1}})$ is taking into account two constraints – constraint in (10a) and $\mathbf{z} \in \Omega$. Here, we present an alternative convex relaxation of the original problem, which avoids these issues by introducing auxiliary optimization variables and relaxing the equality constraint in (4b) without the need for the constraint set Ω.

First, recall from the previous subsection that the constraint in (4b) can be rewritten as

$$\mathbf{p}^{-1} \circ \boldsymbol{\lambda} + \mathbf{p}^{-1} \circ B\mathbf{p} = \boldsymbol{\lambda} + B\mathbf{p} + \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta}. \quad (5)$$

Since any solution must satisfy $\mathbf{p} \in (0, 1]^N$, we introduce following auxiliary variables and rewrite the equality constraint in (5): for fixed $\mathbf{y} \in \mathbb{R}^N_+$, define

$$\mathbf{p} := e^{-\mathbf{y}},\ \mathbf{t} := \boldsymbol{\lambda} \circ e^{\mathbf{y}},\ U := \mathtt{diag}(e^{\mathbf{y}})B\mathtt{diag}(e^{-\mathbf{y}}). \quad (13)$$

Using these new variables, (5) can be rewritten as follows.

$$\mathbf{t} + U\mathbf{1} = \boldsymbol{\lambda} + B\mathbf{p} + \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} \quad (14)$$

Then, problem (P) is equivalent to the following problem.

$$(\mathrm{P3}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{p}\geq\mathbf{0},\mathbf{y},\mathbf{t},U} \quad f(\mathbf{s},\mathbf{p}) = w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}$$
$$\text{s.t.} \quad (13), (14)$$

The equivalent problem $(\mathrm{P3})$ is still nonconvex due to the constraints in (13). We can relax these equality constraints with the following inequality convex constraints.

$$\mathbf{1} \geq \mathbf{p} \geq e^{-\mathbf{y}},\ \mathbf{t} \geq \boldsymbol{\lambda} \circ e^{\mathbf{y}},\ U \geq \mathtt{diag}(e^{\mathbf{y}})B\mathtt{diag}(e^{-\mathbf{y}}) \quad (15)$$

This leads to the following second convex relaxation.

$$(\mathrm{P_{R2}}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{p},\mathbf{y}\geq\mathbf{0},\mathbf{t},U} \quad f(\mathbf{s},\mathbf{p}) = w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}$$
$$\text{s.t.} \quad (14), (15)$$

We can express the constraints in (15) as a following set of at most $2N + m$ exponential cone constraints:

$$(p_i, 1, -y_i) \in \mathcal{K}_\mathrm{exp} \quad \text{for all } i \in \mathcal{A} \quad (16a)$$
$$(t_i, 1, y_i + \log\lambda_i) \in \mathcal{K}_\mathrm{exp} \quad \text{for all } i \in \Psi_{\boldsymbol{\lambda}} \quad (16b)$$
$$(u_{ij}, 1, y_i - y_j + \log b_{ij}) \in \mathcal{K}_\mathrm{exp} \quad \text{for all } (i,j) \in \mathcal{E} \quad (16c)$$

where $\mathcal{K}_\mathrm{exp} := \mathrm{cl}(\{(x_1, x_2, x_3) \,|\, x_1 \geq x_2 e^{x_3/x_2}, x_2 > 0\})$, and $\Psi_{\boldsymbol{\lambda}} := \{i \in \mathcal{A} \mid \lambda_i > 0\}$. These constraints can be handled efficiently by conic optimization solvers, e.g., MOSEK [1].

Remark 3. We demonstrate below that, somewhat surprisingly, the convex relaxations in $(\mathrm{P_{R1}})$ and $(\mathrm{P_{R2}})$ are in fact equivalent. Moreover, although one may suspect that the size of $(\mathrm{P_{R2}})$ with $4N + m$ variables and $3N + m$ constraints is much larger than the size of $(\mathrm{P_{R1}})$, the constraints of $(\mathrm{P_{R2}})$ are much easier to handle numerically. We will provide numerical results to illustrate this in Section 7.

Analogously to Theorem 2, the following theorem tells us how to find a feasible point of the problem (P), using an optimal point of problem $(\mathrm{P_{R2}})$. In addition, it asserts that the two convex relaxations $(\mathrm{P_{R1}})$ and $(\mathrm{P_{R2}})$ are equivalent in that their optimal values coincide and we can find an optimal point of $(\mathrm{P_{R1}})$ from an optimal point of $(\mathrm{P_{R2}})$.

Theorem 3. *Suppose* $\mathbf{x}_{\mathrm{R}}^{+} := (\mathbf{s}^{+}, \mathbf{p}^{+}, \mathbf{y}^{+}, \mathbf{t}^{+}, U^{+})$ *is an optimal point of* $(\mathrm{P_{R2}})$*. Then, we have*

$$f(\mathbf{s}^{+}, \mathbf{p}^{+}) \leq f^{*} \leq f(\mathbf{s}', \mathbf{p}'), \tag{17}$$

where f^{*} *and* $(\mathbf{s}', \mathbf{p}')$ *are the optimal value and a feasible point, respectively, of the original problem (P) with*

$$\mathbf{p}' = e^{-\mathbf{y}^{+}} \;\; \textit{and} \;\; \mathbf{s}' = \mathbf{s}^{+} + \mathtt{diag}(\boldsymbol{\alpha}^{-1}) B (\mathbf{p}^{+} - \mathbf{p}').$$

Moreover, the last two constraints of (15) *are active at* $\mathbf{x}^{+}$*, i.e.,*

$$\mathbf{t}^{+} = \boldsymbol{\lambda} \circ e^{\mathbf{y}^{+}} \textit{ and } U^{+} = \mathtt{diag}(e^{\mathbf{y}^{+}}) B \mathtt{diag}(e^{-\mathbf{y}^{+}}). \tag{18}$$

Finally, $(\mathbf{s}^{+}, \mathbf{p}^{+}, \mathbf{t}^{+} + U^{+}\mathbf{1})$ *is an optimal point of* $(\mathrm{P_{R1}})$*.*

Proof. Please see Appendix D of [23] for a proof. □

As a direct consequence of the theorem, the suboptimality of $(\mathbf{s}', \mathbf{p}')$ can be assessed using the gap $f(\mathbf{s}', \mathbf{p}') - f(\mathbf{s}^{+}, \mathbf{p}^{+})$. Similar to Lemma 3, the condition in (11) provides a sufficient condition for this gap to be zero, i.e., the convex relaxation in $(\mathrm{P_{R2}})$ is exact.

5 Upper bounds on Optimal Value

The previous section described (i) how we can formulate a convex relaxation of problem (P), which provides a lower bound on the optimal value of (P), using two different techniques and (ii) how to find a feasible solution to (P) using an optimal point of a convex relaxation.

Although the convex relaxation $(\mathrm{P_{R1}})$ or $(\mathrm{P_{R2}})$ may be exact under certain conditions, this is not true in general. In addition, $(\mathrm{P_{R1}})$ may not scale well due to the constraint in (10a); see also Remark 2 above and numerical results in Section 7. For these reasons, we also propose efficient algorithms for finding a local minimizer of the nonconvex problem (P) in this section. These algorithms provide an upper bound on the optimal value, which, together with the optimal value of a convex relaxation when available, can be used to offer a bound on the optimality gap.

5.1 Reduced Gradient Method

Among different nonconvex optimization approaches, we first choose the RGM [8], [17] because it is well suited to the problem (P) and, more importantly, is scalable.

5.1.1 Main Algorithm

First, together with Theorem 1, the implicit function theorem tells us that the condition $\mathbf{g}(\mathbf{s}, \mathbf{p}) = \mathbf{0}$ in (4b) defines a continuous mapping $\mathbf{p}^{*} : \mathbf{s} \in \mathbb{R}_{+}^{N} \mapsto \mathbf{p}^{*}(\mathbf{s}) \in (0,1)^{N}$ such that $\mathbf{g}(\mathbf{s}, \mathbf{p}^{*}(\mathbf{s})) = 0$. Thus, problem (P) can be transformed to a reduced problem only with optimization variables $\mathbf{s}$:

$$\min_{\mathbf{s} \in \mathbb{R}_{+}^{N}} \quad F(\mathbf{s}) := w(\mathbf{s}) + \mathbf{c}^{\mathsf{T}} \mathbf{p}^{*}(\mathbf{s}). \tag{19}$$

Suppose that $(\mathbf{s}^{\star}, \mathbf{p}^{\star})$ is a feasible point of (P). Then, the gradient of F at $\mathbf{s}^{\star}$ is equal to

$$\nabla F(\mathbf{s}^{\star}) = \nabla w(\mathbf{s}^{\star}) + J(\mathbf{s}^{\star})^{\mathsf{T}} \mathbf{c},$$

where $J(\mathbf{s}^{\star}) = \left[\partial p_i^{*}(\mathbf{s}^{\star}) / \partial s_j\right]$. This matrix can be computed by totally differentiating $\mathbf{g}(\mathbf{s}, \mathbf{p}^{*}(\mathbf{s})) = \mathbf{0}$ at $\mathbf{s}^{\star}$: the calculation of total derivative yields

$$M(\mathbf{s}^{\star}) J(\mathbf{s}^{\star}) = -\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{p}^{\star}) \tag{20}$$

with $M(\mathbf{s}^{\star}) = \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}^{\star} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{\star}) - \mathtt{diag}(\mathbf{1} - \mathbf{p}^{\star}) B$. The following lemma shows that $M(\mathbf{s}^{\star})$ is nonsingular.

Lemma 4. The matrix $M(\mathbf{s}^{\star})$ is a nonsingular M-matrix.

Proof. First, note that $M(\mathbf{s}^{\star})$ is a Z-matrix. Second, after some algebra, the constraint $\mathbf{g}(\mathbf{s}^{\star}, \mathbf{p}^{\star}) = \mathbf{0}$ is equivalent to $M(\mathbf{s}^{\star})\mathbf{p}^{\star} = \boldsymbol{\lambda} + \mathbf{p}^{\star} \circ (B\mathbf{p}^{\star})$. Since $\mathbf{p}^{\star} > \mathbf{0}$, we have $\boldsymbol{\lambda} + \mathbf{p}^{\star} \circ (B\mathbf{p}^{\star}) > \mathbf{0}$. Thus, Lemma 1-(g) implies that $M(\mathbf{s}^{\star})$ is a nonsingular M-matrix. □

As a result, $J(\mathbf{s}^{\star}) = -M(\mathbf{s}^{\star})^{-1} \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{p}^{\star})$ from (20) and the gradient of F is given by

$$\nabla F(\mathbf{s}^{\star}) = \nabla w(\mathbf{s}^{\star}) - \boldsymbol{\alpha} \circ \mathbf{p}^{\star} \circ \left(\left(M(\mathbf{s}^{\star}) \right)^{-\mathsf{T}} \mathbf{c} \right).$$

Hence, we can apply the (projected) gradient descent method on the reduced problem in (19). For instance, [3, Proposition 2.3.3] shows that this method converges to a stationary point under step sizes $\{\gamma_t\}_{t \geq 0}$ chosen by the Armijo backtracking line search.

Note that, after each update of $\mathbf{s}$ during a search, we need to compute the corresponding $\mathbf{p}$ so that $(\mathbf{s}, \mathbf{p})$ is feasible for the problem (P). As mentioned earlier, this can be done by using the fixed point iteration in (2).

Our proposed algorithm is provided in Algorithm 2.

Algorithm 2: Reduced Gradient Method

1 **init**: $t = 0$, feasible $(\mathbf{s}^{(0)}, \mathbf{p}^{(0)})$
2 **while** *stopping cond. not met* **do**
3 $\quad M^{(t)} \leftarrow \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}^{(t)} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)}) - \mathtt{diag}(\mathbf{1} - \mathbf{p}^{(t)}) B$
4 $\quad \mathbf{u} \leftarrow (M^{(t)})^{-\mathsf{T}} \mathbf{c}$
5 $\quad \gamma_t \leftarrow$ LINE_SEARCH
6 $\quad \mathbf{s}^{(t+1)} \leftarrow \left[\mathbf{s}^{(t)} - \gamma_t \left(\nabla w(\mathbf{s}^{(t)}) - \boldsymbol{\alpha} \circ \mathbf{p}^{(t)} \circ \mathbf{u} \right) \right]_{+}$
7 $\quad \mathbf{p}^{(t+1)} \leftarrow \mathbf{p}^{*}(\mathbf{s}^{(t+1)})$ using (2)
8 $\quad t \leftarrow t + 1$

5.1.2 Computational Complexity and Issues

For large systems, a naive evaluation of the gradient ∇F, which requires the inverse matrix $\left(M(\mathbf{s}^{\star}) \right)^{-\mathsf{T}}$, becomes computationally expensive, if not infeasible. For this reason, we develop an efficient subroutine for computing ∇F. This is possible because our algorithm only requires $\mathbf{u}$ (in line 4 of Algorithm 2), not the matrix $\left(M(\mathbf{s}^{\star}) \right)^{-\mathsf{T}}$.

For fixed $t \in \mathbb{N} := \{0, 1, \ldots\}$, the vector $\mathbf{u}$ is the solution to a set of linear equations $M^{\mathsf{T}}\mathbf{u} = \mathbf{c}$, where the matrix M^{T} tends to be sparse for most real graphs $\mathcal{G}$. Thus, there are several efficient algorithms for solving them. In this paper, we employ the power method: let $M = D - E$, where D and E denote the diagonal part and off-diagonal part of M, respectively. Then, the linear equations are equivalent to $\mathbf{c} = D\mathbf{u} - E^{\mathsf{T}}\mathbf{u}$. Since D is invertible, the following fixed point relation holds:

$$\mathbf{u} = D^{-1} E^{\mathsf{T}} \mathbf{u} + D^{-1} \mathbf{c} =: G(\mathbf{u}) \tag{21}$$

As $M \in \mathbb{M}_{+}^{N \times N}$ (Lemma 4), Lemma 1-(d) tells us that $M = D - E$ is a convergent splitting and the mapping G in (21) is a contraction mapping with coefficient $\rho(D^{-1}E^{\mathsf{T}}) < 1$. Hence, the iteration $\mathbf{u}_{k+1} = G(\mathbf{u}_k)$ converges to the solution $\mathbf{u}$ exponentially fast. Moreover, this iteration is highly scalable because $E = \mathtt{diag}(\mathbf{1} - \mathbf{p}) B$ is sparse, requiring only $O(|\mathcal{E}|)$ memory space and $O(|\mathcal{E}|)$ operations per iteration.

5.2 Sequential Convex Programming Method

In this subsection, we will develop a second efficient algorithm for finding a local minimizer of the original nonconvex problem (P). This will provide another upper bound on the optimal value we can use to provide an optimality gap together with a lower bound from convex relaxations.

Our algorithm is based on SCP applied to the original formulation in (4a)–(4b). To this end, we successively convexify the constraint (4b) using first order approximations. The novelty of our approach is to linearize (only) the terms $p_i p_j$ and then employ either the convexity result in Lemma 2 or the exponential cone formulation as in Section 4.

At each iteration $t \in \mathbb{N}$, we replace the terms $p_i p_j$ with their first order Taylor expansion at $\mathbf{p}^{(t)}$, resulting in the following partially linearized equality constraint functions:

$$\begin{aligned}\mathbf{g}^{(t)}(\mathbf{s},\mathbf{p}) = \boldsymbol{\lambda} + \mathbf{p}^{(t)} \circ (B\mathbf{p}^{(t)}) - \big(\boldsymbol{\lambda} + B\mathbf{p}^{(t)}\big) \circ \mathbf{p} \\ + (\mathbf{1} - \mathbf{p}^{(t)}) \circ (B\mathbf{p}) - (\boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta}) \circ \mathbf{p}\end{aligned} \tag{22}$$

The partial linearization error is equal to $\mathbf{g}^{(t)}(\mathbf{s},\mathbf{p}) - \mathbf{g}(\mathbf{s},\mathbf{p}) = (\mathbf{p}-\mathbf{p}^{(t)}) \circ B(\mathbf{p}-\mathbf{p}^{(t)})$. When $\mathbf{p}$ is close to $\mathbf{p}^{(t)}$, this error will likely be 'small' and we expect the linearization step to be acceptable. This allows us to approximate the constraint (4b) with $\mathbf{g}^{(t)}(\mathbf{s},\mathbf{p}) = \mathbf{0}$, which can be rewritten as

$$\big(L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})\big)\mathbf{p} = \boldsymbol{\lambda}^{(t)}, \tag{23}$$

where $\boldsymbol{\lambda}^{(t)} = \boldsymbol{\lambda} + \mathbf{p}^{(t)} \circ (B\mathbf{p}^{(t)})$, and

$$L^{(t)} = \mathtt{diag}(\boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)}) - \mathtt{diag}(\mathbf{1} - \mathbf{p}^{(t)})B.$$

This gives us the following subproblem we need to solve at each iteration $t \in \mathbb{N}$.

$$(\mathrm{S1}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{p}\geq\mathbf{0}} \quad \big\{w(\mathbf{s}) + \mathbf{c}^{\mathsf{T}}\mathbf{p} \mid (23) \text{ holds}\big\} \tag{24}$$

The solution at the t-th iteration is then used to construct a new constraint for the $(t+1)$-th iteration, and we repeat this procedure until some stopping condition is met. Unfortunately, the problem in (24) is still nonconvex. But, as we show below, under certain conditions, it can be transformed to a convex problem, which can be solved efficiently.

5.2.1 Convex Formulation Based on M-matrix

First, we show that if $(\mathbf{s},\mathbf{p})$ is feasible for the subproblem in (24), then $L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}) \in \mathbb{M}_+^{N\times N}$: note that this is always a Z-matrix. Also, from (22) and (23), any feasible $\mathbf{p}$ for the subproblem must be positive, and given $\mathbf{p}^{(t)} > \mathbf{0}$, we have $\boldsymbol{\lambda}^{(t)} > \mathbf{0}$ from its definition. Because $\mathbf{p}$ and $\boldsymbol{\lambda}^{(t)}$ are positive, together with condition (23), Lemma 1-(g) implies that $L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}) \in \mathbb{M}_+^{N\times N}$. This in turn tells us from Lemma 1-(b) that its inverse exists and is nonnegative. Consequently,

$$\mathbf{p} = \big(L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})\big)^{-1}\boldsymbol{\lambda}^{(t)} > \mathbf{0}. \tag{25}$$

This allows us to reformulate the subproblem in (24) as follows: we replace $\mathbf{p}$ with the above expression in (25) and introduce a new constraint that $\mathbf{s}$ belongs to a feasible set

$$\Omega^{(t)} = \big\{\mathbf{s} \in \mathbb{R}_+^N \mid L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}) \in \mathbb{M}_+^{N\times N}\big\}.$$

Note that $\Omega^{(t)}$ is convex (the proof follows similar arguments in Appendix B of [23]). The partially linearized subproblem in (24) can now be written as

$$(\mathrm{P_L}) \qquad \min_{\mathbf{s}\in\Omega^{(t)}} \quad J(\mathbf{s}) := w(\mathbf{s}) + \zeta(\mathbf{s}), \tag{26}$$

where $\zeta(\mathbf{s}) := \mathbf{c}^{\mathsf{T}}\big(L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})\big)^{-1}\boldsymbol{\lambda}^{(t)}$ is a convex and decreasing function on $\Omega^{(t)}$ in view of Lemma 2. Thus, the problem $(\mathrm{P_L})$ is convex at every iteration t. Note that, when solving $(\mathrm{P_L})$, we need to ensure the constraint $\mathbf{s} \in \Omega^{(t)}$ is satisfied. This can be done in a manner similar to that discussed in subsection 4.1.1.

After computing an optimal point of $(\mathrm{P_L})$ at the t-th iteration, which we denote by $\mathbf{s}^{(t+1)}$, we then find $\mathbf{p}^{(t+1)}$ satisfying $\mathbf{g}(\mathbf{s}^{(t+1)},\mathbf{p}^{(t+1)}) = \mathbf{0}$ (constraint (4b)) using the fixed point iteration in (2). Thus, we obtain a feasible solution $(\mathbf{s}^{(t+1)},\mathbf{p}^{(t+1)})$ to the original problem (P) after each iteration $t \in \mathbb{N}$. The proposed algorithm based on this approach is provided in Algorithm 3 below.

Algorithm 3: Sequential Convex Programming

init: $t = 0$, $\mathbf{p}^{(0)} \in [0,1]^N$
while *stopping cond. not met* **do**
  $\boldsymbol{\lambda}^{(t)} \leftarrow \boldsymbol{\lambda} + \mathbf{p}^{(t)} \circ (B\mathbf{p}^{(t)})$
  $L^{(t)} \leftarrow \mathtt{diag}(\boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)}) - \mathtt{diag}(\mathbf{1} - \mathbf{p}^{(t)})B$
  $\mathbf{s}^{(t+1)} \leftarrow \arg\min_{\mathbf{s}\in\Omega^{(t)}} w(\mathbf{s}) + \mathbf{c}^{\mathsf{T}}\big(L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})\big)^{-1}\boldsymbol{\lambda}^{(t)}$
  $\mathbf{p}^{(t+1)} \leftarrow \mathbf{p}^*(\mathbf{s}^{(t+1)})$ using (2)
  $t \leftarrow t + 1$

Remark 4. (*Complexity*) For small and medium-sized networks, the subproblem can be solved using off-the-shelf numerical convex solvers, e.g., interior point methods. In this paper, we use an interior-point method to solve the subproblem $(\mathrm{P_L})$, which employs the Newton's algorithm on a sequence of equality constrained problems. Since the number of variables is $O(N)$ and the number of constraints is also $O(N)$, the worst case complexity is $O(N^3)$ [5].

For large networks, we take advantage of the fact that we do not need to solve $(\mathrm{P_L})$ exactly at each iteration. Thus, we can use simple approximations of $\Omega^{(t)}$ and employ computationally cheaper methods to solve $(\mathrm{P_L})$. For example, we can follow the same gradient-based approach in [19] for solving the subproblem, where the gradient of $\zeta(\mathbf{s})$ given by

$$\nabla\zeta(\mathbf{s}) = -(S^{-\mathsf{T}}\mathbf{c}) \circ \boldsymbol{\alpha} \circ (S^{-1}\boldsymbol{\lambda}^{(t)}),$$

where $S = L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})$, can be computed efficiently using the power method as explained in subsection 5.1.2.

5.2.2 Convex Formulation Based on Exponential Cones

As discussed above, if $(\mathbf{s},\mathbf{p})$ is feasible for problem (S1), $L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})$ is a nonsingular M-matrix and $\mathbf{p} > \mathbf{0}$; see also (25). Thus, we can introduce a new variable $\mathbf{y}$ satisfying

$$\mathbf{p} = e^{-\mathbf{y}}.$$

Then, (23) becomes $\big(L^{(t)} + \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s})\big)e^{-\mathbf{y}} = \boldsymbol{\lambda}^{(t)}$, which, after left-multiplying both sides by $\mathtt{diag}(e^{\mathbf{y}})$, is equivalent to

$$\boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)} = \mathtt{diag}(e^{\mathbf{y}})B^{(t)}e^{-\mathbf{y}} + \mathtt{diag}(e^{\mathbf{y}})\boldsymbol{\lambda}^{(t)},$$

where $B^{(t)} = \texttt{diag}(\mathbf{1} - \mathbf{p}^{(t)})B$. As a result, the subproblem is equivalent to the following problem.

$$\begin{aligned}(\text{S2}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{y},\mathbf{t},U} \quad & f^{(t)} = w(\mathbf{s}) + \mathbf{c}^\mathsf{T} e^{-\mathbf{y}} \\ \text{s.t.} \quad & \mathbf{t} + U\mathbf{1} = \boldsymbol{\alpha}\circ\mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)} \\ & \mathbf{t} = \boldsymbol{\lambda}^{(t)} \circ e^{\mathbf{y}} \\ & U = \texttt{diag}(e^{\mathbf{y}})B^{(t)}\texttt{diag}(e^{-\mathbf{y}})\end{aligned}$$

A convex relaxation of (S2) can be obtained by replacing the last two equality constraints with inequality constraints.

$$\begin{aligned}(\text{S}_\text{R1}) \qquad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{y},\mathbf{t},U} \quad & f^{(t)} = w(\mathbf{s}) + \mathbf{c}^\mathsf{T} e^{-\mathbf{y}} \\ \text{s.t.} \quad & \mathbf{t} + U\mathbf{1} = \boldsymbol{\alpha}\circ\mathbf{s} + \boldsymbol{\delta} + \boldsymbol{\lambda} + B\mathbf{p}^{(t)} && \text{(27a)} \\ & \mathbf{t} \geq \boldsymbol{\lambda}^{(t)} \circ e^{\mathbf{y}} && \text{(27b)} \\ & U \geq \texttt{diag}(e^{\mathbf{y}})B^{(t)}\texttt{diag}(e^{-\mathbf{y}}) && \text{(27c)}\end{aligned}$$

It turns out that this convex relaxation is always exact.

Theorem 4. *The convex relaxation* (S_R1) *is exact.*

Proof. A proof can be found in Appendix E of [23]. □

We end this subsection by noting that this convex formulation is in fact equivalent to the one based on M-matrix in (26). This can be shown using similar arguments used in the proof of Theorem 3 and is omitted here.

6 Special Case: $\boldsymbol{\lambda} = \mathbf{0}$

In practice, we expect that the systems experience primary attacks infrequently and $\boldsymbol{\lambda}$ is small, and that steady-state infection probabilities are not large. For this reason, we consider a limit case of our problem as $\boldsymbol{\lambda} \to \mathbf{0}$ with diminishing primary attack rates. As we show, studying the special case with $\boldsymbol{\lambda} = \mathbf{0}$ reveals additional insights into the steady-state behavior and provides an approximate upper bound on the system cost when $\boldsymbol{\lambda} \approx \mathbf{0}$, which can be computed easily.

This case reduces to a problem that has been studied by previous works, in which the adjustable curing rate is equal to $\delta_i/q_i(s_i)$ for each $i \in \mathcal{A}$.[3] A key difference between this case and when $\boldsymbol{\lambda} \gneq \mathbf{0}$ is that Theorem 1 cannot be applied to guarantee the uniqueness of an equilibrium because the assumption $\boldsymbol{\lambda} \gneq \mathbf{0}$ is violated. It turns out that this difference has significant effects on our problem, as it will be clear.

6.1 Preliminary

In the absence of primary attacks, if no system is infected at the beginning, obviously they will remain at the state. However, if some systems are infected initially, there are two possible outcomes based on the security investments $\mathbf{s}$.

Case 1: $\rho\big(\text{diag}(\boldsymbol{\alpha}\circ\mathbf{s} + \boldsymbol{\delta})^{-1}B\big) \leq 1$ **–** In this case, the unique (stable) equilibrium of (1) is $\mathbf{p}_\text{se}(\mathbf{s}) = \mathbf{0}$. Thus, as $t \to \infty$, $\mathbf{p}(t) \to \mathbf{0}$ and all systems become free of infection.

Case 2: $\rho\big(\text{diag}(\boldsymbol{\alpha}\circ\mathbf{s} + \boldsymbol{\delta})^{-1}B\big) > 1$ **–** In this case, there are two equilibria of (1) – one stable equilibrium $\mathbf{p}_\text{se}(\mathbf{s}) > \mathbf{0}$ and one unstable equilibrium $\mathbf{0}$: (a) if $\mathbf{p}(0) \neq \mathbf{0}$, although there are no primary attacks, we have $\mathbf{p}(t) \to \mathbf{p}_\text{se}(\mathbf{s})$. As a result, somewhat surprisingly, infections continue to transmit among the systems indefinitely and do not go away; and (b) if $\mathbf{p}(0) = \mathbf{0}$, obviously $\mathbf{p}(t) = \mathbf{0}$ for all $t \in \mathbb{R}_+$.

Based on this observation, we define a function $\mathbf{p}_\text{se} : \mathbb{R}_+^N \to [0,1]^N$, where $\mathbf{p}_\text{se}(\mathbf{s})$ is the aforementioned stable equilibrium of (1) for the given security investment vector $\mathbf{s} \in \mathbb{R}_+^N$. It is shown [23, Appendix I] that $\mathbf{p}_\text{se}$ is a continuous function over $\mathbb{R}_+^N$. This tells us that, if we start with $\mathbf{p}(0) \neq \mathbf{0}$, for any given security investments $\mathbf{s} \geq \mathbf{0}$, our steady-state cost is given by $w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}_\text{se}(\mathbf{s})$. For this reason, we are interested in the following optimization problem.

$$\min_{\mathbf{s}\geq\mathbf{0}} \Big\{ w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}_\text{se}(\mathbf{s}) \Big\} \qquad (28)$$

We denote the optimal value and the optimal set of (28) by f_0^* and $\mathbb{S}_0^*$, respectively. Based on the above discussion, we have the following simple observation.

Theorem 5. *If* $\rho\big(\texttt{diag}(\boldsymbol{\delta})^{-1}B\big) \leq 1$*, then* $\mathbf{s}^* = \mathbf{0}$ *is the optimal point. Otherwise,* $\rho\big(\texttt{diag}(\boldsymbol{\alpha}\circ\mathbf{s}^* + \boldsymbol{\delta})^{-1}B\big) \geq 1$ *for all* $\mathbf{s}^* \in \mathbb{S}_0^*$.

Proof. The theorem follows directly from the above discussion and the monotonicity of the spectral radius of nonnegative matrices [10, Thm 8.1.18]: if $A, B \in \mathbb{R}_+^{n\times n}$ such that $A \geq B$, then $\rho(A) \geq \rho(B)$. □

Remark 5. The theorem rules out the case where $\rho\big(\texttt{diag}(\boldsymbol{\alpha}\circ\mathbf{s}^* + \boldsymbol{\delta})^{-1}B\big) < 1$ for some $\mathbf{s}^* \neq \mathbf{0}$. It suggests that if the recovery rates of all the systems are sufficiently large, no additional investments are needed. Otherwise, at any solution $\mathbf{s}^* \in \mathbb{S}_0^*$, the spectral radius is either (i) at the threshold (of one) or (ii) strictly above the threshold. In case (i), $w(\mathbf{s}^*)$ is also the smallest investment cost to suppress the spread in that $\lim_{t\to\infty}\mathbf{p}(t) = \mathbf{0}$ for all $\mathbf{p}(0)$. We will show in subsection 6.2 below how to compute this minimum investment for suppression, denoted by C^*. On the other hand, case (ii) corresponds to an endemic state, i.e., $\lim_{t\to\infty}\mathbf{p}(t) = \mathbf{p}_\text{se}(\mathbf{s}^*) > \mathbf{0}$ when $\mathbf{p}(0) \neq \mathbf{0}$. In this case, we will demonstrate that we can find upper and lower bounds on the optimal cost using our techniques in Sections 4 and 5.

In order to facilitate our discussion, we introduce the following related optimization problem with a *fictitious* constraint on security investments. The goal of imposing a fictitious budget constraint is not to investigate a problem with a budget constraint; instead, it is used to facilitate the determination of a (nearly) optimal point of (28) as we will show.

$$\min_{\mathbf{s}\geq\mathbf{0}} \quad \Big\{ w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}_\text{se}(\mathbf{s}) \mid w(\mathbf{s}) \leq C \Big\} \qquad (29)$$

We define a function $f_0 : \mathbb{R}_+ \to \mathbb{R}_+$, where $f_0(C)$ is the optimal value of the above optimization problem for a given budget C. Clearly, the function f_0 is continuous and nonincreasing, and $\lim_{C\to\infty} f_0(C) = f_0^*$, i.e., problem (29) reduces to (28) by letting $C \to \infty$.

Suppose

$$\mathbf{s}^* \in \arg\min_{\mathbf{s}\in\mathbb{S}_0^*} w(\mathbf{s}) \;\text{ and }\; w^* := w(\mathbf{s}^*) = \min_{\mathbf{s}\in\mathbb{S}_0^*} w(\mathbf{s}). \qquad (30)$$

Obviously, w^* is the minimum security investments necessary to minimize the total cost in (28), and $\mathbf{s}^*$ is an optimal

3. In the previous studies [21], [25], [29], security investments affect the curing rates rather than the breach probability, i.e., they determine how quickly each system can recover from an infection, but do not change the infection probability of systems.

point with the smallest security investments. Then, because f_0 is nonincreasing, for any $C \geq w^*$, we have

$$f_0^* \leq f_0(C) \leq f_0(w^*) \leq w(\mathbf{s}^*) + \mathbf{c}^\mathsf{T}\mathbf{p}_{\rm se}(\mathbf{s}^*) = f_0^*, \quad (31)$$

which implies $f_0(C) = f_0(w^*) = f_0^*$. On the other hand,

$$f_0^* = f_0(w^*) < f_0(C) \text{ if } C < w^*, \quad (32)$$

where the strict inequality follows from the definition of $\mathbf{s}^*$ in (30); any $\mathbf{s}$ with $w(\mathbf{s}) < w^*$ is not an optimal point and, as a result, we have $f_0^* < w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p}_{\rm se}(\mathbf{s})$.

The inequalities in (31) and (32) tell us the following: increasing the security budget C reduces the total cost $f_0(C)$ while $C \leq w^*$. On the other hand, beyond w^*, increasing the budget will not reduce the cost any more as $f_0(w^*) = f_0^*$.

6.2 Bounds on the Optimal Value of (28)

From the discussion at the beginning of subsection 6.1, it is clear that the spectral radius of the matrix $\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta})^{-1}B$ plays an important role in the dynamics and the determination of a stable equilibrium of (1). For this reason, we find it convenient to define the following problem:

$$\min_{\mathbf{s}\geq\mathbf{0}} \big\{ w(\mathbf{s}) \mid \rho(\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta})^{-1}B) \leq 1 \big\} \quad (33)$$

Let C^* be the optimal value of this optimization problem, which is the aforementioned minimum investments needed for suppression. We show in Appendix H of [23] that this optimization problem can be transformed into a convex (exponential cone) problem and, thus, can be solved efficiently.

The following lemma points out an important fact that we will make use of in the remainder of the section.

Lemma 5. The optimal value C^* of (33) is an upper bound on f_0^*, i.e., $f_0^* \leq C^*$.

Proof. We know that any optimal point $\tilde{\mathbf{s}}_0$ of (33) satisfies $w(\tilde{\mathbf{s}}_0) = C^*$ and $\mathbf{p}_{\rm se}(\tilde{\mathbf{s}}_0) = \mathbf{0}$. Therefore, the total cost achieved by $\tilde{\mathbf{s}}_0$ is equal to $C^* + \mathbf{c}^\mathsf{T}\mathbf{0} = C^* \geq f_0(C^*) \geq f_0^*$. □

From this lemma and the definition of $\mathbf{s}^*$ in (30), we have

$$w^* = w(\mathbf{s}^*) \leq f_0^* \leq C^*. \quad (34)$$

Obviously, this also implies $f_0^* = f_0(C)$ for all $C \geq C^*$.

In a special case when the equalities in (34) hold, we have $f_0^* = w^* = C^*$, and an optimal point of the optimization problem in (33), say $\tilde{\mathbf{s}}_0$, is optimal for the problem in (28) with $\mathbf{p}_{\rm se}(\tilde{\mathbf{s}}_0) = \mathbf{0}$. Also, $\rho\big(\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}^* + \boldsymbol{\delta})^{-1}B\big) = 1$ and the equality holds in the second part of Theorem 5. But, in general these equalities may not hold, in which case we must have $w^* < f_0^* < C^*$ and $\rho\big(\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}^* + \boldsymbol{\delta})^{-1}B\big) > 1$.

Because w^* is unknown beforehand, we cannot determine which case holds. However, if we solve the constrained problem in (29) with $C = C^* - \epsilon$ for small positive ϵ, either (a) the optimal point $\mathbf{s}_C^*$ we obtain is an optimal point of (28) if $w(\mathbf{s}_C^*) < C$ or (b) $C = C^* - \epsilon \leq w^* \leq f_0^* \leq C^*$ if $w(\mathbf{s}_C^*) = C$. Note that in the latter case, we have $C^* - f_0^* \leq \epsilon$, and $\tilde{\mathbf{s}}_0$ is ϵ-suboptimal for (28).

At first glance, one may suspect that solving (29) is as difficult as solving (28) because both share the same nonconvex objective function. However, (29) enjoys a few numerical advantages: first, unlike (28), the constrained problem (29) with $w(\mathbf{s}) \leq C < C^*$ admits only $\mathbf{p}_{\rm se}(\mathbf{s}) > \mathbf{0}$, which can be computed using the fixed point iteration in (2). Second, perhaps more importantly, because the stable equilibrium is guaranteed to be strictly positive, it allows us to employ the approaches in Sections 4 and 5 in order to compute upper and lower bounds on the optimal value $f_0(C)$. In the process of finding these bounds, we also demonstrate an interesting observation that we can bound the gap $C^* - f_0^*$ under certain conditions.

Before we proceed with discussing the bounds, let us remark on the choice of ϵ. Theoretically, we want ϵ as small as possible because it determines the ϵ-suboptimality of $\tilde{\mathbf{s}}_0$ when the constraint is active (case (b) above). In practice, however, ϵ should not be too small because it can cause numerical issues when $C^* = w^*$ and $\mathbf{p}_{\rm se}(\mathbf{s}^*) = \mathbf{0}$ (or $\mathbf{p}_{\rm se}(\mathbf{s}^*) \approx \mathbf{0}$ when $C^* \approx w^*$). This is because our approaches to finding upper and lower bounds on the optimal cost in Sections 4 and 5 rely on the strict condition that $\mathbf{p}_{\rm se}(\mathbf{s}) > \mathbf{0}$.

6.2.1 Lower Bound via Convex Relaxation

Borrowing a similar approach used in subsection 4.2, we can formulate the following convex relaxation of (29):

$$\begin{aligned}
(\mathrm{P_{R3}}) \quad \min_{\mathbf{s}\geq\mathbf{0},\mathbf{p},\mathbf{y}\geq\mathbf{0},U} \quad & f(\mathbf{s},\mathbf{p}) = w(\mathbf{s}) + \mathbf{c}^\mathsf{T}\mathbf{p} && \text{(35a)}\\
\text{s.t.} \quad & w(\mathbf{s}) \leq C \\
& U\mathbf{1} = B\mathbf{p} + \boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta} \\
& \mathbf{p} \geq e^{-\mathbf{y}} && \text{(35b)}\\
& U \geq \mathtt{diag}(e^{\mathbf{y}})B\mathtt{diag}(e^{-\mathbf{y}}) && \text{(35c)}
\end{aligned}$$

We denote the optimal value of $(\mathrm{P_{R3}})$ by $f_L(C)$. The inequalities in (35b) and (35c) can be expressed as exponential cone constraints as done in (16a) and (16c), respectively. Thus, this convex relaxation can be solved efficiently.

Clearly, $f_L(C)$ is nonincreasing on $[0, C^*)$. Also, $f_L(C) \leq f_0(C)$ for all $C \in [0, C^*)$. Let $f_L^* := \lim_{C\to C^*} f_L(C)$. Together with the earlier inequality in (34), we have

$$f_L^* \leq f_0^* \leq C^*. \quad (36)$$

More can be said regarding these bounds as follows.

Theorem 6. *Suppose $C = C^* - \epsilon$ and $(\mathbf{s}_L, \mathbf{p}_L, \mathbf{y}_L, U_L)$ is an optimal point of $(\mathrm{P_{R3}})$. Let $C_L = w(\mathbf{s}_L)$. Then, we have $f_L^* = f_L(C_L) = f_L(C) \leq f_0^*$ if $C_L < C$, and*

$$C^* - f_0^* \leq \epsilon + \epsilon\big(f_L(0) - f_L^*\big)/C^* \text{ if } C_L = C. \quad (37)$$

Proof. Please see Appendix F of [23] for a proof. □

This result tells us that, when $C_L = C$, any optimal point of (33) is $O(\epsilon)$-suboptimal for the original problem in (28). Also, since $f_L(0) \leq f_0(0) = \mathbf{c}^\mathsf{T}\mathbf{p}_{\rm se}(\mathbf{0})$, the bound in (37) is upper bounded by $\epsilon\big(1 + \frac{\mathbf{c}^\mathsf{T}\mathbf{p}_{\rm se}(\mathbf{0})}{C^*}\big)$, which can be computed before solving $(\mathrm{P_{R3}})$. Therefore, a natural question that arises is: *Can we determine if the condition $C_L = C$ holds for some $C < C^*$ without having to solve the convex relaxation $(\mathrm{P_{R3}})$?*

The following theorem offers a (partial) answer to this question by providing a sufficient condition for the condition $C_L = C$ to hold. For a given budget constraint $C \in \mathbb{R}_+$, define $\mathbb{S}_C = \{\mathbf{s} \in \mathbb{R}_+^N \mid w(\mathbf{s}) \leq C\}$.

Theorem 7. *Suppose that every $\mathbf{s} \in \mathbb{S}_C$ satisfies*

$$B^\mathsf{T}\mathtt{diag}(\boldsymbol{\alpha})^{-1}\nabla w(\mathbf{s}) \lneq \mathbf{c}. \quad (38)$$

Then, $C_L = C$. If (38) *holds for all* $\mathbf{s} \in \mathbb{S}_{C^*}$, *then* $f_0^* = C^*$.

Proof. A proof can be found in Appendix G of [23]. □

Note that the condition (38) can be verified prior to solving the relaxed problem $(\mathrm{P_{R3}})$. In the case that condition (38) holds for all $\mathbf{s} \in \mathbb{S}_{C^*}$, any optimal point $\mathbf{s}_0$ of (33) is also optimal for our original problem in (28).

6.2.2 Upper Bound via the Reduced Gradient Method

In order to use the RGM for the problem in (29), we first need to introduce a following modification to Algorithm 2: replace line 5 of Algorithm 2 with

$$\mathbf{s}^{(t+1)} \leftarrow \mathcal{P}_{\mathbb{S}_C}\big[\mathbf{s}^{(t)} + \gamma_t(\boldsymbol{\alpha} \circ \mathbf{p}^{(t)} \circ \mathbf{u})\big], \tag{39}$$

where $\mathcal{P}_{\mathbb{S}_C}[\cdot]$ denotes the Euclidean projection onto $\mathbb{S}_C$. When w is simple, this projection step can be very efficient.

The results in Section 5 still hold in this case. In particular, at any feasible point $(\mathbf{s}^\star, \mathbf{p}^\star)$ of problem (P) with $\boldsymbol{\lambda} = \mathbf{0}$ such that $\mathbf{s}^\star \in \mathbb{S}_C$, the matrix

$$M(\mathbf{s}^\star) = \mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}^\star + \boldsymbol{\delta} + B\mathbf{p}^\star) - \mathtt{diag}(\mathbf{1} - \mathbf{p}^\star)B,$$

which arises from totally differentiating the constraint $\mathbf{g}(\mathbf{s}, \mathbf{p}) = \mathbf{0}$, is still a nonsingular M-matrix. This can be verified by noting that $M(\mathbf{s}^\star)$ satisfies $M(\mathbf{s}^\star)\mathbf{p}^\star = \mathbf{p}^\star \circ (B\mathbf{p}^\star) > \mathbf{0}$, where the positivity follows from $\mathbf{p}^\star > \mathbf{0}$ because we require that $\mathbf{s}^\star \in \mathbb{S}_C$ with $C < C^*$. As a result, we can use Algorithm 2 with an efficient evaluation of reduced gradient as shown in subsection 5.1.2 and the projection step as described above.

Remark 6. We summarize how to find a good solution to (28) when $\rho\big(\mathtt{diag}(\boldsymbol{\delta})^{-1}B\big) > 1$ based on the above discussion: first, find a pair $(\mathbf{s}_0, C^*)$ of the optimal point and optimal value of (33). If (38) holds for all $\mathbf{s} \in \mathbb{S}_{C^*}$, $\mathbf{s}_0$ is an optimal point of (28). Otherwise, solve $(\mathrm{P_{R3}})$ with $C = C^* - \epsilon$ (for small ϵ) and let $(\mathbf{s}_L, f_L)$ be the pair of its optimal point and optimal value. If $w(\mathbf{s}_L) = C$, adopt $\mathbf{s}_0$ as a solution to (28) with $\mathtt{opt\_gap} \leq \epsilon(1 + \frac{\mathbf{c}^\mathsf{T}\mathbf{p}_{\mathrm{se}}(\mathbf{0})}{C^*})$. Otherwise, solve (29) using RGM and adopt its solution $\mathbf{s}_U$ as a solution to (28) with $\mathtt{opt\_gap} \leq f_U - f_L$, where $f_U = w(\mathbf{s}_U) + \mathbf{c}^\mathsf{T}\mathbf{p}_{\mathrm{se}}(\mathbf{s}_U)$.

7 Numerical Results

In this section, we provide some numerical results that demonstrate the performance of the proposed algorithms. Our numerical studies are carried out in MATLAB (version 9.5) on a laptop with 8GB RAM and a 2.4GHz Intel Core i5 processor. We consider 5 different strongly connected scale-free networks with the power law parameter for node degrees set to 1.5, and the minimum and maximum node degrees equal to 2 and $\lceil 3 \log N \rceil$, respectively, in order to ensure network connectivity with high probability.

For all considered networks, we fix $\alpha_i = 1$ and $\delta_i = 0.1$ for all $i \in \mathcal{A}$. The infection rates $\beta_{j,i}$, $(j,i) \in \mathcal{E}$, are modeled using i.i.d. Uniform(0,1) random variables. We choose

$$w(\mathbf{s}) = \mathbf{1}^\mathsf{T}\mathbf{s} \quad \text{and} \quad \mathbf{c} = \nu B^\mathsf{T}\mathbf{1} + 2\mathbf{c}_{\mathrm{rand}},$$

where the elements of $\mathbf{c}_{\mathrm{rand}}$ are given by i.i.d. Uniform(0,1) random variables and $\nu \geq 0$ is a varying parameter. We select $\mathbf{c}$ above, in order to reflect an observation that nodes which support more neighbors should, on the average, have larger economic costs modeled by c_i^e (Section 3-A). We consider two separate cases: $\boldsymbol{\lambda} > \mathbf{0}$ and $\boldsymbol{\lambda} = \mathbf{0}$.

7.1 Case $\boldsymbol{\lambda} > \mathbf{0}$

We generate $\boldsymbol{\lambda}$ using i.i.d. Uniform(0,1) random variables for each network, set $\nu \in \{0, 0.5, 1\}$, and apply 5 schemes described below. The results (averaged over 10 runs) are summarized in Table 1, and a more detailed description of the simulation setups can be found in [23, Appendix J]. Here, the reported optimality gap is the relative optimality gap given by $\mathtt{opt\_gap} = (f^{\mathrm{cur}} - f_{\mathrm{R}}^*)/f_{\mathrm{R}}^*$, where $f_{\mathrm{R}}^* := \min(f_{\mathrm{R1}}^*, f_{\mathrm{R2}}^*)$, f_{R1}^* and f_{R2}^* are the optimal values of $(\mathrm{P_{R1}})$ and $(\mathrm{P_{R2}})$, respectively, and f^{cur} is the cost achieved by the solution found by the algorithm under consideration. Although the two convex relaxations are equivalent, their numerical solutions are not necessarily identical and we take a conservative lower bound given by the minimum of the two values. When the optimal value of a convex relaxation is unavailable, we take the other optimal value. Also, the column t_s indicates the total runtime.

- `M-matrix + OPTI`: We solve the relaxed problem based on M-matrix in subsection 4.1 using Algorithm 1, where line 3 utilizes an interior point method from the OPTI package [7], and consider the feasible point $(\tilde{\mathbf{s}}, \tilde{\mathbf{p}})$ in Theorem 2. The column iter shows the pair of (i) the number of outer updates (each corresponding to an approximation $\tilde{\Omega}(\underline{\mathbf{z}}^{(t)})$ of the set Ω) and (ii) the average number of inner interior-point iterations inside outer updates. As we can see, the algorithm runtime does not scale well with the network size; for the case $(N, |\mathcal{E}|) = (2001, 12076)$, the solver failed to converge within an hour.
- `K-Exp + MOSEK`: We solve the relaxed problem $(\mathrm{P_{R2}})$ with exponential cone constraints using the MOSEK package [1] and consider the feasible point $(\mathbf{s}', \mathbf{p}')$ in Theorem 3. The column iter indicates the number of interior point iterations. As expected, this method enjoys smaller runtimes and, hence, has a computational advantage over the `M-matrix + OPTI` scheme.
- `RGM + ARMIJO`: We use the RGM in Algorithm 2 to find a local minimizer $(\mathbf{s}^\star, \mathbf{p}^\star)$. The column iter shows the pair of (a) the number of gradient updates and (b) the maximum number of fixed point iterations needed for evaluating $\mathbf{u}$ in line 4 and $\mathbf{p}^*$ in line 7, denoted by $\bar{k}_{\mathrm{fp}}$. In our study, the reported values of $\bar{k}_{\mathrm{fp}}$ are all relatively small as expected from our earlier discussions (Theorem 1 and subsection 5.1.2).
- `M-Matrix SCP`: We use Algorithm 3 to find a local minimizer. Here, we solve the convex optimization subproblem in line 5 of Algorithm 3 approximately, using the OPTI package for $N \leq 10^3$ and a gradient descent method for $N > 10^3$. The column iter shows the pair of (a) the number of outer linearization updates and (b) the average inner steps of either the interior-point solver or gradient descent method. However, this approach does not scale well due to the M-matrix based relaxation as shown in Table 1.
- `K-Exp SCP`: For this algorithm, we replace the convex optimization subproblem in line 5 of Algorithm 3 with the formulation in (27) and solve it approximately using MOSEK. The column iter shows the pair of (a) the number of outer linearization updates and (b) the average inner interior-point steps. As we can see from Table 1, this approach achieves similar opt_gap as `RGM + ARMIJO`, but the runtime is roughly two orders higher. Compared

TABLE 1
Numerical Results ($\boldsymbol{\lambda} > \mathbf{0}$).

$\nu = 0$	M-Matrix + OPTI			K-Exp + MOSEK			RGM + ARMIJO			K-Exp SCP			M-Matrix SCP		
$N, \lvert\mathcal{E}\rvert$	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)
$100, 474$	$8.89\text{e}-2$	$3, 15$	0.41	$8.89\text{e}-2$	13	0.09	$1.16\text{e}-2$	$12, 8$	0.00	$1.16\text{e}-2$	$4, 14$	0.41	$1.16\text{e}-2$	$4, 11$	0.13
$200, 1014$	$1.13\text{e}-1$	$3, 18$	1.71	$1.13\text{e}-1$	15	0.21	$1.31\text{e}-2$	$8, 7$	0.00	$1.31\text{e}-2$	$4, 16$	0.77	$1.31\text{e}-2$	$4, 13$	0.23
$499, 2738$	$1.36\text{e}-1$	$3, 27$	27.8	$1.36\text{e}-1$	16	0.66	$1.26\text{e}-2$	$8, 8$	0.01	$1.26\text{e}-2$	$5, 18$	3.53	$1.26\text{e}-2$	$4, 13$	0.63
$999, 5750$	$1.41\text{e}-1$	$4, 36$	428	$1.41\text{e}-1$	18	1.76	$1.40\text{e}-2$	$11, 8$	0.03	$1.40\text{e}-2$	$5, 20$	8.86	$1.40\text{e}-2$	$4, 17$	2.80
$2001, 12076$	n/a			$1.47\text{e}-1$	17	4.78	$1.37\text{e}-2$	$17, 7$	0.15	$1.37\text{e}-2$	$4, 19$	16.1	$1.37\text{e}-2$	$4, 108$	4.63

$\nu = 0.5$	M-Matrix + OPTI			K-Exp + MOSEK			RGM + ARMIJO			K-Exp SCP			M-Matrix SCP		
$N, \lvert\mathcal{E}\rvert$	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)
$100, 474$	$1.60\text{e}-2$	$2, 20$	0.33	$1.60\text{e}-2$	11	0.08	$3.24\text{e}-3$	$20, 8$	0.01	$3.24\text{e}-3$	$6, 16$	0.58	$3.24\text{e}-3$	$5, 16$	0.22
$200, 1014$	$1.16\text{e}-2$	$3, 20$	1.88	$1.16\text{e}-2$	12	0.18	$1.98\text{e}-3$	$21, 7$	0.01	$1.98\text{e}-3$	$5, 15$	0.94	$1.98\text{e}-3$	$4, 20$	0.37
$499, 2738$	$1.26\text{e}-2$	$3, 29$	31.4	$1.26\text{e}-2$	13	0.57	$2.26\text{e}-3$	$14, 10$	0.02	$2.26\text{e}-3$	$5, 17$	3.84	$2.26\text{e}-3$	$5, 18$	1.08
$999, 5750$	$1.52\text{e}-2$	$3, 29$	229	$1.52\text{e}-2$	15	1.60	$2.33\text{e}-3$	$19, 11$	0.06	$2.33\text{e}-3$	$6, 17$	9.66	$2.33\text{e}-3$	$5, 24$	5.50
$2001, 12076$	n/a			$1.56\text{e}-2$	15	4.53	$2.33\text{e}-3$	$25, 8$	0.22	$2.33\text{e}-3$	$5, 18$	21.3	$2.33\text{e}-3$	$5, 208$	10.7

$\nu = 1$	M-Matrix + OPTI			K-Exp + MOSEK			RGM + ARMIJO			K-Exp SCP			M-Matrix SCP		
$N, \lvert\mathcal{E}\rvert$	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)	opt_gap	iter	t_s (s)
$100, 474$	$3.32\text{e}-9$	$1, 24$	0.21	$2.0\text{e}-10$	11	0.08	$7.58\text{e}-8$	$80, 9$	0.02	$6.75\text{e}-7$	$15, 12$	1.20	$5.12\text{e}-7$	$15, 18$	0.84
$200, 1014$	$2.88\text{e}-9$	$1, 34$	1.03	$8.0\text{e}-10$	13	0.20	$1.52\text{e}-7$	$94, 8$	0.05	$1.02\text{e}-6$	$16, 14$	2.92	$8.49\text{e}-7$	$16, 21$	1.59
$499, 2738$	$2.91\text{e}-9$	$1, 35$	11.5	$1.5\text{e}-10$	11	0.51	$1.06\text{e}-7$	$116, 10$	0.16	$1.03\text{e}-6$	$18, 16$	12.8	$8.09\text{e}-7$	$18, 24$	5.68
$999, 5750$	$2.57\text{e}-9$	$2, 51$	278	$3.61\text{e}-9$	12	1.30	$1.57\text{e}-7$	$126, 14$	0.39	$1.20\text{e}-6$	$18, 17$	29.4	$9.41\text{e}-7$	$18, 27$	23.9
$2001, 12076$	n/a			$1.50\text{e}-9$	12	3.76	$1.25\text{e}-7$	$140, 9$	1.35	$1.07\text{e}-6$	$19, 17$	76.7	$2.79\text{e}-5$	$16, 663$	105

to `M-Matrix SCP`, its performance, both in terms of the quality of solution and runtime, is comparable.

We summarize observations. First, as ν increases and infection costs become larger, as expected from Lemma 3, the optimality gap diminishes and becomes negligible when $\nu = 1$. Second, the upper bounds from local minimizers are very close to the lower bound f_{R}^*, even when the relaxation may not be exact (for $\nu = 0$ and 0.5). Moreover, they lead to optimal points when the relaxation is exact. This suggests that the algorithms can practically find global solutions to the original problem. Finally, Algorithm 2 based on `RGM`, is highly scalable: despite a larger number of required iterations compared to all other schemes, the total runtime t_s is much smaller and is a fraction of that of Algorithm 1 or 3; we note that we did not optimize step sizes; we instead used the same parameters in all cases.

We also tried `sqp` and `interior-point` solvers in MATLAB for problem (P), but found them to be very inefficient compared to our approaches to finding local optimizers. For example, for the case $(N, |\mathcal{E}|) = (999, 5750)$ and $\nu = 0.5$, while RGM runs in only a fraction of a second, `sqp` takes 19 iterations in 102 seconds to achieve the same opt_gap as RGM, and `interior-point` terminates after 125 iterations in 68 seconds with twice the opt_gap of RGM.

7.2 Case $\boldsymbol{\lambda} = \mathbf{0}$

In this subsection, we study the scenario with no primary attacks, using the scale-free network with 499 nodes from the previous subsection. We consider the value of ν in $\{0.6, 0.8, 1\}$ in order to obtain more informative numbers.

Following the steps outlined in Section 6, we first find the optimal value C^* and an optimal point $\mathbf{s}_0$ of problem (33), using MOSEK. When $\nu = 1$, condition (38) in Theorem 7 holds for all $\mathbf{s} \in \mathbb{R}_+^N$ and, thus, we have $f_0^* = C^*$ with $\mathbf{s}_0$ being an optimal point of the original problem in (28).

Second, for $\nu < 1$, we consider the problem in (29) with $\epsilon = 0.01C^*$ or, equivalently, $C_1 = 0.99C^*$, and find a lower bound $f_L(C_1)$, which is the optimal value of the relaxed problem $(\mathrm{P_{R3}})$, using MOSEK. For $\nu = 0.6$ and 0.8, we found that the constraint $w(\mathbf{s}) \leq C_1$ is inactive at the optimal point and, consequently, $f_L^* = f_L(C_1)$. In addition, using the projected RGM described in subsection 6.2.2, we also compute an upper bound $f_U(C_1)$ on the optimal value f_0^* and then consider the gap $\Delta f_0 := \min\{f_U(C_1), C^*\} - f_L^*$.

We plot in Fig. 1 both the upper bound $f_U(C)$ and the lower bound $f_L(C)$ of the optimal value of problem in (29) as a function of C over the interval $[0, 0.99C^*]$. There are several observations we make from the plots.

o1. When $\nu = 0.6$ and the infection costs $\mathbf{c}$ are small, the left plot shows $f_U(C) < C^*$, which tells us that the local minimizer found by the projected RGM of the problem in (29) with the given security budget C, is better than the optimal point of (33). The plot also indicates that both the upper and lower bounds quickly reach a plateau less than C^* with increasing C. This likely suggests that the optimal security investment at an optimal point of (28) is significantly smaller than C^*.

o2. As we discussed just before subsection 6.2.1, the left plot highlights the practical usefulness of our approach: by considering the problem in (29) for $C = C^* - \epsilon$ with small positive ϵ, we can quickly estimate the optimal security investments, i.e., whether or not they are close to C^* (when the security budget constraint is active at $\mathbf{s}_C^*$) or is equal to $w(\mathbf{s}_C^*)$ (when the constraint is inactive at $\mathbf{s}_C^*$), without suffering from the numerical issues mentioned earlier.

o3. As ν increases, so do the bounds $f_U(C_1)$ and $f_L(C_1)$ (normalized by C^*). For larger values of ν with higher infection costs (middle and right plots), the upper bound $f_U(C_1)$ obtained from a local minimizer is slightly larger than another upper bound C^*, suggesting that the budget

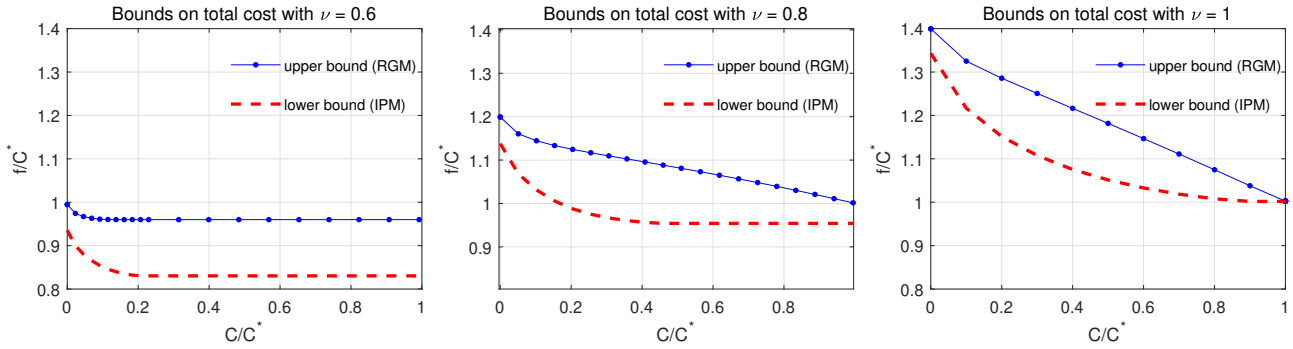

Fig. 1. Upper and lower bounds on the optimal value of the problem in (29) with varying *fictitious* security budget C ($\nu = 0.6, 0.8$ and 1.0).

constraint is active at a local minimizer returned by the projected RGM. This suggests that, when infection costs are high, $\mathbf{s}_0$, which may overinvest compared to an optimal point, may still be a good feasible point. Furthermore, as ν gets close to one, eventually $\mathbf{s}_0$ becomes an optimal point for the problem in (28) as shown in the right plot for $\nu = 1$. This is expected because as the infection costs become larger, the system operator has an incentive to invest more in security.

o4. Although we do not report detailed numbers here, both the upper and lower bounds can be computed efficiently using RGM and MOSEK; the required computational time is always less than 2 seconds for each run, suggesting that the RGM may be a good method for identifying suitable security investments for large systems.

8 Discussion

8.1 Constraints on Security Investments

As mentioned in Section 3, we imposed only non-negativity constraints on security investments $\mathbf{s}$ in problem (P) and subsequent problems. In this subsection, we discuss how additional constraint(s) on $\mathbf{s}$, such as a budget constraint, affect our main results reported in Sections 4 through 6.

8.1.1 Case with $\boldsymbol{\lambda} \gneq \mathbf{0}$

Suppose that the security investments $\mathbf{s}$ are required to lie in some convex set $\mathcal{S} \subset \mathbb{R}^N_+$ in problem (P). Then, the relaxed problems $(\mathrm{P_{R1}})$ and $(\mathrm{P_{R2}})$ are still convex. However, the pair $(\tilde{\mathbf{s}}, \tilde{\mathbf{p}})$ (resp. $(\mathbf{s}', \mathbf{p}')$) in Theorem 2 (resp. 3) is a feasible point of problem (P) if and only if $\tilde{\mathbf{s}} \in \mathcal{S}$ (resp. $\mathbf{s}' \in \mathcal{S}$). For this reason, the convex relaxations $(\mathrm{P_{R1}})$ and $(\mathrm{P_{R2}})$ are exact if $\tilde{\mathbf{s}}$ and $\mathbf{s}'$ lie in $\mathcal{S}$ and the condition (11) in Lemma 3 holds. In addition, in order to ensure that $\mathbf{s}^{(t+1)}$ belongs to $\mathcal{S}$, line 6 of Algorithm 2 needs to be modified as follows:

$$\mathbf{s}^{(t+1)} \leftarrow \mathcal{P}_{\mathcal{S}}\big[\mathbf{s}^{(t)} - \gamma_t\big(\nabla w(\mathbf{s}^{(t)}) - \boldsymbol{\alpha} \circ \mathbf{p}^{(t)} \circ \mathbf{u}\big)\big],$$

where $\mathcal{P}_{\mathcal{S}}[\cdot]$ denotes the projection operator onto $\mathcal{S}$.

8.1.2 Case with $\boldsymbol{\lambda} = \mathbf{0}$

The finding in Theorem 5 continues to hold when the minimum element $\mathbf{s}_{\min}$ of the set $\mathcal{S}$ exists, with the minimum element $\mathbf{s}_{\min}$ being the unique optimal point of the problem in (28) when $\rho\big(\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s}_{\min} + \boldsymbol{\delta})^{-1} B\big) \leq 1$. Hence, when the recovery rates $\boldsymbol{\delta}$ are sufficiently large, only the minimum investments given by $\mathbf{s}_{\min}$ are needed. Obviously, when $\mathcal{S} = \mathbb{R}^N_+$, the minimum element is $\mathbf{0}$. Also, for a general constraint set $\mathcal{S}$, problem (33) is not guaranteed to be feasible, i.e., $\rho(\mathtt{diag}(\boldsymbol{\alpha} \circ \mathbf{s} + \boldsymbol{\delta})^{-1} B) > 1$ for all $\mathbf{s} \in \mathcal{S}$. This means that $\mathbf{p}_{\mathrm{se}}(\mathbf{s}) > \mathbf{0}$ for all $\mathbf{s} \in \mathcal{S}$ and our methods in Sections 4 and 5 can be applied directly to problem (28).

8.2 Relaxation of Irreducibility of B

Although we suspect that irreducibility of matrix B is a reasonable assumption for many systems of interest, such as enterprise intranets, some systems may not satisfy this assumption. For this reason, here we discuss how relaxing this assumption affects our results.

Note that the irreducibility of B is used to (i) ensure the existence of a unique equilibrium $\mathbf{p}^*(\mathbf{s}) \in (0,1)^N$ of (1) as shown in Theorem 1, and (ii) make use of Lemma 2 for our M-matrix based convex formulations.

We relax the assumption that B is irreducible and instead assume that, for every system $i \in \mathcal{A}$, either $\lambda_i > 0$ or there is another system $j \in \mathcal{A} \setminus \{i\}$ with $\lambda_j > 0$ and a directed path to i in $\mathcal{G}$. Then, the main results in Theorem 1 still hold, i.e., there is a unique equilibrium $\mathbf{p}^*(\mathbf{s}) \in (0,1)^N$ of (1) which is strictly positive and can be computed via iteration (2). Moreover, our formulations and results based on exponential cones, including the convex relaxation $(\mathrm{P_{R2}})$ and Theorem 3, are still valid because they rely only on the positivity of $\mathbf{p}(\mathbf{s})$. However, the convexity of problem $(\mathrm{P_{R1}})$ is not guaranteed and requires extending Lemma 2. Finally, when the aforementioned assumption does not hold, the problem is more complicated; our results cannot be applied directly, and it is still an open problem.

9 Conclusions

We studied the problem of determining suitable security investments for hardening interdependent component systems of large systems against malicious attacks and infections. Our formulation aims to minimize the average aggregate costs of a system operator based on the steady-state analysis. We showed that the resulting optimization problem is nonconvex, and proposed a set of algorithms for finding a good solution; two approaches are based on convex relaxations, and the other two look for a local minimizer based on RGM and SCP. In addition, we derived a sufficient condition under which the convex relaxations are exact. Finally, we evaluated the proposed algorithms and demonstrated that, although the original problem is

nonconvex, local minimizers found by the RGM and SCP methods are good solutions with only small optimality gaps. In addition, as predicted by our analytical results, when the infection costs are high, the optimal points of convex relaxations solve the original nonconvex problem.

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] Dimitri P Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.

[4] 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.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

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

[7] Jonathan Currie and David I Wilson. OPTI: lowering the barrier between open source optimizers and the industrial MATLAB user. *Foundations of Computer-aided Process Operations*, 24:32, 2012.

[8] Daniel Gabay and David G Luenberger. Efficiently converging minimization methods based on the reduced gradient. *SIAM J. Control Optim.*, 14(1):42–61, 1976.

[9] Eric Gourdin, Jasmina Omic, and Piet Van Mieghem. Optimization of network protection against virus spread. In *Proc. of DRCN*, pages 86–93, 2011.

[10] Roger Horn and Charles Johnson. *Matrix Analysis*. Cambridge University Press, 1985.

[11] 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.

[12] Libin Jiang, Venkat Anantharam, and Jean Walrand. How bad are selfish investments in network security? *IEEE/ACM Trans. Netw.*, 19(2):549–560, April 2011.

[13] Mohammad M. Khalili, Parinaz Naghizadeh, and Mingyan Liu. Designing cyber insurance policies: the role of pre-screening and security interdependence. *IEEE Trans. Inf. Forensics Security*, 13(9):2226–2239, September 2018.

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

[15] Howard Kunreuther and Geoffrey Heal. Interdependent security. *J. Risk and Uncertainty*, 26(2/3):231–249, 2003.

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

[17] Leon S Lasdon, Richard L Fox, and Margery W Ratner. Nonlinear optimization using the generalized reduced gradient method. *Revue française d'automatique, informatique, recherche opérationnelle. Recherche opérationnelle*, 8(V3):73–103, 1974.

[18] 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.

[19] Van Sy Mai and Eyad H Abed. Optimizing leader influence in networks through selection of direct followers. *IEEE Trans. Autom. Control*, 64(3):1280–1287, 2018.

[20] Van Sy Mai and Abdella Battou. Asynchronous distributed matrix balancing and application to suppressing epidemic. In *Proc. of American Control Conf.*, pages 2777–2782. IEEE, 2019.

[21] 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.

[22] Van Sy Mai, Richard La, and Abdella Battou. Optimal cybersecurity investments for SIS model. In *Proc. of IEEE Global Communications Conf.*, 2020.

[23] Van Sy Mai, Richard La, and Abdella Battou. Optimal Cybersecurity Investments in Large Networks Using SIS Model: Algorithm Design. *arXiv preprint arXiv:2005.07257*, 2020. https://arxiv.org/abs/2005.07257.

[24] Pratyusa K. Manadhata and Jeannette M. Wing. An attack surface metric. *IEEE Trans. softw. eng.*, 37(3):371–386, May-June 2011.

[25] Piet Van Mieghem, Jasmina Omic, and Robert Kooij. Virus spread in networks. *IEEE/ACM Trans. Netw.*, 17(1):1–14, February 2009.

[26] Erik Miehling, Mohammad Rasouli, and Demosthenis Teneketzis. A POMDP approach to the dynamic defense of large-scale cyber networks. *IEEE Trans. Inf. Forensics Security*, 13(10):2490–2505, October 2018.

[27] 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.

[28] Rafail Ostrovsky, Yuval Rabani, and Arman Yousefi. Matrix balancing in L_p norms: bounding the convergence rate of Osborne's iteration. In *Proc. ACM-SIAM Symp. Discrete Algorithms*, 2017.

[29] 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.

[30] Ranjan Pal, Leena Golubchik, Konstantinos Psounis, and Pan Hui. Security pricing as enabler of cyber-insurance: a first look at differentiated pricing markets. *IEEE Trans. Dependable Secure Comput.*, 16(2):358–372, March/April 2019.

[31] J Peña. A stable test to check if a matrix is a nonsingular M-matrix. *Math. Comp.*, 73(247):1385–1392, 2004.

[32] Robert J Plemmons. M-matrix characterizations. I–nonsingular M-matrices. *Linear Algebra Its Appl.*, 18(2):175–188, 1977.

[33] 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.

[34] 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.

[35] Oleg Sheyner and Jeannette M. Wing. Tools for generating and analyzing attack graphs. In *Proc. of Int. Symposium on Formal Methods for Components and Objects*, 2003.

Van Sy Mai received his B.E. degree in Electrical Engineering from the Hanoi University of Technology in 2008, his M.E. degree in Electrical Engineering from the Chulalongkorn University in 2010, and his Ph.D. degree in Electrical and Computer Engineering from the University of Maryland in 2017. Since 2017, he has been a guest researcher at the National Institute of Standards and Technology.

Richard J. La received his Ph.D. degree in Electrical Engineering from the University of California, Berkeley in 2000. Since 2001 he has been on the faculty of the Department of Electrical and Computer Engineering at the University of Maryland, where he is currently a Professor. He is currently an associate editor for IEEE/ACM Transactions on Networking, and served as an associate editor for IEEE Transactions on Information Theory and IEEE Transactions on Mobile Computing.

Abdella Battou is the Division Chief of the Advanced Network Technologies Division, within The Information Technology Lab at NIST. He also leads the Cloud Computing Program. His research areas in Information and Communications Technology (ICT) include cloud computing, high performance optical networking, information centric networking, and more recently quantum networking. From 2009 to 2012, prior to joining NIST, he served as the Executive Director of The Mid-Atlantic Crossroads (MAX) GigaPop. From 2000 to 2009, he was Chief Technology Officer, and Vice President of Research and Development for Lambda OpticalSystems. Dr. Battou holds a PhD and MSEE in Electrical Engineering, and MA in Mathematics all from the Catholic University of America.