Securing Infrastructure Facilities: When Does Proactive Defense Help?

Dynamic Games and Applications
https://doi.org/10.1007/s13235-018-0280-8

Securing Infrastructure Facilities: When Does Proactive Defense Help?

Manxi Wu[1] · **Saurabh Amin**[1]

Abstract
Infrastructure systems are increasingly facing new security threats due to the vulnerabilities of cyber-physical components that support their operation. In this article, we investigate how the infrastructure operator (defender) should prioritize the investment in securing a set of facilities in order to reduce the impact of a strategic adversary (attacker) who can target a facility to increase the overall usage cost of the system. We adopt a game-theoretic approach to model the defender-attacker interaction and study two models: normal form game—where both players move simultaneously—and sequential game—where attacker moves after observing the defender's strategy. For each model, we provide a complete characterization of how the set of facilities that are secured by the defender in equilibrium vary with the costs of attack and defense. Importantly, our analysis provides a sharp condition relating the cost parameters for which the defender has the first-mover advantage. Specifically, we show that to fully deter the attacker from targeting any facility, the defender needs to proactively secure all "vulnerable facilities" at an appropriate level of effort. We illustrate the outcome of the attacker–defender interaction on a simple transportation network. We also suggest a dynamic learning setup to understand how this outcome can affect the ability of imperfectly informed users to make their decisions about using the system in the post-attack stage.

Keywords Infrastructure security · Normal form game · Sequential game

1 Introduction

In this article, we consider the problem of strategic allocation of defense effort to secure one or more facilities of an infrastructure system that is prone to a targeted attack by a malicious adversary. The setup is motivated by the recent incidents and projected threats to critical infrastructures such as transportation, electricity, and urban water networks [38,41,43]. Two of the well-recognized security concerns faced by infrastructure operators are: (i) How to

✉ Manxi Wu
manxiwu@mit.edu

Saurabh Amin
amins@mit.edu

[1] Massachusetts Institute of Technology, Cambridge, MA, USA

Published online: 30 August 2018

prioritize investments among facilities that are heterogeneous in terms of the impact that their compromise can have on the overall efficiency (or usage cost) of the system; and (ii) Whether or not an attacker can be fully deterred from launching an attack by proactively securing some of the facilities. Our work addresses these questions by focusing on the most basic form of strategic interaction between the system operator (defender) and an attacker, modeled as a normal form (simultaneous) or a sequential (Stackelberg) game. The normal form game is relevant to situations in which the attacker cannot directly observe the chosen security plan, whereas the sequential game applies to situations where the defender proactively secures some facilities, and the attacker can observe the defense strategy.

In recent years, many game-theoretical models have been proposed to study problems in cyber-physical security of critical infrastructure systems; see [5,35] for a survey of these models. These models are motivated by the questions of strategic network design [25,34,45], intrusion detection [4,18,23,46], interdependent security [6,39], network interdiction [22,49], and attack-resilient estimation and control [17,47].

Our model is relevant for assessing strategic defense decisions for an infrastructure system viewed as a collection of facilities. In our model, each facility is considered as a distinct entity for the purpose of investment in defense, and multiple facilities can be covered by a single investment strategy. The attacker can target a single facility and compromise its operation, thereby affecting the overall operating efficiency of the system. Both players choose randomized strategies. The performance of the system is evaluated by a usage cost, whose value depends on the actions of both players. In particular, if an undefended facility is targeted by the attacker, it is assumed to be compromised, and this outcome is reflected as a change in the usage cost. Naturally, the defender aims to maintain a low usage cost, while the attacker wishes to increase the usage cost. The attacker (resp. defender) incurs a positive cost in targeting (resp. securing) a unit facility. Thus, both players face a trade-off between the usage cost and the attack/defense costs, which results in qualitatively different equilibrium regimes.

We analyze both normal form and sequential games in the above-mentioned setting. First, we provide a complete characterization of the equilibrium structure in terms of the relative vulnerability of different facilities and the costs of defense/attack for both games. Secondly, we identify ranges of attack and defense costs for which the defender gets the first-mover advantage by investing in proactive defense. Furthermore, we relate the outcome of this game (post-attack stage) to a dynamic learning problem in which the users of the infrastructure system are not fully informed about the realized security state (i.e., the identity of compromised facility).

We now outline our main results. To begin our analysis, we make the following observations. Analogous to [24], we can represent the defender's mixed strategy by a vector with elements corresponding to the probabilities for each facility being secured. The defender's mixed strategy can also be viewed as her effort on each facility. Moreover, the attacker/defender only targets/secures facilities whose disruption will result in an increase in the usage cost (Proposition 1). If the increase in the usage cost of a facility is larger than the cost of attack, then we say that it is a vulnerable facility.

Our approach to characterizing Nash equilibrium (NE) of the normal form game is based on the fact that it is strategically equivalent to a zero-sum game. Hence, the set of attacker's equilibrium strategies can be obtained as the optimal solution set of a linear optimization program (Proposition 2). For any given attack cost, we show that there exists a threshold cost of defense, which distinguishes two equilibrium regime types, named as type I and type II regimes. Theorem 1 shows that when the defense cost is lower than the cost threshold (type I regimes), the total attack probability is positive but less than 1, and all vulnerable facilities

are secured by the defender with positive probability. On the other hand, when the defense cost is higher than the threshold (type II regimes), the total attack probability is 1, and some vulnerable facilities are not secured at all.

We develop a new approach to characterize the subgame perfect equilibrium (SPE) of the sequential game, noting that the strategic equivalence to zero-sum game no longer holds in this case. In this game, the defender, as the first mover, either proactively secures all vulnerable facilities with a threshold security effort so that the attacker does not target any facility, or leaves at least one vulnerable facility secured with an effort less than the threshold while the total attack probability is 1. For any attack cost, we establish another threshold cost of the defense, which is strictly higher than the corresponding threshold in the normal form game. This new threshold again distinguishes the equilibrium strategies into two regime types, named as type $\widetilde{\mathrm{I}}$ and type $\widetilde{\mathrm{II}}$ regimes. Theorem 2 shows that when the defense cost is lower than the cost threshold (type $\widetilde{\mathrm{I}}$ regimes), the defender can fully deter the attacker by proactively securing all vulnerable facilities with the threshold security effort. On the other hand, when the defense cost is higher than the threshold (type $\widetilde{\mathrm{II}}$ regimes), the defender maintains the same level of security effort as that in NE, while the total attack probability is 1.

Our characterization shows that both NE and SPE satisfy the following intuitive properties: (i) Both the defender and attacker prioritize the facilities that results in a high usage cost when compromised; (ii) the attack and defense costs jointly determine the set of facilities that are targeted or secured in equilibrium. On the one hand, as the attack cost decreases, more facilities are vulnerable to attack. On the other hand, as the defense cost decreases, the defender secures more facilities with positive effort, and eventually when the defense cost is below a certain threshold (defined differently in each game), all vulnerable facilities are secured with a positive effort; (iii) each player's equilibrium payoff is non-decreasing in the opponent's cost, and non-increasing in her own cost.

It is well known in the literature on two-player games that so long as both players can choose mixed strategies, the equilibrium utility of the first mover in a sequential game is no less than that in a normal form game [8, pp. 126], [48]. However, cases can be found where the first-mover advantage changes from positive to zero when the attacker's observed signal of the defender's strategy is associated with a noise [7]. In the security game setting, the paper [12] analyzed a game where there are two facilities, and the attacker's valuation of each facility is private information. They identify a condition under which the defender's equilibrium utility is strictly higher when his strategy can be observed by the attacker. In contrast, our model considers multiple facilities and assumes that both players have complete information of the usage cost of each facility.

In fact, for our model, we are able to provide sharp conditions under which proactive defense strictly increases the defender's utility. Given any attack cost, unless the defense cost is "relatively high" (higher than the threshold cost in the sequential game), proactive defense is advantageous in terms of strictly improving the defender's utility and fully deterring the attack. However, if the defense cost is "relatively medium" (lower than the threshold cost in sequential game, but higher than that in the normal form game), a higher security effort on each vulnerable facility is required to gain the first-mover advantage. Finally, if the defense cost is "relatively low" (lower than the threshold cost in the normal form game), then the defender can gain advantage by simply making the first move with the same level of security effort as that in the normal form game.

Note that our approach to characterizing NE and SPE can be readily extended to models with facility-dependent cost parameters and less than perfect defense. We conjecture that a different set of techniques will be required to tackle the more general situation in which

the attacker can target multiple facilities at the same time; see [22] for related work in this direction. However, even when the attacker targets multiple facilities, one can find game parameters for which the defender is always strictly better off in the sequential game.

Finally, we provide a brief discussion on rational learning dynamics, aimed at understanding how the outcome of the attacker–defender interaction—which may or may not result in compromise of a facility (state)—affects the ability of system users to learn about the realized state through a repeated use of the system. A key issue is that the uncertainty about the realized state can significantly impact the ability of users to make decisions to ensure that their long-term cost corresponds to the true usage cost of the system. We explain this issue using a simple transportation network as an example, in which rational travelers (users) need to learn about the identity of the facility that is likely to be compromised using imperfect information about the attack and noisy realizations of travel time in each stage of a repeated routing game played over the network.

The results reported in this article contribute to the study on the allocation of defense resources on facilities against strategic adversaries, as discussed in [12,40]. The underlying assumption that drives our analysis is that an attack on each facility can be treated independently for the purpose of evaluating its impact on the overall usage cost. Other papers that also make this assumption include [2,11,13,16]. Indeed, when the impact of facility compromises is related to the network structure, facilities can no longer be treated independently, and the network structure becomes a crucial factor in analyzing the defense strategy [26]. Additionally, network connections can also introduce the possibility of cascading failure among facilities, which is addressed in [1,29]. These settings are not covered by our model.

The paper is structured as follows: In Sect. 2, we introduce the model of both games and discuss the modeling assumptions. We provide preliminary results to facilitate our analysis in Sect. 3. Section 4 characterizes NE, and Sect. 5 characterizes SPE. Section 6 compares both games. We discuss some extensions of our model and briefly introduce dynamic aspects in Sect. 7.

All proofs are included in Appendix.

2 The Model

2.1 Attacker–Defender Interaction: Normal Form Versus Sequential Games

Consider an infrastructure system modeled as a set of components (facilities) $\mathcal{E}$. To defend the system against an external malicious attack, the system operator (defender) can secure one or more facilities in $\mathcal{E}$ by investing in appropriate security technology. The set of facilities in question can include cyber or physical elements that are crucial to the functioning of the system. These facilities are potential targets for a malicious adversary whose goal is to compromise the overall functionality of the system by gaining unauthorized access to certain cyber-physical elements. The security technology can be a combination of proactive mechanisms (authentication and access control) or reactive ones (attack detection and response). Since our focus is on modeling the strategic interaction between the attacker and defender at a system level, we do not consider the specific functionalities of individual facilities or the protection mechanisms offered by various technologies.

We now introduce our game-theoretic model. Let us denote a pure strategy of the defender as $s_d \subseteq \mathcal{E}$, with $s_d \in S_d = 2^{\mathcal{E}}$. The cost of securing any facility is given by the parameter $p_d \in \mathbb{R}_{>0}$. Thus, the total defense cost incurred in choosing a pure strategy s_d is $|s_d| \cdot p_d$,

where $|s_d|$ is the cardinality of s_d (i.e., the number of secured facilities). The attacker chooses to target a single facility $e \in \mathcal{E}$ or not to attack. We denote a pure strategy of the attacker as $s_a \in S_a = \mathcal{E} \cup \{\emptyset\}$. The cost of an attack is given by the parameter $p_a \in \mathbb{R}_{>0}$, and it reflects the effort that attacker needs to spend in order to successfully target a single facility and compromise its operation.

We assume that prior to the attack, the usage cost of the system is $C_\emptyset$. This cost represents the level of efficiency with which the defender is able to operate the system for its users. A higher usage cost reflects lower efficiency. If a facility e is targeted by the attacker but not secured by the defender, we consider that e is compromised and the usage cost of the system changes to C_e. Therefore, given any pure strategy profile (s_d, s_a), the usage cost after the attacker–defender interaction, denoted $C(s_d, s_a)$, can be expressed as follows:

$$C(s_d, s_a) = \begin{cases} C_e, & \text{if } s_a = e, \text{ and } s_d \not\ni e, \\ C_\emptyset, & \text{otherwise.} \end{cases} \tag{1}$$

To study the effect of timing of the attacker–defender interaction, prior literature on security games has studied both normal form game and sequential game [5]. We study both models in our setting. In the normal form game, denoted Γ, the defender and the attacker move simultaneously. On the other hand, in the sequential game, denoted $\widetilde{\Gamma}$, the defender moves in the first stage and the attacker moves in the second stage after observing the defender's strategy. We allow both players to use mixed strategies. In Γ, we denote the defender's mixed strategy as $\sigma_d \triangleq (\sigma_d(s_d))_{s_d \in S_d} \in \Delta(S_d)$, where $\sigma_d(s_d)$ is the probability that the set of secured facilities is s_d. Similarly, a mixed strategy of the attacker is $\sigma_a \triangleq (\sigma_a(s_a))_{s_a \in S_a} \in \Delta(S_a)$, where $\sigma_a(s_a)$ is the probability that the realized action is s_a. Let $\sigma = (\sigma_d, \sigma_a)$ denote a mixed strategy profile. In $\widetilde{\Gamma}$, the defender's mixed strategy $\widetilde{\sigma}_d \triangleq (\widetilde{\sigma}_d(s_d))_{s_d \in S_d} \in \Delta(S_d)$ is defined analogously to that in Γ. The attacker's strategy is a map from $\Delta(S_d)$ to $\Delta(S_a)$, denoted by $\widetilde{\sigma}_a(\widetilde{\sigma}_d) \triangleq (\widetilde{\sigma}_a(s_a, \widetilde{\sigma}_d))_{s_a \in S_a} \in \Delta(S_a)$, where $\widetilde{\sigma}_a(s_a, \widetilde{\sigma}_d)$ is the probability that the realized action is s_a when the defender's strategy is $\widetilde{\sigma}_d$. A strategy profile in this case is denoted as $\widetilde{\sigma} = (\widetilde{\sigma}_d, \widetilde{\sigma}_a(\widetilde{\sigma}_d))$.

The defender's utility is comprised of two parts: the negative of the usage cost as given in (1) and the defense cost incurred in securing the system. Similarly, the attacker's utility is the usage cost net the attack cost. For a pure strategy profile (s_d, s_a), the utilities of defender and attacker can be, respectively, expressed as follows:

$$u_d(s_d, s_a) = -C(s_d, s_a) - p_d \cdot |s_d|, \quad u_a(s_d, s_a) = C(s_d, s_a) - p_a \cdot \mathbb{1}\{s_a \neq \emptyset\}.$$

For a mixed strategy profile (σ_d, σ_a), the expected utilities can be written as:

$$U_d(\sigma_d, \sigma_a) = \sum_{s_d \in S_d} \sum_{s_a \in S_a} u_d(s_d, s_a) \cdot \sigma_a(s_a) \cdot \sigma_d(s_d) = -\mathbb{E}_\sigma[C] - p_d \cdot \mathbb{E}_{\sigma_d}[|s_d|], \tag{2a}$$

$$U_a(\sigma_d, \sigma_a) = \sum_{s_d \in S_d} \sum_{s_a \in S_a} u_a(s_d, s_a) \cdot \sigma_a(s_a) \cdot \sigma_d(s_d) = \mathbb{E}_\sigma[C] - p_a \cdot \mathbb{E}_{\sigma_a}[|s_a|], \tag{2b}$$

where $\mathbb{E}_\sigma[C]$ is the expected usage cost, and $\mathbb{E}_{\sigma_d}[|s_d|]$ (resp. $\mathbb{E}_{\sigma_a}[|s_a|]$) is the expected number of defended (resp. targeted) facilities, i.e.:

$$\mathbb{E}_\sigma[C] = \sum_{s_a \in S_a} \sum_{s_d \in S_d} C(s_d, s_a) \cdot \sigma_a(s_a) \cdot \sigma_d(s_d),$$

$$\mathbb{E}_{\sigma_d}[|s_d|] = \sum_{s_d \in S_d} |s_d| \sigma_d(s_d), \quad \mathbb{E}_{\sigma_a}[|s_a|] = \sum_{e \in \mathcal{E}} \sigma_a(e).$$

An equilibrium outcome of the game Γ is defined in the sense of Nash equilibrium (NE). A strategy profile $\sigma^* = (\sigma_d^*, \sigma_a^*)$ is a NE if:

$$U_d(\sigma_d^*, \sigma_a^*) \geq U_d(\sigma_d, \sigma_a^*), \quad \forall \sigma_d \in \Delta(S_d),$$
$$U_a(\sigma_d^*, \sigma_a^*) \geq U_a(\sigma_d^*, \sigma_a), \quad \forall \sigma_a \in \Delta(S_a).$$

In the sequential game $\widetilde{\Gamma}$, the solution concept is that of a subgame perfect equilibrium (SPE), which is also known as Stackelberg equilibrium. A strategy profile $\widetilde{\sigma}^* = (\widetilde{\sigma}_d^*, \widetilde{\sigma}_a^*(\widetilde{\sigma}_d))$ is a SPE if:

$$U_d(\widetilde{\sigma}_d^*, \widetilde{\sigma}_a^*(\widetilde{\sigma}_d^*)) \geq U_d(\widetilde{\sigma}_d, \widetilde{\sigma}_a^*(\widetilde{\sigma}_d)), \quad \forall \widetilde{\sigma}_d \in \Delta(S_d), \tag{3a}$$
$$U_a(\widetilde{\sigma}_d, \widetilde{\sigma}_a^*(\widetilde{\sigma}_d)) \geq U_a(\widetilde{\sigma}_d, \widetilde{\sigma}_a(\widetilde{\sigma}_d)), \quad \forall \widetilde{\sigma}_d \in \Delta(S_d), \quad \forall \widetilde{\sigma}_a(\widetilde{\sigma}_d) \in \Delta(S_a). \tag{3b}$$

Since both S_d and S_a are finite sets, and we consider mixed strategies, both NE and SPE exist.

2.2 Model Discussion

One of our main assumptions is that the attacker's capability is limited to targeting at most one facility, while the defender can invest in securing multiple facilities. Although this assumption appears to be somewhat restrictive, it enables us to derive analytical results on the equilibrium structure for a system with multiple facilities. The assumption can be justified in situations where the attacker can only target system components in a localized manner. Thus, a facility can be viewed as a set of collocated components that can be simultaneously targeted by the attacker. For example, in a transportation system, a facility can be a vulnerable link (edge), or a set of links that are connected by a vulnerable node (an intersection or a hub). In Sect. 7.1, we briefly discuss the issues in solving a more general game where multiple facilities can be simultaneously targeted by the attacker.

Secondly, our model assumes that the costs of attack and defense are identical across all facilities. We make this assumption largely to avoid the notational burden of analyzing the effect of facility-dependent attack/defense cost parameters on the equilibrium structures. In fact, as argued in Sect. 7.1, the qualitative properties of equilibria still hold when cost parameters are facility-dependent. However, characterizing the equilibrium regimes in this case can be quite tedious and may not necessarily provide new insights on strategic defense investments.

Thirdly, we allow both players to choose mixed strategies. Indeed, mixed strategies are commonly considered in security games as a pure NE may not always exist. A mixed strategy entails a player's decision to introduce randomness in her behavior, i.e., the manner in which a facility is targeted (resp. secured) by the attacker (resp. defender). Consider, for example, the problem of inspecting a transportation network facing risk of a malicious attack. In this problem, a mixed strategy can be viewed as randomized allocation of inspection effort on subsets of facilities. Mixed strategy of the attacker can be similarly interpreted.

Fourthly, we assume that the defender has the technological means to perfectly secure a facility. In other words, an attack on a secured facility cannot impact its operation. As we will see in Sect. 3, the defender's mixed strategy can be viewed as the level of security effort on each facility, where the effort level 1 (maximum) means perfect security, and 0 (minimum) means no security. Under this interpretation, the defense cost p_d is the cost of perfectly securing a unit facility (i.e., with maximum level of effort), and the expected defense cost is p_d scaled by the security effort defined by the defender's mixed strategy.

Fifthly, we do not consider a specific functional form for modeling the usage cost. In our model, for any facility $e \in \mathcal{E}$, the difference between the post-attack usage cost C_e and the pre-attack cost $C_\emptyset$ represents the change of the usage cost of the system when e is compromised. This change can be evaluated based on the type of attacker–defender interaction one is interested in studying. For example, in situations when attack on a facility results in its complete disruption, one can use a connectivity-based metric such as the number of active source–destination paths or the number of connected components to evaluate the usage cost [25,26]. On the other hand, in situations when facilities are congestible resources and an attack on a facility increases the users' cost of accessing it, the system's usage cost can be defined as the average cost for accessing (or routing through) the system. This cost can be naturally evaluated as the user cost in a Wardrop equilibrium [13], although socially optimal cost has also been considered in the literature [3].

Finally, we note that for the purpose of our analysis, the usage cost as given in (1) fully captures the impact of player' actions on the system. For any two facilities $e, e^{'} \in \mathcal{E}$, the ordering of C_e and $C_{e'}$ determines the relative scale of impact of the two facilities. As we show in Sects. 4–5, the order of cost functions in the set $\{C_e\}_{e\in\mathcal{E}}$ plays a key role in our analysis approach. Indeed, the usage cost is intimately linked with the network structure and way of operation (e.g., how individual users are routed through the network and how their costs are affected by a compromised facility). Barring a simple (yet illustrative) example, we do not elaborate further on how the network structure and/or the functional form of usage cost changes the interpretations of equilibrium outcome. We also do not discuss the computational aspects of arriving at the ordering of usage costs.

3 Rationalizable Strategies and Aggregate Defense Effort

We introduce two preliminary results that are useful in our subsequent analysis. Firstly, we show that the defender's strategy can be equivalently represented by a vector of facility-specific security effort levels. Secondly, we identify the set of rationalizable strategies of both players.

For any defender's mixed strategy $\sigma_d \in \Delta(S_d)$, the corresponding *security effort vector* is $\rho(\sigma_d) = (\rho_e(\sigma_d))_{e\in\mathcal{E}}$, where $\rho_e(\sigma_d)$ is the probability that facility e is secured:

$$\rho_e(\sigma_d) = \sum_{s_d \ni e} \sigma_d(s_d). \tag{4}$$

In other words, $\rho_e(\sigma_d)$ is the level of security effort exerted by the defender on facility e under the security plan σ_d. Since $\sigma_d(s_d) \geq 0$ for any $s_d \in S_d$, we obtain that $0 \leq \rho_e(\sigma_d) = \sum_{s_d\ni e} \sigma_d(s_d) \leq \sum_{s_d\in S_d} \sigma_d(s_d) = 1$. Hence, any σ_d induces a valid vector of probabilities $\rho \in [0,1]^{|\mathcal{E}|}$. In fact, any vector $\rho \in [0,1]^{|\mathcal{E}|}$ can be induced by at least one feasible σ_d. The following lemma provides a way to explicitly construct one such feasible strategy.

Lemma 1 *Consider any feasible security effort vector $\rho \in [0,1]^{|\mathcal{E}|}$. Let m be the number of distinct positive values in ρ, and define $\rho_{(i)}$ as the i-th largest distinct value in ρ, i.e., $\rho_{(1)} > \cdots > \rho_{(m)}$. The following defender's strategy is feasible and induces ρ:*

$$\sigma_d(\{e \in \mathcal{E} | \rho_e \geq \rho_{(i)}\}) = \rho_{(i)} - \rho_{(i+1)}, \quad \forall i = 1, \ldots, m-1 \tag{5a}$$

$$\sigma_d(\{e \in \mathcal{E} | \rho_e \geq \rho_{(m)}\}) = \rho_{(m)}, \tag{5b}$$

$$\sigma_d(\emptyset) = 1 - \rho_{(1)}. \tag{5c}$$

For any remaining $s_d \in S_d$, $\sigma_d(s_d) = 0$.

We now re-express the player utilities in (2) in terms of $(\rho(\sigma_d), \sigma_a)$ as follows:

$$\begin{aligned} U_d(\sigma_d, \sigma_a) &= -\sum_{s_a \in S_a} \left(\sum_{s_d \in S_d} \sigma_d(s_d) C(s_d, s_a) \right) \sigma_a(s_a) - \left(\sum_{s_d \in S_d} |s_d| \sigma_d(s_d) \right) p_d \\ &= -\sum_{e \in \mathcal{E}} \left(\sum_{s_d \in S_d} \sigma_d(s_d) C(s_d, e) \right) \sigma_a(e) - C_\emptyset \sigma_a(\emptyset) - \left(\sum_{e \in \mathcal{E}} \rho_e(\sigma_d) \right) p_d \\ &\overset{(1)}{=} -\sum_{e \in \mathcal{E}} \left(\left(\sum_{s_d \ni e} \sigma_d(s_d) \right) C_\emptyset + \left(1 - \sum_{s_d \ni e} \sigma_d(s_d) \right) C_e \right) \sigma_a(e) - C_\emptyset \sigma_a(\emptyset) \\ &\quad - \left(\sum_{e \in \mathcal{E}} \rho_e(\sigma_d) \right) p_d \\ &= -\sum_{e \in \mathcal{E}} \left(\rho_e(\sigma_d) \left((C_\emptyset - C_e) \sigma_a(e) + p_d \right) + C_e \sigma_a(e) \right) - C_\emptyset \sigma_a(\emptyset), \end{aligned} \tag{6a}$$

$$U_a(\sigma_d, \sigma_a) = \sum_{e \in \mathcal{E}} \left(\rho_e(\sigma_d) \left(C_\emptyset - C_e \right) \sigma_a(e) + C_e \sigma_a(e) \right) + C_\emptyset \sigma_a(\emptyset) - \left(\sum_{e \in \mathcal{E}} \sigma_a(e) \right) p_a. \tag{6b}$$

Thus, for any given attack strategy and any two defense strategies, if the induced security effort vectors are identical, then the corresponding utility for each player is also identical. Henceforth, we denote the player utilities as $U_d(\rho, \sigma_a)$ and $U_a(\rho, \sigma_a)$ and use σ_d and $\rho_e(\sigma_d)$ interchangeably in representing the defender's strategy. For the sequential game $\widetilde{\Gamma}$, we analogously denote the security effort vector given the strategy $\widetilde{\sigma}_d$ as $\tilde{\rho}(\widetilde{\sigma}_d)$, and the defender's utility (resp. attacker's utility) as $\widetilde{U}_d(\tilde{\rho}, \widetilde{\sigma}_a)$ (resp. $\widetilde{U}_a(\tilde{\rho}, \widetilde{\sigma}_a)$).

We next characterize the set of rationalizable strategies. Note that the post-attack usage cost C_e can increase or remain the same or even decrease, in comparison with the pre-attack cost $C_\emptyset$. Let the facilities whose damage results in an increased usage cost be grouped in the set $\bar{\mathcal{E}}$. Similarly, let $\widehat{\mathcal{E}}$ denote the set of facilities such that a damage to any one of them has no effect on the usage cost. Finally, the set of remaining facilities is denoted as $\check{\mathcal{E}}$. Thus,

$$\bar{\mathcal{E}} \triangleq \{e \in \mathcal{E} | C_e > C_\emptyset\}, \tag{7a}$$

$$\widehat{\mathcal{E}} \triangleq \{e \in \mathcal{E} | C_e = C_\emptyset\}, \tag{7b}$$

$$\check{\mathcal{E}} \triangleq \{e \in \mathcal{E} | C_e < C_\emptyset\}. \tag{7c}$$

Clearly, $\bar{\mathcal{E}} \cup \widehat{\mathcal{E}} \cup \check{\mathcal{E}} = \mathcal{E}$. The following proposition shows that in a rationalizable strategy profile, the defender does not secure facilities that are not in $\bar{\mathcal{E}}$, and the attacker only considers targeting the facilities that are in $\bar{\mathcal{E}}$.

Proposition 1 *The rationalizable action sets for the defender and attacker are given by* $2^{\bar{\mathcal{E}}}$ *and* $\bar{\mathcal{E}} \cup \{\emptyset\}$*, respectively. Hence, any equilibrium strategy profile* (ρ^*, σ_a^*) *in* Γ *(resp.* $(\tilde{\rho}^*, \widetilde{\sigma}_a^*)$ *in* $\widetilde{\Gamma}$*) satisfies:*

$$\begin{aligned} &\rho_e^* = \sigma_a^*(e) = 0, \qquad \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}, \\ &\tilde{\rho}_e^* = \widetilde{\sigma}_a^*(e, \tilde{\rho}) = 0, \quad \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}, \quad \forall \tilde{\rho} \in [0, 1]^{\mathcal{E}}. \end{aligned}$$

If $\bar{\mathcal{E}} = \emptyset$, then the attacker/defender does not attack/secure any facility in equilibrium. Henceforth, to avoid triviality, we assume $\bar{\mathcal{E}} \neq \emptyset$. Additionally, we define a partition of facilities in $\bar{\mathcal{E}}$ such that all facilities with identical C_e are grouped in the same set. Let the number of distinct values in $\{C_e\}_{e\in\bar{\mathcal{E}}}$ be K, and $C_{(k)}$ denote the k-th highest distinct value in the set $\{C_e\}_{e\in\bar{\mathcal{E}}}$. Then, we can order the usage costs as follows:

$$C_{(1)} > C_{(2)} > \cdots > C_{(K)} > C_{\emptyset}. \tag{8}$$

We denote $\bar{\mathcal{E}}_{(k)}$ as the set of facilities such that if any $e \in \bar{\mathcal{E}}_{(k)}$ is damaged, the usage cost $C_e = C_{(k)}$, i.e., $\bar{\mathcal{E}}_{(k)} \triangleq \left\{e \in \bar{\mathcal{E}} | C_e = C_{(k)}\right\}$. We also define $E_{(k)} \triangleq |\bar{\mathcal{E}}_{(k)}|$. Clearly, $\cup_{k=1}^{K} \bar{\mathcal{E}}_{(k)} = \bar{\mathcal{E}}$, and $\sum_{k=1}^{K} E_{(k)} = |\bar{\mathcal{E}}|$. Facilities in the same group have identical impact on the infrastructure system when compromised.

4 Normal Form Game Γ

In this section, we provide complete characterization of the set of NEs for any given attack and defense cost parameters in game Γ. In Sect. 4.1, we show that Γ is strategically equivalent to a zero-sum game, and hence the set of attacker's equilibrium strategies can be solved by a linear program. In Sect. 4.2, we show that the space of cost parameters $(p_a, p_d) \in \mathbb{R}^2_{>0}$ can be partitioned into qualitatively distinct equilibrium regimes.

4.1 Strategic Equivalence to Zero-Sum Game

Our notion of strategic equivalence is the same as the best response equivalence defined in Rosenthal [42]. If Γ and another game Γ^0 are strategically equivalent, then given any strategy of the defender (resp. attacker), the set of attacker's (resp. defender's) best responses is identical in the two games. This result forms the basis of characterizing the set of NE.

We define the utility functions of the game Γ^0 as follows:

$$U_d^0(\sigma_d, \sigma_a) = -\mathbb{E}_{\sigma}[C] - \mathbb{E}_{\sigma_d}[|s_d|] \cdot p_d + p_a \cdot \mathbb{E}_{\sigma_a}[|s_a|], \tag{9a}$$

$$U_a^0(\sigma_d, \sigma_a) = \mathbb{E}_{\sigma}[C] + \mathbb{E}_{\sigma_d}[|s_d|] \cdot p_d - p_a \cdot \mathbb{E}_{\sigma_a}[|s_a|]. \tag{9b}$$

Thus, Γ^0 is a zero-sum game. We denote the set of defender's (resp. attacker's) equilibrium strategies in Γ^0 as Σ_d^0 (resp. Σ_a^0).

Lemma 2 *The normal form game Γ is strategically equivalent to the zero-sum game Γ^0. The set of defender's (resp. attacker's) equilibrium strategies in Γ is $\Sigma_d^* \equiv \Sigma_d^0$ (resp. $\Sigma_a^* \equiv \Sigma_a^0$). Furthermore, for any $\sigma_d^* \in \Sigma_d^*$ and any $\sigma_a^* \in \Sigma_a^*$, (σ_d^*, σ_a^*) is an equilibrium strategy profile of Γ.*

Based on Lemma 2, the set of attacker's equilibrium strategies Σ_a^* can be expressed as the optimal solution set of a linear program.

Proposition 2 *The set Σ_a^* is the optimal solution set of the following optimization problem:*

$$\max_{\sigma_a} \quad V(\sigma_a)$$

$$s.t. \quad V(\sigma_a) = \sum_{e\in\bar{\mathcal{E}}} \min\{\sigma_a(e)\cdot(C_\emptyset - p_a) + p_d,\ \sigma_a(e)\cdot(C_e - p_a)\} + \sigma_a(\emptyset)\cdot C_\emptyset, \tag{10a}$$

$$\sum_{e\in\bar{\mathcal{E}}} \sigma_a(e) + \sigma_a(\emptyset) = 1, \tag{10b}$$

$$\sigma_a(\emptyset) \geq 0, \quad \sigma_a(e) \geq 0, \quad \forall e \in \bar{\mathcal{E}}. \tag{10c}$$

Furthermore, (10) is equivalent to the following linear optimization program:

$$\max_{\sigma_a, v} \quad \sum_{e\in\bar{\mathcal{E}}} v_e + \sigma_a(\emptyset)\cdot C_\emptyset$$

$$s.t. \quad \sigma_a(e)\cdot(C_\emptyset - p_a) + p_d - v_e \geq 0, \quad \forall e \in \bar{\mathcal{E}}, \tag{11a}$$

$$\sigma_a(e)\cdot(C_e - p_a) - v_e \geq 0, \quad \forall e \in \bar{\mathcal{E}}, \tag{11b}$$

$$\sum_{e\in\bar{\mathcal{E}}} \sigma_a(e) + \sigma_a(\emptyset) = 1, \tag{11c}$$

$$\sigma_a(\emptyset) \geq 0, \quad \sigma_a(e) \geq 0, \quad \forall e \in \bar{\mathcal{E}}. \tag{11d}$$

where $v = (v_e)_{e\in\bar{\mathcal{E}}}$ is an $|\bar{\mathcal{E}}|$-dimensional variable.

In Proposition 2, the objective function $V(\sigma_a)$ is a piecewise linear function in σ_a. Furthermore, given any σ_a and any $e \in \bar{\mathcal{E}}$, we can write:

$$\begin{aligned}&\min\{\sigma_a(e)\cdot(C_\emptyset - p_a) + p_d,\ \sigma_a(e)\cdot(C_e - p_a)\}\\ =&\begin{cases}\sigma_a(e)\cdot(C_\emptyset - p_a) + p_d & \text{if } \sigma_a(e) > \frac{p_d}{C_e - C_\emptyset},\\ \sigma_a(e)\cdot(C_e - p_a) & \text{if } \sigma_a(e) \leq \frac{p_d}{C_e - C_\emptyset}.\end{cases}\end{aligned} \tag{12}$$

Thus, we can observe that if $\sigma_a(e)$ is equal to $p_d/(C_e - C_\emptyset)$, then $-\sigma_a(e)\cdot C_\emptyset - p_d = -\sigma_a(e)\cdot C_e$, i.e., if a facility e is targeted with the threshold attack probability $p_d/(C_e - C_\emptyset)$, the defender is indifferent between securing e versus not. The following lemma analyzes the defender's best response to the attacker's strategy and shows that no facility is targeted with probability higher than the threshold probability in equilibrium.

Lemma 3 *Given any strategy of the attacker $\sigma_a \in \Delta(S_a)$, for any defender's security effort ρ that is a best response to σ_a, denoted $\rho \in BR(\sigma_a)$, the security effort ρ_e on each facility $e \in \mathcal{E}$ satisfies:*

$$\rho_e \begin{cases} = 0, & \forall e \in \left\{\bar{\mathcal{E}} | \sigma_a(e) < \frac{p_d}{C_e - C_\emptyset}\right\} \cup \widehat{\mathcal{E}} \cup \check{\mathcal{E}},\\ \in [0,1], & \forall e \in \left\{\bar{\mathcal{E}} | \sigma_a(e) = \frac{p_d}{C_e - C_\emptyset}\right\},\\ = 1, & \forall e \in \left\{\bar{\mathcal{E}} | \sigma_a(e) > \frac{p_d}{C_e - C_\emptyset}\right\}.\end{cases} \tag{13}$$

Furthermore, in equilibrium, the attacker's strategy σ_a^* satisfies:

$$\sigma_a^*(e) \leq \frac{p_d}{C_e - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}, \tag{14a}$$

$$\sigma_a^*(e) = 0, \quad \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}. \tag{14b}$$

Lemma 3 highlights a key property of NE: The attacker does not target at any facility $e \in \bar{\mathcal{E}}$ with probability higher than the threshold $p_d/(C_e - C_\emptyset)$, and the defender allocates a nonzero security effort only on the facilities that are targeted with the threshold probability.

Intuitively, if a facility e were to be targeted with a probability higher than the threshold $p_d/(C_e - C_\emptyset)$, then the defender's best response would be to secure that facility with probability 1, and the attacker's expected utility will be $-C_\emptyset - p_a\sigma_a(e)$, which is smaller than $-C_\emptyset$ (utility of no attack). Hence, the attacker would be better off by choosing the no attack action.

Now, we can rewrite $V(\sigma_a)$ as defined in (10) as follows:

$$V(\sigma_a) \overset{(14)}{=} \sum_{e \in \{\bar{\mathcal{E}} | \sigma_a(e) \leq \frac{p_d}{C_e - C_\emptyset}\}} \sigma_a(e)\,(C_e - p_a) + C_\emptyset \cdot \sigma_a(\emptyset), \tag{15}$$

and the set of attacker's equilibrium strategies maximizes this function.

4.2 Characterization of NE in Γ

We are now in the position to introduce the equilibrium regimes. Each regime corresponds to a range of cost parameters such that the qualitative properties of equilibrium (i.e., the set of facilities that are targeted and secured) do not change in the interior of each regime.

We say that a facility e is *vulnerable* if $C_e - p_a > C_\emptyset$. Therefore, given any attack cost p_a, the set of vulnerable facilities is given by $\{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$. Clearly, only vulnerable facilities are likely targets of the attacker. If $p_a > C_{(1)} - C_\emptyset$, then there are no vulnerable facilities. In contrast, if $p_a < C_{(1)} - C_\emptyset$, we define the following threshold for the per-facility defense cost:

$$\bar{p}_d(p_a) \overset{\Delta}{=} \frac{1}{\sum_{e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}} \frac{1}{C_e - C_\emptyset}}. \tag{16}$$

We can check that for any $i = 1, \ldots, K-1$ (resp. $i = K$), if $C_{(i+1)} - C_\emptyset \leq p_a < C_{(i)} - C_\emptyset$ (resp. $0 < p_a < C_{(K)} - C_\emptyset$), then

$$\bar{p}_d(p_a) = \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1}. \tag{17}$$

Recall from Lemma 3 that $\sigma_a^*(e)$ is upper bounded by the threshold attack probability $p_d/(C_e - C_\emptyset)$. If the defense cost $p_d < \bar{p}_d(p_a)$, then $\sum_{k=1}^{i} \frac{E_{(k)} p_d}{C_{(k)} - C_\emptyset} < 1$, which implies that even when the attacker targets each vulnerable facility with the threshold attack probability, the total probability of attack is still smaller than 1. Thus, the attacker must necessarily choose not to attack with a positive probability. On the other hand, if $p_d > \bar{p}_d(p_a)$, then the no attack action is not chosen by the attacker in equilibrium.

Following the above discussion, we introduce two types of regimes depending on whether or not p_d is higher than the threshold $\bar{p}_d(p_a)$. In type I regimes, denoted as $\{\Lambda^i | i = 0, \ldots, K\}$, the defense cost $p_d < \bar{p}_d(p_a)$, whereas in type II regimes, denoted as $\{\Lambda_j | j = 1, \ldots, K\}$, the defense cost $p_d > \bar{p}_d(p_a)$. Hence, we say that p_d is "relatively low" (resp. "relatively high") in comparison with p_a in type I regimes (resp. type II regimes). We formally define these $2K + 1$ regimes as follows:

(a) Type I regimes Λ^i, $i = 0, \ldots, K$:

– If $i = 0$:

$$p_a > C_{(1)} - C_\emptyset, \text{ and } p_d > 0 \tag{18}$$

– If $i = 1, \ldots, K-1$:

$$C_{(i+1)} - C_\emptyset < p_a < C_{(i)} - C_\emptyset, \text{ and } 0 < p_d < \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} \tag{19}$$

– If $i = K$:

$$0 < p_a < C_{(K)} - C_\emptyset, \text{ and } 0 < p_d < \left(\sum_{k=1}^{K} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} \tag{20}$$

(b) Type II regimes, Λ_j, $j = 1, \ldots, K$:

– If $j = 1$:

$$0 < p_a < C_{(1)} - C_\emptyset, \text{ and } p_d > \left(\frac{E_{(1)}}{C_{(1)} - C_\emptyset} \right)^{-1} \tag{21}$$

– If $j = 2, \ldots, K$:

$$0 < p_a < C_{(j)} - C_\emptyset, \text{ and } \left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} < p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} \tag{22}$$

We now characterize equilibrium strategy sets Σ_d^* and Σ_a^* in the interior of each regime.[1]

Theorem 1 *The set of NE in each regime is as follows:*

(a) *Type I regimes* Λ^i*:*

– *If* $i = 0$,

$$\rho_e^* = 0, \quad \forall e \in \mathcal{E} \tag{23a}$$

$$\sigma_a^*(\emptyset) = 1. \tag{23b}$$

– *If* $i = 1, \ldots, K$,

$$\rho_e^* = \frac{C_{(k)} - p_a - C_\emptyset}{C_{(k)} - C_\emptyset}, \qquad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, i \tag{24a}$$

$$\rho_e^* = 0, \qquad \forall e \in \mathcal{E} \setminus \left(\cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)} \right), \tag{24b}$$

$$\sigma_a^*(e) = \frac{p_d}{C_{(k)} - C_\emptyset}, \qquad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, i, \tag{24c}$$

$$\sigma_a^*(\emptyset) = 1 - \sum_{e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}} \sigma_a^*(e). \tag{24d}$$

[1] For the sake of brevity, we omit the discussion of equilibrium strategies when cost parameters lie exactly on the regime boundary, although this case can be addressed using the approach developed in this article.

(b) *Type II regimes* Λ_j*:*

– $j = 1$*:*

$$\rho_e^* = 0, \qquad \forall e \in \mathcal{E}, \tag{25a}$$

$$0 \le \sigma_a^*(e) \le \frac{p_d}{C_{(1)} - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}_{(1)}, \tag{25b}$$

$$\sum_{e \in \bar{\mathcal{E}}_{(1)}} \sigma_a^*(e) = 1. \tag{25c}$$

– $j = 2, \ldots, K$*:*

$$\rho_e^* = \frac{C_{(k)} - C_{(j)}}{C_{(k)} - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, j-1, \tag{26a}$$

$$\rho_e^* = 0, \qquad \forall e \in \mathcal{E} \setminus \left(\cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)} \right), \tag{26b}$$

$$\sigma_a^*(e) = \frac{p_d}{C_{(k)} - C_\emptyset}, \qquad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, j-1 \tag{26c}$$

$$0 \le \sigma_a^*(e) \le \frac{p_d}{C_{(j)} - C_\emptyset}, \qquad \forall e \in \bar{\mathcal{E}}_{(j)}, \tag{26d}$$

$$\sum_{e \in \bar{\mathcal{E}}_{(j)}} \sigma_a^*(e) = 1 - \sum_{k=1}^{j-1} \frac{p_d \cdot E_{(k)}}{C_{(k)} - C_\emptyset}. \tag{26e}$$

Let us discuss the intuition behind the proof of Theorem 1.

Recall from Proposition 2 and Lemma 3 that the set of attacker's equilibrium strategies Σ_a^* is the set of feasible mixed strategies that maximizes $V(\sigma_a)$ in (15), and the attacker never targets at any facility $e \in \mathcal{E}$ with probability higher than the threshold $p_d/(C_e - C_\emptyset)$. Also recall that the costs $\{C_{(k)}\}_{k=1}^K$ are ordered according to (8). Thus, in equilibrium, the attacker targets the facilities in $\bar{\mathcal{E}}_{(k)}$ with the threshold attack probability starting from $k = 1$ and proceeding to $k = 2, 3, \ldots K$ until either all the vulnerable facilities are targeted with the threshold attack probability (and no attack is chosen with remaining probability), or the total attack probability reaches 1.

Again, from Lemma 3, we know that the defender secures the set of facilities that are targeted with the threshold attack probability with positive effort. The equilibrium level of security effort ensures that the attacker gets an identical utility in choosing any pure strategy in the support of σ_a^*, and this utility is higher or equal to that of choosing any other pure strategy.

The distinctions between the two regime types are summarized as follows:

(1) In type I regimes, the defense cost $p_d < \bar{p}_d(p_a)$. The defender secures all vulnerable facilities with a positive level of effort. The attacker targets at each vulnerable facility with the threshold attack probability, and the total probability of attack is less than 1.
(2) In type II regimes, the defense cost $p_d > \bar{p}_d(p_a)$. The defender only secures a subset of targeted facilities with positive level of security effort. The attacker chooses the facilities in decreasing order of $C_e - C_\emptyset$ and targets each of them with the threshold probability until the attack resource is exhausted, i.e., the total probability of attack is 1.

5 Sequential Game $\widetilde{\Gamma}$

In this section, we characterize the set of SPEs in the game $\widetilde{\Gamma}$ for any given attack and defense cost parameters. The sequential game $\widetilde{\Gamma}$ is no longer strategically equivalent to a zero-sum game. Hence, the proof technique we used for equilibrium characterization in game Γ does not work for the game $\widetilde{\Gamma}$. In Sect. 5.1, we analyze the attacker's best response to the defender's security effort vector. We also identify a threshold level of security effort which determines whether or not the defender achieves full attack deterrence in equilibrium. In Sect. 5.2, we present the equilibrium regimes which govern the qualitative properties of SPE.

5.1 Properties of SPE

By definition of SPE, for any security effort vector $\tilde{\rho} \in [0,1]^{|\mathcal{E}|}$ chosen by the defender in the first stage, the attacker's equilibrium strategy in the second stage is a best response to $\tilde{\rho}$, i.e., $\widetilde{\sigma}_a^*(\tilde{\rho})$ satisfies (3b). As we describe next, the properties of SPE crucially depend on a threshold security effort level defined as follows:

$$\widehat{\rho}_e \triangleq \frac{C_e - p_a - C_\emptyset}{C_e - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}. \tag{27}$$

The following lemma presents the best response correspondence $BR(\tilde{\rho})$ of the attacker:

Lemma 4 *Given any $\tilde{\rho} \in [0,1]^{|\mathcal{E}|}$, if $\tilde{\rho}$ satisfies $\tilde{\rho}_e \geq \widehat{\rho}_e$, for all $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, then $BR(\tilde{\rho}) = \Delta(\bar{\mathcal{E}}^* \cup \{\emptyset\})$, where:*

$$\bar{\mathcal{E}}^* \triangleq \left\{ \bar{\mathcal{E}} \,|C_e - p_a > C_\emptyset, \quad \tilde{\rho}_e = \widehat{\rho}_e \right\}. \tag{28}$$

Otherwise, $BR(\tilde{\rho}) = \Delta(\bar{\mathcal{E}}^\diamond)$, where:

$$\bar{\mathcal{E}}^\diamond \triangleq \underset{e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}}{argmax} \{\tilde{\rho}_e C_\emptyset + (1 - \tilde{\rho}_e) C_e\}. \tag{29}$$

In words, if each vulnerable facility e is secured with an effort higher or equal to the threshold effort $\widehat{\rho}_e$ in (27), then the attacker's best response is to choose a mixed strategy with support comprised of all vulnerable facilities that are secured with the threshold level of effort (i.e., $\bar{\mathcal{E}}^*$ as defined in (28)) and the no attack action. Otherwise, the support of attacker's strategy is comprised of all vulnerable facilities (pure actions) that maximize the expected usage cost (see (29)). In particular, no attack action is not chosen in attacker's best response.

Now recall that any SPE $(\tilde{\rho}^*, \widetilde{\sigma}_a^*(\tilde{\rho}^*))$ must satisfy both (3a) and (3b). Thus, for an equilibrium security effort $\tilde{\rho}^*$, an attacker's best response $\widetilde{\sigma}_a(\tilde{\rho}^*) \in BR(\tilde{\rho}^*)$ is an equilibrium strategy only if both these constraints are satisfied. The next lemma shows that depending on whether the defender secures each vulnerable facility e with the threshold effort $\widehat{\rho}_e$ or not, the total attack probability in equilibrium is either 0 or 1. Thus, the defender being the first mover determines whether the attacker is fully deterred from conducting an attack or not. Additionally, in SPE, the security effort on each vulnerable facility e is no higher than the threshold effort $\widehat{\rho}_e$, and the security effort on any other edge is 0.

Lemma 5 *Any SPE $(\tilde{\rho}^*, \widetilde{\sigma}_a^*(\tilde{\rho}^*))$ of the game $\widetilde{\Gamma}$ satisfies the following property:*

$$\sum_{e \in \bar{\mathcal{E}}} \widetilde{\sigma}_a^*(e, \tilde{\rho}^*) = \begin{cases} 0, & \text{if } \tilde{\rho}_e^* \geq \widehat{\rho}_e, \quad \forall e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}, \\ 1, & \text{otherwise.} \end{cases}$$

Additionally, for any $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* \leq \widehat{\rho}_e$. *For any* $e \in \mathcal{E} \setminus \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* = 0$.

The proof of this result is based on the analysis of the following three cases:
Case 1 There exists at least one facility $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$ such that $\tilde{\rho}_e^* < \widehat{\rho}_e$. In this case, by applying Lemma 4, we know that $\widetilde{\sigma}_a^*(\tilde{\rho}^*) \in BR(\tilde{\rho}^*) = \Delta(\bar{\mathcal{E}}^\diamond)$, where $\bar{\mathcal{E}}^\diamond$ is defined in (29). Hence, the total attack probability is 1.
Case 2 For any $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* > \widehat{\rho}_e$. In this case, the set $\bar{\mathcal{E}}^*$ defined in (28) is empty. Hence, Lemma 4 shows that the total attack probability is 0.
Case 3 For any $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* \geq \widehat{\rho}_e$, and the set $\bar{\mathcal{E}}^*$ in (28) is non-empty. Again from Lemma 4, we know that $\widetilde{\sigma}_a^*(\tilde{\rho}^*) \in BR(\tilde{\rho}^*) = \Delta(\bar{\mathcal{E}}^* \cup \{\emptyset\})$. Now assume that the attacker chooses to target at least one facility $e \in \bar{\mathcal{E}}^*$ with a positive probability in equilibrium. Then, the defender can deviate by slightly increasing the security effort on each facility in $\bar{\mathcal{E}}^*$. By introducing such a deviation, the defender's security effort satisfies the condition of Case 2, where the total attack probability is 0. Hence, this results in a higher utility for the defender. Therefore, in any SPE $\left(\tilde{\rho}^*, \widetilde{\sigma}_a^*(\tilde{\rho}^*)\right)$, one cannot have a second-stage outcome in which the attacker targets facilities in $\bar{\mathcal{E}}^*$. We can thus conclude that the total attack probability must be 0 in this case.

In both Cases 2 and 3, we say that the attacker is *fully deterred*.

Clearly, these three cases are exhaustive in that they cover all feasible security effort vectors, and hence we can conclude that the total attack probability in equilibrium is either 0 or 1. Additionally, since the attacker is fully deterred when each vulnerable facility is secured with the threshold effort, the defender will not further increase the security effort beyond the threshold effort on any vulnerable facility. That is, only Cases 1 and 3 are possible in equilibrium.

5.2 Characterization of SPE

Recall that in Sect. 4, type I and type II regimes for the game Γ can be distinguished based on a threshold defense cost $\bar{p}_d(p_a)$. It turns out that in $\widetilde{\Gamma}$, there are still $2K + 1$ regimes. Again, each regime denotes distinct ranges of cost parameters and can be categorized either as type $\widetilde{\text{I}}$ or type $\widetilde{\text{II}}$. However, in contrast to Γ, the regime boundaries in this case are more complicated; in particular, they are nonlinear in the cost parameters p_a and p_d.

To introduce the boundary $\widetilde{p}_d(p_a)$, we need to define the function $p_d^{ij}(p_a)$ for each $i = 1, \ldots, K$ and $j = 1, \ldots, i$ as follows:

$$p_d^{ij}(p_a) = \begin{cases} \dfrac{C_{(1)} - C_\emptyset}{\sum_{k=1}^{i} E_{(k)} - \sum_{k=1}^{i} \frac{p_a E_{(k)}}{C_{(k)} - C_\emptyset}}, & \text{if } j = 1, \\ \dfrac{C_{(j)} - C_\emptyset}{(C_{(j)} - C_\emptyset) \cdot \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right) + \sum_{k=j}^{i} E_{(k)} - \sum_{k=1}^{i} \frac{p_a E_{(k)}}{C_{(k)} - C_\emptyset}}, & \text{if } j = 2, \ldots, i. \end{cases} \tag{30}$$

For any $i = 1, \ldots, K$, and any attack cost $C_{(i+1)} - C_\emptyset \leq p_a < C_{(i)} - C_\emptyset$ (or $0 < p_a < C_{(K)} - C_\emptyset$ if $i = K$), the threshold $\widetilde{p}_d(p_a)$ is defined as follows:

$$\widetilde{p}_d(p_a) = \begin{cases} p_d^{ij}(p_a), & \text{if } \frac{\sum_{k=j+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}} \le p_a < \frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}, \text{ and } j = 1, \ldots, i-1, \\ p_d^{ii}(p_a), & \text{if } 0 \le p_a < \frac{E_{(i)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}. \end{cases} \tag{31}$$

Lemma 6 *Given any attack cost* $0 \le p_a < C_{(1)} - C_\emptyset$, *the threshold* $\widetilde{p}_d(p_a)$ *is a strictly increasing and continuous function of* p_a.

Furthermore, for any $0 < p_a < C_{(1)} - C_\emptyset$, $\widetilde{p}_d(p_a) > \bar{p}_d(p_a)$. *If* $p_a = 0$, $\widetilde{p}_d(0) = \bar{p}_d(0)$. *If* $p_a \to C_{(1)} - C_\emptyset$, $\widetilde{p}_d(p_a) \to +\infty$.

Since $\widetilde{p}_d(p_a)$ is a strictly increasing and continuous function of p_a, the inverse function $\widetilde{p}_d^{-1}(p_d)$ is well defined. Now we are ready to formally define the regimes for the game $\widetilde{\Gamma}$:

(1) Type $\widetilde{\text{I}}$ regimes $\widetilde{\Lambda}^i$, $i = 0, \ldots, K$:

– If $i = 0$:

$$p_a > C_{(1)} - C_\emptyset, \text{ and } \quad p_d > 0. \tag{32}$$

– If $i = 1, \ldots, K-1$:

$$C_{(i+1)} - C_\emptyset < p_a < C_{(i)} - C_\emptyset, \text{ and } \quad 0 < p_d < \widetilde{p}_d(p_a). \tag{33}$$

– If $i = K$:

$$0 < p_a < C_{(K)} - C_\emptyset, \text{ and } \quad 0 < p_d < \widetilde{p}_d(p_a). \tag{34}$$

(2) Type $\widetilde{\text{II}}$ regimes $\widetilde{\Lambda}_j$, $j = 1, \ldots, K$:

– If $j = 1$:

$$0 < p_a < \widetilde{p}_d^{-1}(p_d), \text{ and } \quad p_d > \left(\frac{E_{(1)}}{C_{(1)} - C_\emptyset} \right)^{-1} \tag{35}$$

– If $j = 2, \ldots, K$:

$$0 < p_a < \widetilde{p}_d^{-1}(p_d), \text{ and } \quad \left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} < p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} \tag{36}$$

Analogous to the discussion in Sect. 4.2, we say p_d is "relatively low" in type $\widetilde{\text{I}}$ regimes, and "relatively high" in type $\widetilde{\text{II}}$ regimes. We now provide full characterization of SPE in each regime.

Theorem 2 *The defender's equilibrium security effort vector* $\tilde{\rho}^* = \left(\tilde{\rho}_e^*\right)_{e \in \mathcal{E}}$ *is unique in each regime. Specifically, SPE in each regime is as follows:*

(1) *Type* $\widetilde{I}$ *regimes* $\widetilde{\Lambda}^i$:

– *If* $i = 0$,

$$\tilde{\rho}_e^* = 0, \quad \forall e \in \mathcal{E}, \tag{37a}$$

$$\widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}) = 1, \quad \forall \tilde{\rho} \in [0,1]^{|\mathcal{E}|}. \tag{37b}$$

– *If* $i = 1, \ldots, K$,

$$\tilde{\rho}_e^* = \frac{C_{(k)} - p_a - C_\emptyset}{C_{(k)} - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, i, \tag{38a}$$

$$\tilde{\rho}_e^* = 0, \qquad \forall e \in \mathcal{E} \setminus \left(\cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}\right), \tag{38b}$$

$$\tilde{\sigma}_a^*(\emptyset, \tilde{\rho}^*) = 1, \tag{38c}$$

$$\tilde{\sigma}_a^*(\tilde{\rho}) \in BR(\tilde{\rho}), \qquad \forall \tilde{\rho} \in [0,1]^{|\mathcal{E}|} \setminus \tilde{\rho}^*. \tag{38d}$$

(2) *Type* $\widetilde{\mathrm{II}}$ *regimes* $\widetilde{\Lambda}_j$*:*

– *If* $j = 1$,

$$\tilde{\rho}_e^* = 0, \qquad \forall e \in \mathcal{E}, \tag{39a}$$

$$\tilde{\sigma}_a^*(\tilde{\rho}^*) \in \Delta(\bar{\mathcal{E}}_{(1)}), \tag{39b}$$

$$\tilde{\sigma}_a^*(\tilde{\rho}) \in BR(\tilde{\rho}), \quad \forall \tilde{\rho} \in [0,1]^{|\mathcal{E}|} \setminus \tilde{\rho}^*. \tag{39c}$$

– *If* $j = 2, \ldots, K$,

$$\tilde{\rho}_e^* = \frac{C_{(k)} - C_{(j)}}{C_{(k)} - C_\emptyset}, \quad \forall e \in \bar{\mathcal{E}}_{(k)}, \quad \forall k = 1, \ldots, j-1, \tag{40a}$$

$$\tilde{\rho}_e^* = 0, \qquad \forall e \in \mathcal{E} \setminus \left(\cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)}\right), \tag{40b}$$

$$\tilde{\sigma}_a^*(\tilde{\rho}^*) \in \Delta\left(\cup_{k=1}^{j} \bar{\mathcal{E}}_{(k)}\right), \tag{40c}$$

$$\tilde{\sigma}_a^*(\tilde{\rho}) \in BR(\tilde{\rho}), \qquad \forall \tilde{\rho} \in [0,1]^{|\mathcal{E}|} \setminus \tilde{\rho}^*. \tag{40d}$$

In our proof of Theorem 2 (see Appendix C), we take the approach by first constructing a partition of the space $(p_a, p_d) \in \mathbb{R}^2_{>0}$ defined in (52), and then characterizing the SPE for cost parameters in each set in the partition (Lemmas 7–8). Theorem 2 follows directly by regrouping/combining the elements of this partition such that each of the new partitions has qualitatively identical equilibrium strategies.

From the discussion of Lemma 5, we know that only Cases 1 and 3 are possible in equilibrium and that in any SPE, the security effort on each vulnerable facility e is no higher than the threshold effort $\widehat{\rho}_e$. It turns out that for any attack cost, depending on whether the defense cost is lower or higher than the threshold cost $\widetilde{p}_d(p_a)$, the defender either secures each vulnerable facility with the threshold effort given by (31) (type $\widetilde{\mathrm{I}}$ regime), or there is at least one vulnerable facility that is secured with effort strictly less than the threshold (type $\widetilde{\mathrm{II}}$ regimes):

– In type $\widetilde{\mathrm{I}}$ regimes, the defense cost $p_d < \widetilde{p}_d(p_a)$. The defender secures each vulnerable facility with the threshold effort $\widehat{\rho}_e$. The attacker is fully deterred.
– In type $\widetilde{\mathrm{II}}$ regimes, the defense cost $p_d > \widetilde{p}_d(p_a)$. The defender's equilibrium security effort is identical to that in NE of the normal form game Γ. The total attack probability is 1.

6 Comparison of Γ and $\widetilde{\Gamma}$

Section 6.1 deals with the comparison of players' equilibrium utilities in the two games. In Sect. 6.2, we compare the equilibrium regimes and discuss the distinctions in equilibrium

properties of the two games. We identify situations in which the defender has first-mover advantage.

6.1 Comparison of Equilibrium Utilities

The equilibrium utilities in both games are unique and can be directly derived using Theorems 1 and 2. We denote the equilibrium utilities of the defender and attacker in regime Λ^i (resp. Λ_j) as $U_d^{\Lambda^i}$ and $U_a^{\Lambda^i}$ (resp. $U_d^{\Lambda_j}$ and $U_a^{\Lambda_j}$) in Γ, and $\widetilde{U}_d^{\widetilde{\Lambda}^i}$ and $\widetilde{U}_a^{\widetilde{\Lambda}^i}$ (resp. $\widetilde{U}_d^{\widetilde{\Lambda}_j}$ and $\widetilde{U}_a^{\widetilde{\Lambda}_j}$) in regime $\widetilde{\Lambda}^i$ (resp. $\widetilde{\Lambda}_j$) in $\widetilde{\Gamma}$.

Proposition 3 *In both Γ and $\widetilde{\Gamma}$, the equilibrium utilities are unique in each regime. Specifically,*

(a) *Type I ($\widetilde{I}$) regimes Λ^i ($\widetilde{\Lambda}^i$):*

- *If $i = 0$:*

$$U_d^{\Lambda_0} = \widetilde{U}_d^{\widetilde{\Lambda}^0} = -C_\emptyset, \text{ and } \quad U_a^{\Lambda_0} = \widetilde{U}_a^{\widetilde{\Lambda}^0} = C_\emptyset.$$

- *If $i = 1, \ldots, K$:*

$$U_d^{\Lambda^i} = -C_\emptyset - \left(\sum_{k=1}^{i} E_{(k)} \right) p_d, \qquad \text{and} \quad U_a^{\Lambda^i} = C_\emptyset,$$

$$\widetilde{U}_d^{\widetilde{\Lambda}^i} = -C_\emptyset - \left(\sum_{k=1}^{i} \frac{(C_e - p_a - C_\emptyset)\, E_{(k)}}{C_e - C_\emptyset} \right) p_d, \quad \text{and} \quad \widetilde{U}_a^{\widetilde{\Lambda}^i} = C_\emptyset.$$

(b) *Type II ($\widetilde{II}$) regimes Λ_j ($\widetilde{\Lambda}_j$):*

- *If $j = 1$:*

$$U_d^{\Lambda_1} = \widetilde{U}_d^{\widetilde{\Lambda}_1} = -C_{(1)}, \text{ and } \quad U_a^{\Lambda_1} = \widetilde{U}_a^{\widetilde{\Lambda}_1} = C_{(1)} - p_a.$$

- *If $j = 2, \ldots, K$:*

$$U_d^{\Lambda_j} = \widetilde{U}_d^{\widetilde{\Lambda}_j} = -C_{(j)} - \sum_{k=1}^{j-1} \frac{\left(C_{(k)} - C_{(j)}\right) p_d E_{(k)}}{C_{(k)} - C_\emptyset}, \text{ and } \quad U_a^{\Lambda_j} = \widetilde{U}_a^{\widetilde{\Lambda}_j} = C_{(j)} - p_a.$$

From our results so far, we can summarize the similarities between the equilibrium outcomes in Γ and $\widetilde{\Gamma}$. While most of these conclusions are fairly intuitive, the fact that they are common to both game-theoretic models suggests that the timing of defense investments do not play a role as far as these insights are concerned. Firstly, the support of both players equilibrium strategies tends to contain the facilities, whose compromise results in a high usage cost. The defender secures these facilities with a high level of effort in order to reduce the probability with which they are targeted by the attacker. Secondly, the attack and defense costs jointly determine the set of facilities that are targeted or secured in equilibrium. On the one hand, the set of vulnerable facilities increases as the cost of attack decreases. On the other hand, when the cost of defense is sufficiently high, the attacker tends to conduct an attack with probability 1. However, as the defense cost decreases, the attacker randomizes the attack

on a larger set of facilities. Consequently, the defender secures a larger set of facilities with positive effort, and when the cost of defense is sufficiently small, all vulnerable facilities are secured by the defender. Thirdly, each player's equilibrium payoff is non-decreasing in the opponent's cost, and non-increasing in her own cost. Therefore, to increase her equilibrium payoff, each player is better off as her own cost decreases and the opponent's cost increases.

6.2 First-Mover Advantage

We now focus on identifying parameter ranges in which the defender has the first-mover advantage; i.e., the defender in SPE has a strictly higher payoff than in NE. To identify the first-mover advantage, let us recall the expressions of type I regimes for Γ in (18)–(20) and type $\widetilde{\text{I}}$ regimes for $\widetilde{\Gamma}$ in (32)–(34). Also recall that, for any given cost parameters p_a and p_d, the threshold $\bar{p}_d(p_a)$ (resp. $\widetilde{p}_d(p_a)$) determines whether the equilibrium outcome is of type I or type II regime (resp. type $\widetilde{\text{I}}$ or $\widetilde{\text{II}}$ regime) in the game Γ (resp. $\widetilde{\Gamma}$). Furthermore, from Lemma 6, we know that the cost threshold $\bar{p}_d(p_a)$ in Γ is smaller than the threshold $\widetilde{p}_d(p_a)$ in $\widetilde{\Gamma}$. Thus, for all $i = 1, \ldots, K$, the type I regime Λ^i in Γ is a proper subset of the type $\widetilde{\text{I}}$ regime $\widetilde{\Lambda}^i$ in $\widetilde{\Gamma}$. Consequently, if $p_a < C_{(1)} - C_\emptyset$, we can have one of the following three cases:

- $0 < p_d < \bar{p}_d(p_a)$: The defense cost is relatively low in both Γ and $\widetilde{\Gamma}$. We denote the set of (p_a, p_d) that satisfy this condition as L (*low* defense cost). That is,

$$L \triangleq \{(p_a, p_d) \,|0 < p_d < \bar{p}_d(p_a)\} = \cup_{i=1}^{K} \Lambda^i. \tag{41}$$

- $\bar{p}_d(p_a) < p_d < \widetilde{p}_d(p_a)$: The defense cost is relatively high in Γ, but relatively low in $\widetilde{\Gamma}$. We denote the set of (p_a, p_d) that satisfy this condition as M (*medium* defense cost). That is,

$$M \triangleq \{(p_a, p_d) \,|\bar{p}_d(p_a) < p_d < \widetilde{p}_d(p_a)\} = \cup_{i=1}^{K} \left(\widetilde{\Lambda}^i \setminus \Lambda^i\right). \tag{42}$$

- $p_d > \widetilde{p}_d(p_a)$: The defense cost is relatively high in both Γ and $\widetilde{\Gamma}$. We denote the set of (p_a, p_d) that satisfy this condition as H (*high* defense cost). That is,

$$H \triangleq \{(p_a, p_d) \,|p_d > \widetilde{p}_d(p_a)\} = \cup_{j=1}^{K} \widetilde{\Lambda}_j.$$

We next compare the properties of NE and SPE for cost parameters in each set based on Theorems 1 and 2, and Propositions 3.

- Set L:
  Attacker In Γ, the total attack probability is nonzero but smaller than 1, whereas in $\widetilde{\Gamma}$, the attacker is fully deterred. The attacker's equilibrium utility is identical in both games, i.e., $U_a = \widetilde{U}_a$.
  Defender The defender chooses identical equilibrium security effort in both games, i.e., $\rho^* = \tilde{\rho}^*$, but obtains a higher utility in $\widetilde{\Gamma}$ in comparison with that in Γ, i.e., $U_d < \widetilde{U}_d$.
- Set M:
  Attacker In Γ, the attacker conducts an attack with probability 1, whereas in $\widetilde{\Gamma}$ the attacker is fully deterred. The attacker's equilibrium utility is lower in $\widetilde{\Gamma}$ in comparison with that in Γ, i.e., $U_a > \widetilde{U}_a$.
  Defender The defender secures each vulnerable facility with a strictly higher level of effort in $\widetilde{\Gamma}$ than in Γ, i.e., $\tilde{\rho}_e^* > \rho_e^*$ for each vulnerable facility $e \in \{\mathcal{E} | C_e - p_a > C_\emptyset\}$. The defender's equilibrium utility is higher in $\widetilde{\Gamma}$ in comparison with that in Γ, i.e., $U_d < \widetilde{U}_d$.

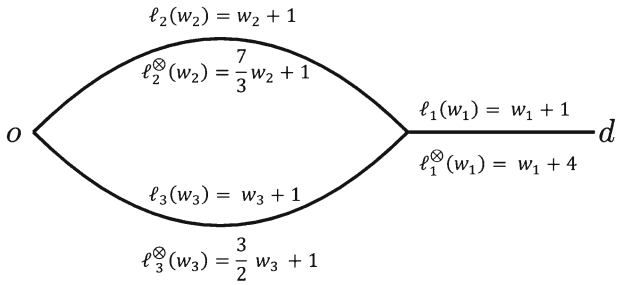

Fig. 1 Three-edge network

– Set H:
Attacker In both games, the attacker conducts an attack with probability 1 and obtains identical utilities, i.e. $U_a = \widetilde{U}_a$.
Defender The defender chooses identical equilibrium security effort in both games, i.e., $\rho^* = \tilde{\rho}^*$, and obtains identical utilities, i.e., $U_d = \widetilde{U}_d$.

Importantly, the key difference between NE and SPE comes from the fact that in $\widetilde{\Gamma}$, the defender as the leading player is able to influence the attacker's strategy in her favor. Hence, when the defense cost is relatively medium or low (both sets M and L), the defender can proactively secure all vulnerable facilities with the threshold effort to fully deter the attack, which results in a higher defender utility in $\widetilde{\Gamma}$ than in Γ. Thus, we say the defender has the first-mover advantage when the cost parameters lie in the set M or L. However, the reason behind the first-mover advantage differs in each set:

– In set M, the defender needs to proactively secure all vulnerable facilities with strictly higher effort in $\widetilde{\Gamma}$ than that in Γ to fully deter the attacker.
– In set L, the defender secures facilities in $\widetilde{\Gamma}$ with the same level of effort as that in Γ, and the attacker is still deterred with probability 1.

On the other hand, in set H, the defense cost is so high that the defender is not able to secure all targeted facilities with an adequately high level of security effort. Thus, the attacker conducts an attack with probability 1 in both games, and the defender no longer has first-mover advantage.

Finally, for the sake of illustration, we compute the parameter sets L, M, and H for transportation network with three facilities (edges); see Fig. 1. If an edge $e \in \mathcal{E}$ is not damaged, then the cost function is $\ell_e(w_e)$, which increases in the edge load w_e. If edge e is successfully compromised by the attacker, then the cost function changes to $\ell_e^{\otimes}(w_e)$, which is higher than $\ell_e(w_e)$ for any edge load $w_e > 0$. The network faces a set of non-atomic travelers with total demand $D = 10$. We define the usage cost in this case as the average cost of travelers in Wardrop equilibrium [21]. Therefore, the usage costs corresponding to attacks to different edges are $C_1 = 20$, $C_2 = 19$, $C_3 = 18$ and the pre-attack usage cost is $C_\emptyset = 17$. From (8), $K = 3$, and $\bar{\mathcal{E}}_{(1)} = \{e_1\}$, $\bar{\mathcal{E}}_{(2)} = \{e_2\}$ and $\bar{\mathcal{E}}_{(3)} = \{e_3\}$. In Fig. 2, we illustrate the regimes of both Γ and $\widetilde{\Gamma}$, and the three sets H, M, and L distinguished by the thresholds $\bar{p}_d(p_a)$ and $\widetilde{p}_d(p_a)$.

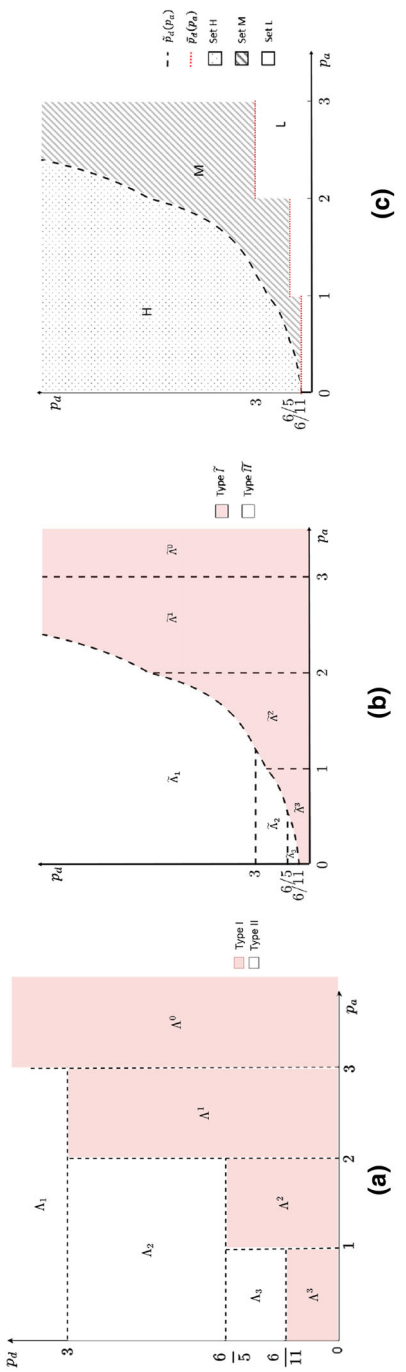

Fig. 2 **a** Regimes of NE in Γ. **b** Regimes of SPE in $\widetilde{\Gamma}$. **c** Comparison of NE and SPE

7 Model Extensions and Dynamic Aspects

In this section, we first discuss how relaxing our modeling assumptions influences our main results. Next, we introduce a dynamic setup in which the users of the infrastructure system face uncertainty about the outcome of attacker–defender interaction (i.e., identity of the compromised facility) and follow a repeated learning procedure to make their usage decisions.

7.1 Relaxing Model Assumptions

Our discussion centers around extending our results when the following modeling aspects are included: facility-dependent cost parameters, less than perfect defense, and attacker's ability to target multiple facilities.

(1) *Facility-dependent attack and defense costs.*
Our techniques for equilibrium characterization of games Γ and $\widetilde{\Gamma}$—as presented in Sects. 4 and 5, respectively—can be generalized to the case when attack/defense costs are non-homogeneous across facilities. We denote the attack (resp. defense) cost for facility $e \in \mathcal{E}$ as $p_{a,e}$ (resp. $p_{d,e}$). However, an explicit characterization of equilibrium regimes in each game can be quite complicated due to the multidimensional nature of cost parameters.
In normal form game Γ, it is easy to show that the attacker's best response correspondence in Lemma 3 holds except that the threshold attack probability for any facility $e \in \bar{\mathcal{E}}$ now becomes $p_{d,e}/(C_e - C_\emptyset)$. The set of vulnerable facilities is given by $\{\mathcal{E} | C_e - p_{a,e} > C_\emptyset\}$. The attacker's equilibrium strategy is to order the facilities in decreasing order of $C_e - p_{a,e}$ and target the facilities in this order each with the threshold probability until either all vulnerable facilities are targeted or the total probability of attack reaches 1. As in Theorem 1, the former case happens when the cost parameters lie in a type I regime, and the latter case happens for type II regimes, although the regime boundaries are more complicated to describe. In equilibrium, the defender chooses the security effort vector to ensure that the attacker is indifferent among choosing any of the pure actions that are in the support of equilibrium attack strategy.
In the sequential game $\widetilde{\Gamma}$, Lemmas 4 and 5 can be extended in a straightforward manner except that the threshold security effort for any vulnerable facility $e \in \{\mathcal{E} | C_e - C_\emptyset > p_{a,e}\}$ is given by $\widehat{\rho}_e = (C_e - p_{a,e} - C_\emptyset)/(C_e - C_\emptyset)$. The SPE for this general case can be obtained analogously to Theorem 2, i.e., comparing the defender's utility of either securing all vulnerable facilities with the threshold effort to fully deter the attack, or choosing a strategy that is identical to that in Γ. These cases happen when the cost parameters lie in (suitably defined) Type $\widetilde{\text{I}}$ and Type $\widetilde{\text{II}}$ regimes, respectively. The main conclusion of our analysis also holds: The defender obtains a higher utility by proactively defending all vulnerable facilities when the facility-dependent cost parameters lie in type $\widetilde{\text{I}}$ regimes.

(2) *Less than perfect defense in addition to facility-dependent cost parameters.*
Now consider that the defense on each facility is only successful with probability $\gamma \in (0, 1)$, which is an exogenous technological parameter. For any security effort vector ρ, the actual probability that a facility e is not compromised when targeted by the attacker is $\gamma\rho_e$. Again our results on NE and SPE in Sects. 4, 5 can be readily extended to this case. However, the expressions for thresholds for attack probability and security effort level need to be modified. In particular, for Γ, in Lemma 3, the threshold attack probability on any facility $e \in \bar{\mathcal{E}}$ is $p_{d,e}/\gamma(C_e - C_\emptyset)$. For $\widetilde{\Gamma}$, the threshold security effort $\widehat{\rho}_e$ for any

vulnerable facility $e \in \{\mathcal{E} | C_e - C_\emptyset > p_{d,e}\}$ is $(C_e - p_{a,e} - C_\emptyset)/\gamma(C_e - C_\emptyset)$. If this threshold is higher than 1 for a particular facility, then the defender is not able to deter the attacker from targeting it.

(3) *Attacker's ability to target multiple facilities.*
If the attacker is not constrained to targeting a single facility, his pure strategy set would be $S_a = 2^{\mathcal{E}}$. Then, for a pure strategy profile (s_d, s_a), the set of compromised facilities is given by $s_a \setminus s_d$, and the usage cost is $C_{s_a \setminus s_d}$. Unfortunately, our approach cannot be straightforwardly applied to this case. This is because the mixed strategies cannot be equivalently represented as probability vectors with elements representing the probability of each facility being targeted or secured. In fact, for a given attacker's strategy, one can find two feasible defender's mixed strategies that induce an identical security effort vector, but result in different players utilities. Hence, the problem of characterizing defender's equilibrium strategies cannot be reduced to characterizing the equilibrium security effort on each facility. Instead, one would need to account for the attack/defense probabilities on all the subsets of facilities in $\mathcal{E}$. This problem is beyond the scope of our paper, although a related work [22] has made some progress in this regard.

Finally, we briefly comment on the model where all the three aspects are included. So long as players' strategy sets are comprised of mixed strategies, the defender's equilibrium utility in $\widetilde{\Gamma}$ must be higher or equal to that in Γ. This is because in $\widetilde{\Gamma}$, the defender can always choose the same strategy as that in NE to achieve a utility that is no less than that in Γ. Moreover, one can show the existence of cost parameters such that the defender has strictly higher equilibrium utility in SPE than in NE. In particular, consider that the attacker's cost parameters $\left(p_{a,e}\right)_{e\in\mathcal{E}}$ satisfy that there is only one vulnerable facility $\bar{e} \in \mathcal{E}$ such that $C_{\bar{e}} - C_\emptyset > p_{a,\bar{e}}$, and the threshold effort on that facility $\widehat{\rho}_{\bar{e}} = \left(C_{\bar{e}} - p_{a,\bar{e}} - C_\emptyset\right)/\gamma(C_{\bar{e}} - C_\emptyset) < 1$. In this case, if the defense cost $p_{d,\bar{e}}$ is sufficiently low, then by proactively securing the facility $\bar{e}$ with the threshold effort $\widehat{\rho}_{\bar{e}}$, the defender can deter the attack completely and obtain a strictly higher utility in $\widetilde{\Gamma}$ than that in Γ. Thus, in this case, the defender gets the first-mover advantage in equilibrium.

7.2 Rational Learning Dynamics

We now discuss an approach for analyzing the dynamics of usage cost after a security attack. Recall that the attacker–defender model enables us to evaluate the vulnerability of individual facilities to a strategic attack for the purpose of prioritizing defense investments. One can view this model as a way to determine the set of possible post-attack states, denoted $s \in S \triangleq \mathcal{E} \cup \{\emptyset\}$. In particular, we consider situations in which the distribution of the system state, denoted $\theta \in \Delta(S)$, is determined by an equilibrium of attacker–defender game (Γ or $\widetilde{\Gamma}$). In Γ, for each $s \in S$, the probability $\theta(s)$ is given as follows:

$$\theta(s) = \begin{cases} \sigma_a^*(e) \cdot (1 - \rho_e^*), & \text{if } s = e, \\ 1 - \sum_{e\in\mathcal{E}} \theta(e), & \text{if } s = \emptyset. \end{cases} \tag{43}$$

For $\widetilde{\Gamma}$, the probability distribution θ can be analogously defined in terms of $\widetilde{\sigma}_a^*$ and $\tilde{\rho}^*$.

Let the realized state be $s = e$; i.e., the facility $e \in \mathcal{E}$ is compromised by the attacker. If this information is known perfectly to all the users immediately after the attack, they can shift their usage choices in accordance with the new state. Then, the cost resulting from the users' choices indeed corresponds to the usage cost $C_s = C_e$, which governs the realized payoffs of both attacker and defender. However, from our results (Theorems 1 and 2), it is apparent that the support of equilibrium player strategies (and hence the support of θ) can

be quite large. Due to inherent limitations in perfectly diagnosing the location of attack, in some situations, the users may not have full knowledge of the realized state. Then, the issues of how users with imperfect information make their decisions in a repeated learning setup and whether or not the long-run usage cost converges to the actual cost C_e become relevant.

To contextualize the above issues, consider the situation in which a transportation system is targeted by an external hacker and that the operation of a single facility is compromised. Furthermore, the nature of attack is such that travelers are not able to immediately know the identity of this facility. This situation can arise when the diagnosis of attack and/or dissemination of information about the attack is imperfect. Examples include cyber-security attacks to transportation facilities that can result in hard-to-detect effects such as compromised traffic signals of a major intersection, or tampering of controllers governing the access to a busy freeway corridor. Then, one can study the problem of learning by rational but imperfectly informed travelers using a repeated routing game model. We now discuss the basic ideas behind the study of this problem. A more rigorous treatment is part of our ongoing work and will be detailed in a subsequent paper.

Let the stages of our repeated routing game be denoted as $t \in T = \{1, 2, \dots\}$. In this game, travelers are imperfectly informed about the network state. In particular, in each stage $t \in T$, they maintain a belief about the state θ^t. The initial belief θ^0 can be different from the prior state distribution θ. However, we require that θ^0 is absolutely continuous with respect to θ [32]:

$$\forall s \in S, \quad \theta(s) > 0, \quad \Rightarrow \quad \theta^0(s) > 0.$$

That is, the initial belief of travelers does not rule out any possible state.

The solution concept we use for this repeated game is *Markov-perfect Equilibrium* (see [37]), in which travelers use routes with the smallest expected cost based on the belief in each stage. Equivalently, the equilibrium routing strategy in stage t is a Wardrop equilibrium of the stage game with belief θ^t [21]. We also consider that at the end of each stage, travelers receive noisy information of the realized costs on routes that are taken. However, no information is available for routes that are not chosen by any traveler. Based on the received information, travelers update their belief of the state using Bayes' rule.

We note that numerous learning schemes have been studied in the literature; for example, fictitious play [15,28,30]; reinforcement learning [10,19,20], and regret minimizations [14, 36]. These learning schemes typically assume that strategies in each stage are determined by a certain function of the history payoff or actions. To explain the learning dynamics in our setup, we consider that in each stage travelers are rational, and they aim to maximize the payoff myopically based on their current belief about other travelers' strategies. The players update their beliefs based on observed actions on the play path. This so-called rational learning dynamics has been investigated in the literature, e.g., [9,27,32,33]. Our model is different from the ones in the literature in that travelers are uncertain about the payoff functions, but correctly anticipate the opponents' strategies. Additionally, the information of the payoff in each stage is noisy and limited (only the realized costs on the taken routes are known).

The game can be understood easily via an example of a transportation network in Fig. 1. In each stage t, travelers with inelastic demand D choose route r_1 $(e_2 - e_1)$ or route r_2 $(e_3 - e_1)$. We denote the equilibrium routing strategy in stage t as $q^{t*}(\theta^t) = \left(q_r^{t*}(\theta^t)\right)_{r \in \{r_1, r_2\}}$, where $q_r^{t*}(\theta^t)$ is the demand of travelers using route r given the belief θ^t. Hence, aggregate flow on edge e_2 (resp. e_3) is $w_2^{t*}(\theta^t) = q_1^{t*}(\theta^t)$ (resp. $w_3^{t*}(\theta^t) = q_2^{t*}(\theta^t)$), and the aggregate flow on edge e_1 is $w_1^{t*}(\theta^t) = D$. Each stage game is a congestion game and hence admits a potential function. The equilibrium routing strategy $q^{t*}(\theta^t)$ can be computed efficiently for this game.

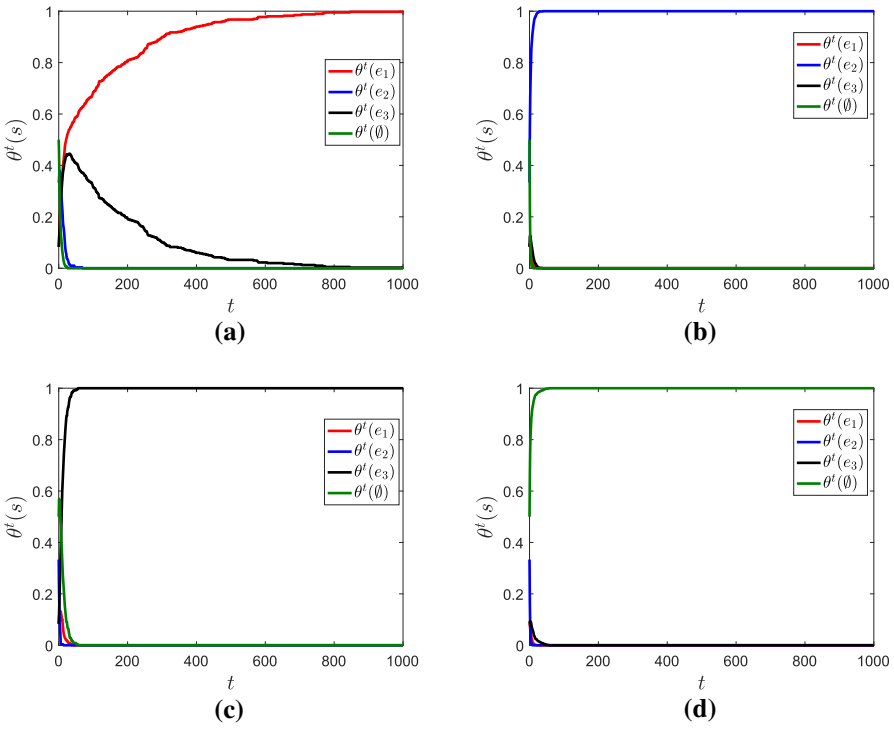

Fig. 3 Rational learning leads to the usage cost of the true state. **a** $s = e_1$. **b** $s = e_2$. **c** $s = e_3$. **d** $s = \emptyset$

Moreover, in each stage, the equilibrium is essentially unique in that the equilibrium edge load is unique for a given belief [44].

The realized cost on each edge $e \in \mathcal{E}$, denoted $c_e^s(w_e^{t*}(\theta^t))$, is equal to the cost (shown in Fig. 1 for the example network) plus a random variable ϵ_e:

$$c_e^s(q^{t*}(\theta^t)) = \begin{cases} \ell_e^{\otimes}(w_e^{t*}(\theta^t)) + \epsilon_e, & \text{if } s = e, \\ \ell_e(w_e^{t*}(\theta^t)) + \epsilon_e, & \text{otherwise.} \end{cases} \tag{44}$$

We illustrate two cases that can arise in rational learning:

– *Long-run usage cost is equal to C_s for any $s \in \{e_1, e_2, e_3, \emptyset\}$.*
Consider the case where the initial belief is $\theta(e_1) = 1/12$, $\theta(e_2) = 1/3$, $\theta(e_3) = 1/12$, $\theta(\emptyset) = 1/2$ (The initial belief can be any probability vector which satisfies the continuity assumption). For any $e \in \mathcal{E}$, the random variable ϵ_e in (44) is distributed as $U[-3, 3]$. The total demand $D = 10$. Figure 3 shows how the belief of each state evolves. We see that eventually travelers learn the true state, and hence the long-run usage cost converges to the actual post-attack usage cost C_s, even though initially all travelers are imperfectly informed about the state.

– *Long-run usage cost is higher than C_s.*
Consider the case when, as a result of attack on edge e_2, the cost function on e_2 changes to $\ell_2^{\otimes}(w_2) = 7/3w_2 + 50$. The total demand is $D = 5$, and the initial belief is $\theta(e_1) = 1/12$, $\theta(e_2) = 1/3, \theta(e_3) = 1/12, \theta(\emptyset) = 1/2$. Starting from this initial belief, travelers exclusively take route r_2, and hence they do not obtain any information about e_2. Even when the realized state is $s = \emptyset$, travelers end up repeatedly taking r_2 as if e_2 is compromised.

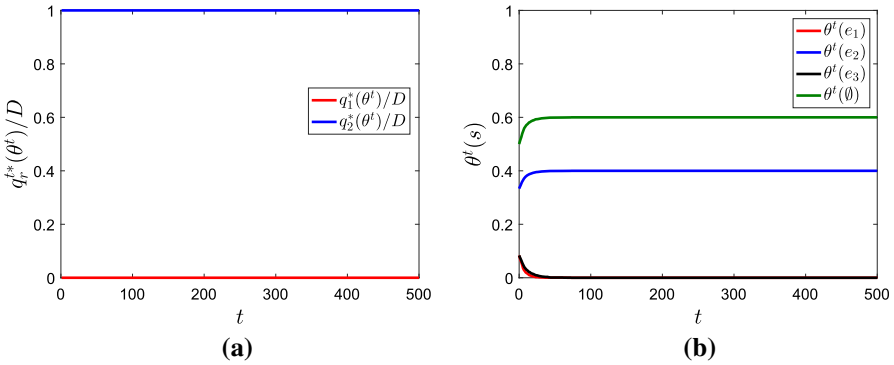

Fig. 4 Learning leads to long-run inefficiency when $\mathbf{s} = \emptyset$: **a** equilibrium routing strategies; **b** beliefs

Thus, the long-run average cost is C_{e_2}, which is higher than the cost corresponding to the true state $C_\emptyset$. Therefore, rational learning dynamics can lead to long-run inefficiency. We illustrate the equilibrium routing strategies and beliefs in each stage in Fig. 4a, b, respectively.

These cases illustrate that if the post-attack state is not perfectly known by the users of the system, then the cost experienced by the users depends on the learning dynamics induced by the repeated play of rational users. Particularly, the learning dynamics can induce a higher usage cost in the long run in comparison with the cost corresponding to the true state. Following previously known results [31], one can argue that if sufficient amount of "off-equilibrium" experiments are conducted by travelers, then the learning will converge to Wardrop equilibrium with the true state. However, such experiments are in general not costless.

As a final remark, we note another implication of proactive defense strategy in ranges of attack/ defense cost parameters where the first-mover advantage holds. In particular, when the cost parameters are in the sets L and M as given in (41)–(42), the attack is completely deterred in the sequential game $\widetilde{\Gamma}$ and there is no uncertainty in the realized state. In such a situation, one does not need to consider uncertainty in the travelers' belief about the true state and the issue of long-run inefficiency due to learning behavior does not arise.

Acknowledgements We are sincerely thankful to Prof. Georges Zaccour and two anonymous referees whose constructive comments helped us to improve our initial manuscript. We thank seminar participants at MIT, HEC Montreal, University of Pennsylvania, and NYU Abu Dhabi for helpful comments. The authors are grateful to Professors Alexandre Bayen, Sanjeev Goyal, Patrick Jaillet, Karl Johansson, Patrick Loiseau, Samer Madanat, Hani Mahmassani, Asu Ozdaglar, Galina Schwartz, Demos Teneketzis, Rakesh Vohra, Dan Work, Georges Zaccour for insightful comments and discussions in the early phase of this research. This work was supported in part by Singapore-MIT Alliance for Research and Technology (SMART) Center for Future Mobility (FM), NSF grant CNS 1239054, NSF CAREER award CNS 1453126.

A Proofs of Section 3

Proof of Lemma 1 We first show that the strategy in (5) is feasible. Since $\rho_{(1)} \leq 1$, and for any $i = 1, \ldots, m-1$, $\rho_{(i)} - \rho_{(i+1)} > 0$, $\sigma_d(s_d)$ is nonnegative for any $s_d \in S_d$. Additionally,

$$\sum_{s_d \in S_d} \sigma_d(s_d) = \sigma_d(\emptyset) + \sum_{i=1}^{m-1} \sigma_d\left(\left\{e \in \mathcal{E} | \rho_e \geq \rho_{(i)}\right\}\right) + \sigma_d\left(\left\{e \in \mathcal{E} | \rho_e \geq \rho_{(m)}\right\}\right)$$
$$= \left(1 - \rho_{(1)}\right) + \sum_{i=1}^{m-1} \left(\rho_{(i)} - \rho_{(i+1)}\right) + \rho_{(m)}$$
$$= 1 - \rho_{(1)} + \rho_{(1)} - \rho_{(m)} + \rho_{(m)}$$
$$= 1.$$

Thus, σ_d in (5) is a feasible strategy of the defender. Now we check that σ_d in (5) indeed induces ρ. Consider any $e \in \mathcal{E}$ such that $\rho_e = 0$. Then, since $e \notin \left\{\mathcal{E} | \rho_e \geq \rho_{(i)}\right\}$ for any $i = 1, \ldots, m$, and $e \notin \emptyset$, for any $s_d \ni e$, we must have $\sigma_d(s_d) = 0$. Thus, $\sum_{s_d \ni e} \sigma_d(s_d) = 0 = \rho_e$. Finally, for any $j = 1, \ldots, m$, consider any $e \in \mathcal{E}$, where $\rho_e = \rho_{(j)}$:

$$\sum_{s_d \ni e} \sigma_d(s_d) = \sum_{i=j}^{m} \sigma_d\left(\left\{e \in \mathcal{E} | \rho_e \geq \rho_{(i)}\right\}\right) = \rho_{(j)}.$$

Therefore, σ_d in (5) induces ρ. □

Proof of Proposition 1 We prove the result by the principal of iterated dominance. We first show that any s_d such that $s_d \nsubseteq \bar{\mathcal{E}}$ is strictly dominated by the strategy $s_d^{'} = s_d \cap \bar{\mathcal{E}}$. Considering any pure strategy of the attacker, $s_a \in \mathcal{E}$, the utilities of the defender with strategy s_d and $s_d^{'}$ are as follows:

$$u_d(s_d, s_a) = -C(s_d, s_a) - |s_d| p_d = -C(s_d, s_a) - (|s_d^{'}| + |s_d \setminus \bar{\mathcal{E}}|) p_d,$$
$$u_d(s_d^{'}, s_a) = -C(s_d^{'}, s_a) - |s_d^{'}| p_d.$$

If $s_a \in \bar{\mathcal{E}}$ or $s_a \notin s_d$ or $s_a = \emptyset$, then $C(s_d, s_a) = C(s_d^{'}, s_a)$, and thus $U_d(s_d, s_a) < U_d(s_d^{'}, s_a)$. If $s_a = e \in s_d \setminus \bar{\mathcal{E}}$, then $e \notin \bar{\mathcal{E}}$, and $C_e \leq C_\emptyset$. We have $C(s_d, s_a) = C_\emptyset \geq C_e = C(s_d^{'}, s_a)$, and thus $U_d(s_d^{'}, s_a) \geq -C(s_d, s_a) - |s_d^{'}| p_d > U_d(s_d, s_a)$. Therefore, any s_d such that $s_d \nsubseteq \bar{\mathcal{E}}$ is a strictly dominated strategy. Hence, in Γ, any equilibrium strategy of the defender satisfies $\sigma_d^*(s_d) = 0$. From (4), we know that $\rho_e^* = 0$ for any $e \in \mathcal{E} \setminus \bar{\mathcal{E}}$.

We denote the set of defender's pure strategies that are not strictly dominated as $\bar{S}_d = \{s_d | s_d \subseteq \bar{\mathcal{E}}\}$. Consider any $s_d \in \bar{S}_d$, we show that any $s_a \in \mathcal{E} \setminus \bar{\mathcal{E}}$ is strictly dominated by strategy $\emptyset$. The utility functions of the attacker with strategy s_a and $\emptyset$ are as follows:

$$u_a(s_d, s_a) = C(s_d, s_a) - p_a,$$
$$u_a(s_d, \emptyset) = C(s_d, \emptyset).$$

Since $s_d \subseteq \bar{\mathcal{E}}$ and $s_a \in \mathcal{E} \setminus \bar{\mathcal{E}}$, $s_a \notin s_d$, thus $C(s_d, s_a) = C_{s_a} \leq C_\emptyset$. However, $C(s_d, \emptyset) = C_\emptyset$ and $p_a > 0$. Therefore, $U_a(s_d, \emptyset) > U_a(s_d, s_a)$. Hence, any $s_a \in \mathcal{E} \setminus \bar{\mathcal{E}}$ is strictly dominated. Hence, in equilibrium, the probability of the attacker choosing facility $e \in \mathcal{E} \setminus \bar{\mathcal{E}}$ is 0 in Γ.

We can analogously argue that in $\widetilde{\Gamma}$, $\tilde{\rho}_e^* = 0$ and $\widetilde{\sigma}_a^*(e, \tilde{\rho}) = 0$ for any $e \in \mathcal{E} \setminus \bar{\mathcal{E}}$. □

B Proofs of Section 4

Proof of Lemma 2 The utility functions of the attacker with strategy σ_a in Γ^0 and Γ are related as follows:

$$U_a^0(\sigma_d, \sigma_a) = U_a(\sigma_d, \sigma_a) + \mathbb{E}_{\sigma_d}[|s_d|] \cdot p_d.$$

Thus, for a given σ_d, any σ_a that maximizes $U_a^0(\sigma_d, \sigma_a)$ also maximizes $U_a(\sigma_d, \sigma_a)$. So the set of best response strategies of the attacker in Γ^0 is identical to that in Γ. Analogously, given any σ_a, the set of best response strategies of the defender in Γ is identical to that in Γ^0. Thus, Γ^0 and Γ are strategically equivalent; i.e., they have the same set of equilibrium strategy profiles. Using the interchangeability property of equilibria in zero-sum games, we directly obtain that for any $\sigma_d^* \in \Sigma_d^*$ and any $\sigma_a^* \in \Sigma_a^*$, (σ_d^*, σ_a^*) is an equilibrium strategy profile. □

Proof of Proposition 2 From Lemma 2, the set of attacker's equilibrium strategies Σ_a^* is the optimal solution of the following maximin problem:

$$\max_{\sigma_a} \quad \min_{s_d \in S_d} \left\{ \sum_{e \in \bar{\mathcal{E}}} (C(s_d, e) + |s_d| p_d - p_a) \cdot \sigma_a(e) + (C(s_d, \emptyset) + |s_d| p_d) \cdot \sigma_a(\emptyset) \right\}$$

$$s.t. \quad \sum_{e \in \bar{\mathcal{E}}} \sigma_a(e) + \sigma_a(\emptyset) = 1, \tag{45a}$$

$$\sigma_a(\emptyset) \geq 0, \quad \sigma_a(e) \geq 0, \quad \forall e \in \bar{\mathcal{E}}. \tag{45b}$$

Given any $s_d \in S_d$, we can express the objective function in (45) as follows:

$$\begin{aligned}
&\sum_{e \in \bar{\mathcal{E}}} (C(s_d, e) + |s_d| p_d - p_a) \cdot \sigma_a(e) + (C(s_d, \emptyset) + |s_d| p_d) \cdot \sigma_a(\emptyset) \\
&= \sum_{e \in \bar{\mathcal{E}}} (C(s_d, e) - p_a) \cdot \sigma_a(e) + C(s_d, \emptyset) \sigma_a(\emptyset) + |s_d| p_d \cdot \left(\sum_{e \in \mathcal{E}} \sigma_a(e) + \sigma_a(\emptyset) \right) \\
&\overset{(45a)}{=} \sum_{e \in \bar{\mathcal{E}}} \sigma_a(e) \cdot (C(s_d, e) - p_a) + |s_d| p_d + \sigma_a(\emptyset) \cdot C_\emptyset \\
&= \sum_{e \in \bar{\mathcal{E}}} \sigma_a(e) \cdot (C(s_d, e) - p_a) + p_d \cdot \left(\sum_{e \in \bar{\mathcal{E}}} \mathbb{1}\{s_d \ni e\} \right) + \sigma_a(\emptyset) \cdot C_\emptyset \\
&= \sum_{e \in \bar{\mathcal{E}}} (\sigma_a(e) \cdot (C(s_d, e) - p_a) + p_d \cdot \mathbb{1}\{s_d \ni e\}) + \sigma_a(\emptyset) \cdot C_\emptyset \\
&\overset{(1)}{=} \sum_{e \in s_d} (\sigma_a(e) \cdot (C_\emptyset - p_a) + p_d) + \sum_{e \in \bar{\mathcal{E}} \setminus s_d} \sigma_a(e) \cdot (C_e - p_a) + \sigma_a(\emptyset) \cdot C_\emptyset.
\end{aligned}$$

Therefore, we can write:

$$\begin{aligned}
&\min_{s_d \in S_d} \left\{ \sum_{e \in \bar{\mathcal{E}}} (C(s_d, e) + |s_d| p_d - p_a) \cdot \sigma_a(e) + (C(s_d, \emptyset) + |s_d| p_d) \cdot \sigma_a(\emptyset) \right\} \\
&= \min_{s_d \in S_d} \left\{ \sum_{e \in s_d} (\sigma_a(e) \cdot (C_\emptyset - p_a) + p_d) + \sum_{e \in \bar{\mathcal{E}} \setminus s_d} \sigma_a(e) \cdot (C_e - p_a) + \sigma_a(\emptyset) \cdot C_\emptyset \right\} \\
&= \sum_{e \in \bar{\mathcal{E}}} \min \{ \sigma_a(e) \cdot (C_\emptyset - p_a) + p_d, \ \sigma_a(e) \cdot (C_e - p_a) \} + \sigma_a(\emptyset) \cdot C_\emptyset \\
&= V(\sigma_a).
\end{aligned}$$

Thus, (45) is equivalent to (10), and Σ_a^* is the optimal solution set of (10)

By introducing an $|\bar{\mathcal{E}}|$-dimensional variable $v = (v_e)_{e\in\bar{\mathcal{E}}}$, (10) can be changed to a linear optimization program (11), and Σ_a^* is the optimal solution set of (11). □

Proof of Lemma 3 We first argue that the defender's best response is in (13). For edge $e \in \mathcal{E}$ such that $\sigma_a(e) < \frac{p_d}{C_e - C_\emptyset}$, we have $(C_\emptyset - C_e)\,\sigma_a(e) + p_d > 0$. Since $\rho \in BR(\sigma_a)$ maximizes $U_d(\sigma_d, \sigma_a)$ as given in (6a), ρ_e must be 0. Additionally, Proposition 1 ensures that for any $e \in \mathcal{E} \setminus \bar{\mathcal{E}}$, ρ_e is 0.

Analogously, if $\sigma_a(e) > \frac{p_d}{C_e - C_\emptyset}$, then $(C_\emptyset - C_e)\,\sigma_a(e) + p_d < 0$, and the best response $\rho_e = 1$. Finally, if $\sigma_a(e) = \frac{p_d}{C_e - C_\emptyset}$, any $\rho_e \in [0, 1]$ can be a best response.

We next prove (14). We show that if a feasible σ_a violates (14a), i.e., there exists a facility, denoted $\bar{e} \in \bar{\mathcal{E}}$ such that $\sigma_a(\bar{e}) > \frac{p_d}{C_{\bar{e}} - C_\emptyset}$, then σ_a cannot be an equilibrium strategy. There are two cases:

(a) There exists another facility $\widehat{e} \in \bar{\mathcal{E}}$ such that $\sigma_a(\widehat{e}) < \frac{p_d}{C_{\widehat{e}} - C_\emptyset}$. Consider an attacker's strategy $\sigma_a^{'}$ defined as follows:

$$\begin{aligned}
&\sigma_a^{'}(e) = \sigma_a(e), \quad \forall e \in \bar{\mathcal{E}} \setminus \{\bar{e}, \widehat{e}\}, \quad \sigma_a^{'}(\emptyset) = \sigma_a(\emptyset),\\
&\sigma_a^{'}(\bar{e}) = \sigma_a(\bar{e}) - \epsilon,\\
&\sigma_a^{'}(\widehat{e}) = \sigma_a(\widehat{e}) + \epsilon,
\end{aligned}$$

where ϵ is a sufficiently small positive number so that $\sigma_a^{'}(\bar{e}) > \frac{p_d}{C_{\bar{e}} - C_\emptyset}$ and $\sigma_a^{'}(\widehat{e}) < \frac{p_d}{C_{\widehat{e}} - C_\emptyset}$. We obtain:

$$V(\sigma_a^{'}) - V(\sigma_a) = \epsilon\,(C_{\widehat{e}} - C_\emptyset) > 0$$

The last inequality holds from (7a) and $\widehat{e} \in \bar{\mathcal{E}}$. Therefore, σ_a cannot be an attacker's equilibrium strategy.

(b) If there does not exist such $\bar{e}$ as defined in case (a), then for any $e \in \bar{\mathcal{E}}$, we have $\sigma_a(e) \geq \frac{p_d}{C_e - C_\emptyset}$. Now consider $\sigma_a^{'}$ as follows:

$$\begin{aligned}
&\sigma_a^{'}(e) = \sigma_a(e), \quad \forall e \in \mathcal{E} \setminus \{\bar{e}\},\\
&\sigma_a^{'}(\bar{e}) = \sigma_a(\bar{e}) - \epsilon,\\
&\sigma_a^{'}(\emptyset) = \sigma_a(\emptyset) + \epsilon,
\end{aligned}$$

where ϵ is a sufficiently small positive number so that $\sigma_a^{'}(\bar{e}) > \frac{p_d}{C_{\bar{e}} - C_\emptyset}$. We obtain:

$$V(\sigma_a^{'}) - V(\sigma_a) = \epsilon\,(C_\emptyset - (C_\emptyset - p_a)) = \epsilon\, p_a > 0.$$

Therefore, σ_a also cannot be an attacker's equilibrium strategy.

Thus, we can conclude from cases (a) and (b) that in equilibrium σ_a^* must satisfy (14a). Additionally, from Proposition 1, (14b) is also satisfied. □

Proof of Theorem 1 We first prove the attacker's equilibrium strategies in each regime. From Proposition 2 and Lemma 3, we know that σ_a^* maximizes $V(\sigma_a)$, which can be equivalently re-written as in (15). We analyze the attacker's equilibrium strategy set in each regime subsequently:

(a) Type I regimes Λ^i:

– $i = 0$:
Since $p_a > C_{(1)} - C_\emptyset$, we must have $C_\emptyset > C_e - p_a$ for any $e \in \bar{\mathcal{E}}$. There is no vulnerable facility, and thus $\sigma_a^*(\emptyset) = 1$.
– $i = 1, \ldots, K$:
Since p_d satisfies (19) or (20), we obtain:

$$\sum_{e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}} \frac{p_d}{C_e - C_\emptyset} = \sum_{k=1}^{i} \frac{p_d \cdot E_{(k)}}{C_{(k)} - C_\emptyset} < 1 \tag{46}$$

Therefore, the set of feasible attack strategies satisfying (24c)–(24d) is a non-empty set. We also know from Lemma 3 that σ_a^* satisfies (14a). Again from (19) or (20), for any $k = 1, \ldots, i$, we have $C_{(k)} - p_a > C_\emptyset$ and for any $k = i+1, \ldots, K$, we have $C_{(k)} - p_a < C_\emptyset$. Since $\{C_{(k)}\}_{k=1}^{K}$ satisfy (8), to maximize $V(\sigma_a)$ in (15), the optimal solution must satisfy (24c)–(24d).

(b) Type II regimes Λ_j:

– $j = 1$: From (21), we know that:

$$1 = \sum_{e \in \bar{\mathcal{E}}_{(1)}} \sigma_a^*(e) < \frac{p_d E_{(1)}}{C_{(1)} - C_\emptyset}. \tag{47}$$

Thus, the set of feasible attack strategies satisfying (25b)–(25c) is a non-empty set. Additionally, from Lemma 3, we know that σ_a^* satisfies (25b). Since $C_{(1)} > C_{(k)}$ for any $k = 2, \ldots, K$, and $C_{(1)} - p_a > C_\emptyset$. From (15) and (47), we know that in equilibrium the attacker targets facilities in $\bar{\mathcal{E}}_{(1)}$ with probability 1. The set of strategies satisfying (25b)–(25c) maximizes (15) and thus is the set of attacker's equilibrium strategies.
– $j = 2, \ldots, K$: From (22), we know that:

$$0 < 1 - \sum_{k=1}^{j-1} \frac{p_d \cdot E_{(k)}}{C_{(k)} - C_\emptyset} < \frac{p_d \cdot E_{(j)}}{C_{(j)} - C_\emptyset}.$$

Thus, the set of feasible attack strategies satisfying (26c)–(26e) is a non-empty set. From Lemma 3, we know that σ_a^* satisfies (26d). Since $\{C_{(k)}\}_{k=1,\ldots,j}$ satisfies the ordering in (8), in order to maximize $V(\sigma_a)$ in (15), σ_a^* must also satisfy (26c) and (26e), and the remaining facilities are not targeted.

We next prove the defender's equilibrium security effort. By definition of Nash equilibrium, the probability vector ρ^* is induced by an equilibrium strategy if and only if it satisfies the following two conditions:

(1) ρ^* is a best response to any $\sigma_a^* \in \Sigma_a^*$.
(2) Any attacker's equilibrium strategy is a best response to ρ^*; i.e., the attacker has identical utilities for choosing any pure strategy in his equilibrium support set, and the utility is no less than that of any other pure strategies.

Note that in both conditions, we require ρ^* to be a best response to *any* attacker's equilibrium strategy. This is because given any $\sigma_a^* \in \Sigma_a^*$, (ρ^*, σ_a^*) is an equilibrium strategy profile (Lemma 2). We now check these conditions in each regime:

(a) Type I regimes Λ^i:

– If $i = 0$:
Since $\sigma_a^*(e) = 0$ for any $e \in \mathcal{E}$. From Lemma 3, the best response of the defender is $\rho_e^* = 0$ for any $e \in \mathcal{E}$.
– If $i = 1, \ldots, K$:
From Lemma 3, we know that $\rho_e^* = 0$ for any $e \in \mathcal{E} \setminus \left(\cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}\right)$. Since $\sigma_a^*(\emptyset) > 0$, ρ_e^* must ensure that the attacker's utility of choosing any facility $e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$ is identical to that of choosing no attack $\emptyset$. Consider any $e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$:

$$
\begin{aligned}
& U_a(\rho^*, e) = U_a(\rho^*, \emptyset), \\
\overset{(6b)}{\Rightarrow} \quad & \rho_e^* (C_\emptyset - p_a) + (1 - \rho_e^*)(C_e - p_a) = C_\emptyset, \\
\Rightarrow \quad & \rho_e^* = \frac{C_e - p_a - C_\emptyset}{C_e - C_\emptyset}, \quad \forall e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}.
\end{aligned}
$$

For any $\bar{e} \in \mathcal{E} \setminus \left(\cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}\right)$, since $\rho_{\bar{e}}^* = 0$, the attacker receives utility $C_{\bar{e}} - p_a$ by targeting $\bar{e}$, which is lower than $C_\emptyset$. Therefore, ρ^* in (24a)–(24b) satisfies both conditions (1) and (2). ρ^* is the unique equilibrium strategy.

(b) Type II regimes Λ_j:

– If $j = 0$:
Consider an attacker's strategy σ_a such that:

$$
\begin{aligned}
& \sigma_a(e) = \frac{1}{E_{(1)}}, \quad \forall e \in \bar{\mathcal{E}}_{(1)}, \\
& \sigma_a(e) = 0, \quad \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}_{(1)}.
\end{aligned}
$$

Since p_d satisfies (21), we know that $\frac{1}{E_{(1)}} < \frac{p_d}{C_{(1)} - C_\emptyset}$. One can check that σ_a satisfies (25b)–(25c), and thus $\sigma_a \in \Sigma_a^*$. Therefore, we know from Lemma 3 that $\rho_e^* = 0$ for any $e \in \mathcal{E}$.
– If $j = 1, \ldots, K$:
Analogous to our discussion for $j = 0$, the following is an equilibrium strategy of the attacker:

$$
\begin{aligned}
& \sigma_a^*(e) = \frac{p_d}{C_e - C_\emptyset}, \quad \forall e \in \cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)}, \\
& \sigma_a^*(e) = \frac{1}{E_{(j)}} \left(1 - \sum_{i=1}^{j-1} \frac{p_d E_{(k)}}{C_{(k)} - C_\emptyset}\right), \quad \forall e \in \bar{\mathcal{E}}_{(j)}, \\
& \sigma_a^*(e) = 0, \quad \forall e \in \mathcal{E} \setminus \left(\cup_{k=1}^{j} \bar{\mathcal{E}}_{(k)}\right).
\end{aligned}
$$

From Lemma 3, we immediately obtain that $\rho_e^* = 0$ for any $e \in \mathcal{E} \setminus \left(\cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)}\right)$.
Furthermore, for any $e \in \cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)}$, the utility of the attacker in choosing e must be the same as the utility for choosing any facility in $\bar{\mathcal{E}}_{(j)}$, which is $C_{(j)} - p_a$. Therefore,

for any $e \in \cup_{k=1}^{j-1} \bar{\mathcal{E}}_{(k)}$, ρ^* satisfies:

$$\begin{aligned} & U_a(\rho^*, e) = C_{(j)} - p_a, \\ \overset{(6b)}{\Rightarrow} \quad & \rho_e^* \left(C_\emptyset - p_a\right) + \left(1 - \rho_e^*\right)\left(C_{(k)} - p_a\right) = C_{(j)} - p_a, \\ \Rightarrow \quad & \rho_e^* = \frac{C_{(k)} - C_{(j)}}{C_{(k)} - C_\emptyset}. \end{aligned}$$

Additionally, for any $e \in \mathcal{E} \setminus \left(\cup_{k=1}^{j} \bar{\mathcal{E}}_{(k)}\right)$, the utility for the attacker targeting e is $C_e - p_a$, which is smaller than $C_{(j)} - p_a$. Thus, both conditions (1) and (2) are satisfied. ρ^* is the unique equilibrium security effort. □

C Proofs of Section 5

Proof of Lemma 4 For any non-vulnerable facility e, the best response strategy $\widetilde{\sigma}_a$ must be such that $\widetilde{\sigma}_a(e, \tilde{\rho}) = 0$ for any $\tilde{\rho}$.

Now consider any $e \in \{\mathcal{E} | C_e - p_a > C_\emptyset\}$. If $\tilde{\rho}_e > \widehat{\rho}_e$, then we can write:

$$U_a(\tilde{\rho}, e) = \tilde{\rho}_e C_\emptyset + (1 - \tilde{\rho}_e) C_e - p_a < C_\emptyset = U_a(\tilde{\rho}, \emptyset). \tag{48}$$

That is, the attacker's expected utility of targeting the facility e is less than the expected utility of no attack. Thus, in any attacker's best response, $\widetilde{\sigma}_a(e, \tilde{\rho}) = 0$ for any such facility e. Additionally, if $\tilde{\rho}_e = \widehat{\rho}_e$, then $U_a(e, \tilde{\rho}) = U_a(\emptyset, \tilde{\rho})$; i.e., the utility of targeting such facility is identical with the utility of choosing no attack and is higher than that of any other pure strategies. Hence, the set of best response strategies of the attacker is $\Delta(\bar{\mathcal{E}}^* \cup \{\emptyset\})$, where $\bar{\mathcal{E}}^*$ is the set defined in (28).

Otherwise, if there exists a facility $e \in \{\mathcal{E} | C_e - p_a > C_\emptyset\}$ such that $\tilde{\rho}_e < \widehat{\rho}_e$, then we obtain:

$$U_a(\tilde{\rho}, e) = \tilde{\rho}_e C_\emptyset + (1 - \tilde{\rho}_e) C_e - p_a > C_\emptyset = U_a(\tilde{\rho}, \emptyset).$$

Thus, no attack cannot be chosen in any best response strategy, which implies that the attacker chooses to attack with probability 1. Finally, $\bar{\mathcal{E}}^\diamond$ is the set of facilities which incur the highest expected utility for the attacker given $\tilde{\rho}$, thus $BR(\tilde{\rho}) = \Delta(\bar{\mathcal{E}}^\diamond)$. □

Proof of Lemma 5 We first prove that the total attack probability is either 0 or 1 in any SPE. We discuss the following three cases separately:

- There exists at least one single facility $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$ such that $\tilde{\rho}_e^* < \widehat{\rho}_e$. Since $\widetilde{\sigma}_a^*(\tilde{\rho}^*) \in BR(\tilde{\rho}^*)$, from Lemma 4, we know that $\sum_{e \in \bar{\mathcal{E}}} \widetilde{\sigma}_a^*(e, \tilde{\rho}^*) = 1$.
- For all $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* > \widehat{\rho}_e$, i.e., the set $\bar{\mathcal{E}}^*$ in (28) is empty. Since $\widetilde{\sigma}_a^*(\tilde{\rho}^*) \in BR(\tilde{\rho}^*)$, from Lemma 4, we know that no edge is targeted in SPE, i.e., $\sum_{e \in \bar{\mathcal{E}}} \widetilde{\sigma}_a^*(e, \tilde{\rho}^*) = 0$.
- For all $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e^* \geq \widehat{\rho}_e$, and the set $\bar{\mathcal{E}}^*$ in (28) is non-empty. For the sake of contradiction, we assume that in SPE, there exists a facility $e \in \bar{\mathcal{E}}^*$ such that $\widetilde{\sigma}_a^*(e, \tilde{\rho}^*) > 0$, i.e. $\widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}^*) < 1$. Then, we can write $U_d(\tilde{\rho}^*, \widetilde{\sigma}_a^*(\tilde{\rho}^*))$ as follows:

$$U_d(\tilde{\rho}^*, \widetilde{\sigma}_a^*(\tilde{\rho}^*)) = -C_\emptyset - (1 - \widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}^*)) p_a - \left(\sum_{e \in \bar{\mathcal{E}}} \tilde{\rho}_e^*\right) p_d. \tag{49}$$

Now, consider $\tilde{\rho}'$ as follows:

$$\tilde{\rho}'_e = \tilde{\rho}^*_e + \epsilon > \widehat{\rho}_e, \qquad \forall e \in \bar{\mathcal{E}}^*,$$
$$\tilde{\rho}'_e = \tilde{\rho}^*_e = 0, \qquad \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}^*,$$

where ϵ is a sufficiently small positive number. Given such a $\tilde{\rho}'$, we know from Lemma 4 that the unique best response is $\widetilde{\sigma}_a(\emptyset, \tilde{\rho}') = 1$. Therefore, the defender's utility is given by:

$$U_d(\tilde{\rho}', \widetilde{\sigma}_a(\tilde{\rho}')) = -C_\emptyset - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}'_e\right) p_d.$$

Additionally,

$$U_d(\tilde{\rho}', \widetilde{\sigma}_a(\tilde{\rho}')) - U_d(\tilde{\rho}^*, \widetilde{\sigma}_a(\tilde{\rho}^*)) = (1 - \widetilde{\sigma}_a(\emptyset, \tilde{\rho}^*))p_a - \epsilon p_d|\bar{\mathcal{E}}^*|.$$

Since ϵ is sufficiently small and $\widetilde{\sigma}_a(\emptyset, \tilde{\rho}^*) < 1$, we obtain that $U_d(\tilde{\rho}', \widetilde{\sigma}_a(\tilde{\rho}')) > U_d(\tilde{\rho}^*, \widetilde{\sigma}_a(\tilde{\rho}^*))$. Therefore, $\tilde{\rho}^*$ cannot be a SPE. We can conclude that in this case, the attacker chooses not to attack with probability 1.

We next show that in any SPE, the defender's security effort on each vulnerable facility e is no higher than the threshold $\widehat{\rho}_e$ defined in (27). Assume for the sake of contradiction that there exists a facility $\bar{e} \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$ such that $\tilde{\rho}_{\bar{e}} > \widehat{\rho}_{\bar{e}}$. We discuss the following two cases separately:

- The set $\widehat{e} \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset, \tilde{\rho}_e < \widehat{\rho}_e\}$ is non-empty. We know from Lemma 4 that $BR(\tilde{\rho}) = \Delta(\bar{\mathcal{E}}^\diamond)$, where the set $\bar{\mathcal{E}}^\diamond$ in (29) is the set of facilities which incur the highest utility for the attacker. Clearly, $\bar{\mathcal{E}}^\diamond \subseteq \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset, \tilde{\rho}_e < \widehat{\rho}_e\}$, and hence $\bar{e} \notin \bar{\mathcal{E}}^\diamond$. We consider $\tilde{\rho}'$ such that $\tilde{\rho}'_{\bar{e}} = \tilde{\rho}_{\bar{e}} - \epsilon$, where ϵ is a sufficiently small positive number, and $\tilde{\rho}'_e = \tilde{\rho}_e$ for any other facilities. Then, $\tilde{\rho}'_{\bar{e}} > \widehat{\rho}_{\bar{e}}$ still holds, and the set $\bar{\mathcal{E}}^\diamond$ does not change. The attacker's best response strategy remains to be $BR(\tilde{\rho}') = \Delta(\bar{\mathcal{E}}^\diamond)$. Hence, the utility of the defender given $\tilde{\rho}'$ increases by ϵp_d compared to that given $\tilde{\rho}$, because the expected usage cost $\mathbb{E}_\sigma[C]$ does not change, but the expected defense cost decreases by ϵp_d. Thus, such $\tilde{\rho}$ cannot be the defender's equilibrium effort.
- For all $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}_e \geq \widehat{\rho}_e$. We have already argued that $\widetilde{\sigma}^*_a(\emptyset, \tilde{\rho}) = 1$ in this case. Since the defense cost $p_d > 0$, if there exists any e such that $\tilde{\rho}_e > \widehat{\rho}_e$, then by decreasing the security effort on e, the utility of the defender increases. Therefore, such $\tilde{\rho}$ cannot be an equilibrium strategy of the defender.

From both cases, we can conclude that for any $e \in \{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$, $\tilde{\rho}^*_e \leq \widehat{\rho}_e$

Finally, any non-vulnerable facilities $e \in \mathcal{E} \setminus \{\mathcal{E} | C_e - p_a > C_\emptyset\}$ will not be targeted; hence, we must have $\tilde{\rho}^*_e = 0$. □

Proof of Lemma 6 We first show that the threshold $\widetilde{p}_d(p_a)$ as given in (31) is a well-defined function of p_a. Given any $0 \leq p_a < C_{(1)} - C_\emptyset$, there is a unique $i \in \{1, \ldots, K\}$ such that $C_{(i+1)} - C_\emptyset \leq p_a < C_{(i)} - C_\emptyset$. Now, we need to show that there is a unique $j \in \{1, \ldots, i\}$ such that $\frac{\sum_{k=j+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}} \leq p_a < \frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}$ (or $0 \leq p_a < \frac{E_{(i)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}$ if $j = i$). Note that functions $\{p_d^{ij}\}_{j=1}^{i}$ are defined on the range $\left[0, \frac{\sum_{k=1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}\right]$. Since $\{C_{(k)}\}_{k=1}^{i}$ satisfies (8), we have:

$$\frac{\sum_{k=1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}} \geq \frac{\sum_{k=1}^{i} E_{(k)}}{\frac{1}{C_{(i)}-C_{\emptyset}} \sum_{k=1}^{i} E_{(k)}} = C_{(i)} - C_{\emptyset}.$$

Hence, for any $C_{(i+1)} - C_{\emptyset} \leq p_a < C_{(i)} - C_{\emptyset}$, the value $\widetilde{p}_d(p_a)$ is defined as $p_d^{ij}(p_a)$ for a unique $j \in \{1, \ldots, i\}$. Therefore, we can conclude that for any $0 \leq p_a < C_{(1)} - C_{\emptyset}$, $\widetilde{p}_d(p_a)$ is a well-defined function.

We next show that $\widetilde{p}_d(p_a)$ is continuous and strictly increasing in p_a. Since for any $i = 1, \ldots, K$, and any $j = 1, \ldots, i$, the function $p_d^{ij}(p_a)$ is continuous and strictly increasing in p_a, $\widetilde{p}_d(p_a)$ must be piecewise continuous and strictly increasing in p_a. It remains to be shown that $\widetilde{p}_d(p_a)$ is continuous at $p_a \in \{C_{(i)} - C_{\emptyset}\}_{i=2}^{K} \cup \left\{ \frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}} \right\}_{j=1,\ldots,i, i=1,\ldots K}$.

We now show that for any $i = 2, \ldots, K$, $\widetilde{p}_d(p_a)$ is continuous at $C_{(i)} - C_{\emptyset}$. Consider $p_a = C_{(i)} - C_{\emptyset} - \epsilon$ where ϵ is a sufficiently small positive number. There is a unique $\hat{j} \in \{1, \ldots, i\}$ such that $\widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a)$. We want to argue that $\hat{j} \neq i$:

$$\begin{aligned} p_a \cdot \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) &= \left(C_{(i)} - C_{\emptyset} - \epsilon \right) \cdot \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) \\ &= E_{(i)} + \sum_{k=1}^{i-1} \frac{\left(C_{(i)} - C_{\emptyset} \right) E_{(k)}}{C_{(k)} - C_{\emptyset}} - \epsilon \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) > E_{(i)}, \\ \Rightarrow \qquad p_a &= C_{(i)} - C_{\emptyset} - \epsilon > \frac{E_{(i)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}} \end{aligned}$$

Thus, $\hat{j} \in \{1, \ldots, i-1\}$, and from (31), $\frac{\sum_{k=\hat{j}+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}} \leq C_{(i)} - C_{\emptyset} - \epsilon < \frac{\sum_{k=\hat{j}}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}}$. Since ϵ is a sufficiently small positive number, we have:

$$\begin{aligned} \sum_{k=\hat{j}+1}^{i} E_{(k)} &\leq \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) \cdot \left(C_{(i)} - C_{\emptyset} - \epsilon \right) \\ &= E_{(i)} + \sum_{k=1}^{i-1} \frac{\left(C_{(i)} - C_{\emptyset} \right) E_{(k)}}{C_{(k)} - C_{\emptyset}} - \epsilon \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) \\ \Rightarrow \quad \sum_{k=\hat{j}+1}^{i-1} E_{(k)} &\leq \sum_{k=1}^{i-1} \frac{\left(C_{(i)} - C_{\emptyset} \right) E_{(k)}}{C_{(k)} - C_{\emptyset}} + \epsilon \left(\sum_{k=1}^{i-1} \frac{E_{(k)}}{C_{(k)} - C_{\emptyset}} \right) \\ \Rightarrow \quad \frac{\sum_{k=\hat{j}+1}^{i-1} E_{(k)}}{\sum_{k=1}^{i-1} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}} &\leq C_{(i)} - C_{\emptyset} + \epsilon. \end{aligned}$$

Analogously, we can check that $C_{(i)} - C_{\emptyset} + \epsilon < \frac{\sum_{k=\hat{j}}^{i-1} E_{(k)}}{\sum_{k=1}^{i-1} \frac{E_{(k)}}{C_{(k)}-C_{\emptyset}}}$. Hence, from (31), when $p_a = C_{(i)} - C_{\emptyset} + \epsilon$, we have $\widetilde{p}_d(p_a) = p_d^{i-1\hat{j}}(p_a)$. Then,

$$\lim_{p_a \to (C_{(i)} - C_\emptyset)^-} \widetilde{p}_d(p_a) = \lim_{\epsilon \to 0} p_d^{i\hat{j}}(C_{(i)} - C_\emptyset - \epsilon)$$

$$\overset{(30)}{=} \frac{C_{(\hat{j})} - C_\emptyset}{\left(C_{(\hat{j})} - C_\emptyset\right) \cdot \left(\sum_{k=1}^{\hat{j}-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right) + \sum_{k=\hat{j}}^{i-1} E_{(k)} - \sum_{k=1}^{i-1} \frac{p_a E_{(k)}}{C_{(k)} - C_\emptyset}}$$

$$= \lim_{\epsilon \to 0} p_d^{i-1\hat{j}}(C_{(i)} - C_\emptyset + \epsilon) = \lim_{p_a \to (C_{(i)} - C_\emptyset)^+} \widetilde{p}_d(p_a).$$

Thus, $\widetilde{p}_d(p_a)$ is continuous at $C_{(i)} - C_\emptyset$ for any $i = 2, \ldots, K$.

For any $i = 1, \ldots, K$, we next show that $\widetilde{p}_d(p_a)$ is continuous at $p_a = \frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}$ for $j = 1, \ldots, i$:

$$\lim_{p_a \to \left(\frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}\right)^-} \widetilde{p}_d(p_a) = p_d^{ij}\left(\frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}\right) = \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1}$$

$$= p_d^{i(j-1)}\left(\frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}\right) = \lim_{p_a \to \left(\frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}}\right)^+} \widetilde{p}_d(p_a).$$

Hence, we can conclude that $\widetilde{p}_d(p_a)$ is continuous and strictly increasing in p_a.

Additionally, for any $i = 1, \ldots, K$, consider any p_a such that $C_{(i+1)} - C_\emptyset < p_a \leq C_{(i)} - C_\emptyset$ (or $0 < p_a \leq C_{(K)} - C_\emptyset$ if $i = K$), then for any $j = 1, \ldots, i$, we have:

$$p_d^{ij}(p_a) \overset{(30)}{=} \frac{C_{(j)} - C_\emptyset}{\left(C_{(j)} - p_a - C_\emptyset\right) \cdot \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right) + \sum_{k=j}^{i} \frac{(C_{(k)} - p_a - C_\emptyset) E_{(k)}}{C_{(k)} - C_\emptyset}}$$

$$> \frac{C_{(j)} - C_\emptyset}{\left(C_{(j)} - C_{(i+1)}\right) \cdot \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right) + \sum_{k=j}^{i} \frac{(C_{(k)} - C_{(i+1)}) E_{(k)}}{C_{(k)} - C_\emptyset}}$$

$$= \frac{C_{(j)} - C_\emptyset}{\left(C_{(j)} - C_{(i+1)}\right) \cdot \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)}$$

$$\overset{(8)}{>} \left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1}$$

$$\overset{(17)}{=} \bar{p}_d(p_a).$$

Therefore, for any $0 < p_a < C_{(1)} - C_\emptyset$, we have:

$$\widetilde{p}_d(p_a) \overset{(31)}{\geq} \min_{j=1,\ldots,i} p_d^{ij}(p_a) > \bar{p}_d(p_a), \tag{50}$$

Finally, if $p_a = 0$, then we know that $\widetilde{p}_d(0) = p_d^{KK}(0)$. From (30), we can check that $p_d^{KK}(0) = \left(\sum_{k=1}^{K} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1} = \bar{p}_d(0)$. If p_a approaches $C_{(1)} - C_\emptyset$, then $\widetilde{p}_d(p_a) = p_d^{11}(p_a)$, and we have:

$$\lim_{p_a \to C_{(1)} - C_\emptyset} \widetilde{p}_d(p_a) \overset{(30)}{=} \lim_{p_a \to C_{(1)} - C_\emptyset} \frac{C_{(1)} - C_\emptyset}{E_{(1)} - \frac{p_a E_{(1)}}{C_{(1)} - C_\emptyset}} = +\infty$$

□

We define the partition as:

$$\mathcal{P} \triangleq \left\{ \left\{ \Lambda^i \right\}_{i=0}^{K}, \left\{ \Lambda_j^i \right\}_{j=1,\ldots,i, i=1,\ldots,K} \right\}, \tag{51}$$

where $\left\{ \Lambda^i \right\}_{i=0}^{K}$ are type I regimes in the normal form game defined in (18)-(20), and Λ_j^i is the set of (p_d, p_a), which satisfy:

$$p_d \in \begin{cases} \left(\left(\frac{E_{(1)}}{C_{(1)} - C_\emptyset} \right)^{-1}, +\infty \right), & \text{if } j = 1, \\ \left(\left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1}, \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset} \right)^{-1} \right), & \text{if } j = 2, \ldots, K, \end{cases} \tag{52a}$$

$$p_a \in \begin{cases} \left(C_{(i+1)} - C_\emptyset, C_{(i)} - C_\emptyset \right), & \text{if } i = 1, \ldots, K-1, \\ \left(0, C_{(K)} - C_\emptyset \right), & \text{if } i = K, \end{cases} \tag{52b}$$

We can check that sets in $\mathcal{P}$ are disjoint and cover the whole space of (p_d, p_a). Lemma 7 characterizes SPE in sets $\left\{ \Lambda^i \right\}_{i=0}^{K}$, and Lemma 8 characterizes SPE in sets $\left\{ \Lambda_j^i \right\}_{j=1,\ldots,i,\, i=1,\ldots,K}$.

Lemma 7 *In $\widetilde{\Gamma}$, for any (p_a, p_d) in the set Λ^i, where $i = 0, \ldots, K$:*

- *If $i = 0$, then SPE is as given in (37).*
- *If $i = 1, \ldots, K$, then SPE is as given in (38).*

Proof of Lemma 7. – If $i = 0$:
The set of vulnerable facilities $\{\bar{\mathcal{E}} | C_e - p_a > C_\emptyset\}$ is empty. Thus, $\widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}) = 1$, and $\tilde{\rho}_e^* = 0$ for all $e \in \mathcal{E}$.

– For any $i = 1, \ldots, K$:
The set of vulnerable facilities is $\cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$. From Lemma 5, we have already known that for any $e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$, $\tilde{\rho}_e^* \leq \widehat{\rho}_e$. Assume for the sake of contradiction that there exists a facility $\bar{e} \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$ such that $\tilde{\rho}_{\bar{e}} < \widehat{\rho}_{\bar{e}}$. From Lemma 4, we know that $\widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}) = 0$, and $BR(\tilde{\rho}) = \Delta(\bar{\mathcal{E}}^\diamond)$, where $\bar{\mathcal{E}}^\diamond$ is in (29). Clearly, $\bar{\mathcal{E}}^\diamond \subseteq \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}$. We define λ as follows:

$$\lambda = \max_{e \in \cup_{k=1}^{i} \bar{\mathcal{E}}_{(k)}} \{\tilde{\rho}_e C_\emptyset + (1 - \tilde{\rho}_e) C_e\} = \tilde{\rho}_e C_\emptyset + (1 - \tilde{\rho}_e) C_e, \quad \forall e \in \bar{\mathcal{E}}^\diamond.$$

The utility of the defender can be written as:

$$U_d(\tilde{\rho}, \widetilde{\sigma}_a^*(\tilde{\rho})) = -\lambda - \left(\sum_{e \in \mathcal{E}} \tilde{\rho}_e \right) \cdot p_d.$$

We now consider $\tilde{\rho}'$ as follows:

$$\begin{aligned} \tilde{\rho}_e' &= \tilde{\rho}_e + \frac{\epsilon}{C_e - C_\emptyset}, && \forall e \in \bar{\mathcal{E}}^\diamond, \\ \tilde{\rho}_e' &= \tilde{\rho}_e, && \forall e \in \mathcal{E} \setminus \bar{\mathcal{E}}^\diamond, \end{aligned}$$

where ϵ is a sufficiently small positive number. Under this deviation, we can check that the set $\bar{\mathcal{E}}^\diamond$ does not change, but λ changes to $\lambda - \epsilon$. Therefore, the defender's utility can be written as:

$$U_d(\tilde{\rho}', \widetilde{\sigma}_a(\tilde{\rho}')) = -\lambda + \epsilon - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}'_e\right)\cdot p_d = -\lambda + \epsilon - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d - \sum_{e\in\bar{\mathcal{E}}^\diamond} \frac{\epsilon\, p_d}{C_e - C_\emptyset}$$

$$= U_d(\tilde{\rho}, \widetilde{\sigma}_a(\tilde{\rho})) + \epsilon\left(1 - \sum_{e\in\bar{\mathcal{E}}^\diamond} \frac{p_d}{C_e - C_\emptyset}\right)$$

$$\geq U_d(\tilde{\rho}, \widetilde{\sigma}_a(\tilde{\rho})) + \epsilon\left(1 - \sum_{e\in\cup_{k=1}^{i}\bar{\mathcal{E}}_{(k)}} \frac{p_d}{C_e - C_\emptyset}\right)$$

$$= U_d(\tilde{\rho}, \widetilde{\sigma}_a(\tilde{\rho})) + \epsilon\left(1 - \sum_{k=1}^{i} \frac{p_d E_{(k)}}{C_e - C_\emptyset}\right) \overset{(19)}{>} U_d(\tilde{\rho}, \widetilde{\sigma}_a(\tilde{\rho})).$$

Therefore, such $\tilde{\rho}$ cannot be an equilibrium strategy profile. We thus know that $\tilde{\rho}^*$ is as given in (38). The attacker's equilibrium strategy can be derived from Lemmas 4 and 5 directly. □

Lemma 8 *For (p_a, p_d) in Λ_j^i, where $i = 1, \ldots, K$, and $j = 1, \ldots, i$, there are two cases of SPE:*

- *If $p_d > p_d^{ij}$, where p_d^{ij} is as given in (30):*
  - *If $j = 1$, then SPE is as given in (39).*
  - *If $j = 2, \ldots, i$, then SPE is as given in (40).*
- *If $p_d < p_d^{ij}$, then the SPE is as given in (38).*

Proof of Lemma 8 Consider cost parameters in the set Λ_j^i defined in (52), where $i = 1, \ldots, K$ and $j = 1, \ldots, i$. The set of vulnerable facilities is $\cup_{k=1}^{i}\bar{\mathcal{E}}_{(k)}$. From Lemma 5, we know that the defender can either secure all vulnerable facilities $e \in \cup_{k=1}^{i}\bar{\mathcal{E}}_{(k)}$ with the threshold effort $\widehat{\rho}_e$ defined in (27), or leave at least one vulnerable facility secured less than the threshold effort. We discuss the two cases separately:

(1) If any $e \in \cup_{k=1}^{i}\bar{\mathcal{E}}_{(k)}$ is secured with the threshold effort $\widehat{\rho}_e$, then from Lemma 5, we know that the total probability of attack is 0. The defender's utility can be written as:

$$U_d(\widehat{\rho}, \widetilde{\sigma}_a^*(\widehat{\rho})) = -C_\emptyset - \left(\sum_{k=1}^{i} \frac{\left(C_{(k)} - p_a - C_\emptyset\right)\cdot E_{(k)}}{C_{(k)} - C_\emptyset}\right)\cdot p_d. \tag{53}$$

(2) If the set $\{\mathcal{E} | C_e - p_a > C_\emptyset, \quad \tilde{\rho}_e < \widehat{\rho}_e\}$ is non-empty, then we define $\widetilde{P}$ as the set of feasible $\tilde{\rho}$ in this case. We denote $\tilde{\rho}^\dagger$ as the secure effort vector that incurs the highest utility for the defender among all $\tilde{\rho} \in \widetilde{P}$. Then, $\tilde{\rho}^\dagger$ can be written as:

$$\tilde{\rho}^\dagger \in \underset{\tilde{\rho}\in\widetilde{P}}{\text{argmax}}\, U_d(\tilde{\rho}, \widetilde{\sigma}_a^*(\tilde{\rho})) = \underset{\tilde{\rho}\in\widetilde{P}}{\text{argmax}}\left(-\mathbb{E}_{(\tilde{\rho}, \widetilde{\sigma}_a^*(\tilde{\rho}))}[C] - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d\right)$$

$$= \underset{\tilde{\rho}\in\widetilde{P}}{\text{argmax}}\left(-\mathbb{E}_{(\tilde{\rho}, \widetilde{\sigma}_a^*(\tilde{\rho}))}[C] - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d + \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e, \tilde{\rho})\right)\cdot p_a\right.$$

$$\left. - \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e, \tilde{\rho})\right)\cdot p_a\right). \tag{54}$$

We know from Lemma 4 that $\widetilde{\sigma}_a^*(\emptyset, \tilde{\rho}) = 0$. Therefore, $\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e, \tilde{\rho}) = 1$, and (54) can be re-expressed as:

$$\begin{aligned}\tilde{\rho}^\dagger \in &\operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \left(-\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a^*(\tilde{\rho}))}[C] - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d + \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e,\tilde{\rho})\right)\cdot p_a - p_a\right)\\ = &\operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \left(-\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a^*(\tilde{\rho}))}[C] - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d + \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e,\tilde{\rho})\right)\cdot p_a\right).\end{aligned}$$

Since in equilibrium, the attacker chooses the best response strategy, we have:

$$\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a^*(\tilde{\rho}))}[C] - \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a^*(e,\tilde{\rho})\right)\cdot p_a = \max_{\widetilde{\sigma}_a\in\Delta(S_a)} \left(\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a)}[C] - \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a(e)\right)\cdot p_a\right). \tag{55}$$

Hence, $\tilde{\rho}^\dagger$ can be re-expressed as:

$$\begin{aligned}\tilde{\rho}^\dagger &\overset{(55)}{=} \operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \left(-\max_{\widetilde{\sigma}_a\in\Delta(S_a)} \left(\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a)}[C] - \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a(e)\right)\cdot p_a\right) - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d\right)\\ &= \operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \left(-\max_{\widetilde{\sigma}_a\in\Delta(S_a)} \left(\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a)}[C] - \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a(e)\right)\cdot p_a + \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d\right)\right)\\ &= \operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \min_{\widetilde{\sigma}_a\in\Delta(S_a)} \left(-\mathbb{E}_{(\tilde{\rho},\widetilde{\sigma}_a)}[C] + \left(\sum_{e\in\mathcal{E}} \widetilde{\sigma}_a(e)\right)\cdot p_a - \left(\sum_{e\in\mathcal{E}} \tilde{\rho}_e\right)\cdot p_d\right)\\ &\overset{(9a)}{=} \operatorname*{argmax}_{\tilde{\rho}\in\widetilde{P}} \min_{\widetilde{\sigma}_a\in\Delta(S_a)} U_d^0(\tilde{\rho},\widetilde{\sigma}_a).\end{aligned}$$

Therefore, $\tilde{\rho}^\dagger$ is the defender's equilibrium strategy in the zero-sum game, which is identical to the equilibrium strategy in the normal form game (recall Lemma 2). From Theorem 1, when p_a and p_d are in Λ_j^i, $\tilde{\rho}^\dagger$ is in (40) (or (39) if $j = 1$). The defender's utility in this case is:

$$U_d(\tilde{\rho}^\dagger, \widetilde{\sigma}_a^*(\tilde{\rho}^\dagger)) = -C_{(j)} - \left(\sum_{k=1}^{j-1} \frac{\left(C_{(k)} - C_{(j)}\right)\cdot E_{(k)}}{C_{(k)} - C_\emptyset}\right)\cdot p_d. \tag{56}$$

Finally, by comparing U_d in (56) and (53), we can check that if $p_d > p_d^{ij}$, then $U_d(\tilde{\rho}^\dagger, \widetilde{\sigma}_a^*(\tilde{\rho}^\dagger)) > U_d(\widehat{\rho}, \widetilde{\sigma}_a^*(\widehat{\rho}))$. Thus, SPE is in (40) (or (39) if $j = 1$). If $p_d < p_d^{ij}$, then $U_d(\tilde{\rho}^\dagger, \widetilde{\sigma}_a^*(\tilde{\rho}^\dagger)) < U_d(\widehat{\rho}, \widetilde{\sigma}_a^*(\widehat{\rho}))$, and SPE is in (38). □

Proof of Theorem 2 (a) Type $\widetilde{\mathrm{I}}$ regimes $\widetilde{\Lambda}^i$:

- If $i = 0$:
  There is no vulnerable facility. Therefore, the attacker chooses not to attack with probability 1, and the defender does not secure any facility. SPE is as given in (37).
- If $i = 1, \ldots, K$:
  Consider any $C_{(i+1)} - C_\emptyset < p_a < C_{(i)} - C_\emptyset$. From Lemma 6, we know that $\widetilde{p}_d(p_a) > \bar{p}_d(p_a)$, where $\widetilde{p}_d(p_a)$ is as defined in (31) and $\bar{p}_d(p_a)$ is as defined in (16). From Lemma 7, we know that SPE is as given in (38) for any $p_d < \bar{p}_d(p_a)$.
  It remains to be shown that for any $\bar{p}_d(p_a) \le p_d < \widetilde{p}_d(p_a)$, SPE is also as given in

(38). For any $C_{(i+1)} - C_\emptyset \le p_a < C_{(i)} - C_\emptyset$, there is a unique $\hat{j} \in \{1, \ldots, i\}$ such that $\frac{\sum_{k=\hat{j}+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}} \le p_a < \frac{\sum_{k=\hat{j}}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}$, and from (31), we have:

$$\widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a) \ge p_d^{i\hat{j}}\left(\frac{\sum_{k=\hat{j}+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}\right) \overset{30}{=} \left(\sum_{k=1}^{\hat{j}} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1},$$

$$\widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a) < p_d^{i\hat{j}}\left(\frac{\sum_{k=\hat{j}}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}\right) = \left(\sum_{k=1}^{\hat{j}-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1}.$$

Consider any $j = \hat{j}+1, \ldots, i$, and any $\left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \le p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1}$, the cost parameters (p_a, p_d) are in the set Λ_j^i as defined in (52). Additionally, from our definition of $\hat{j}$, we know that $p_a > \frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}$. We now show that in Λ_j^i, $p_d < p_d^{ij}(p_a)$:

$$p_d^{ij}(p_a) \overset{30}{>} p_d^{ij}\left(\frac{\sum_{k=j}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}\right) = \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1} \overset{52}{>} p_d.$$

Hence, from Lemma 8, we know that for any $\left(\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \le p_d \le \left(\sum_{k=1}^{\hat{j}} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1}$, SPE is as given in (38). For any $\left(\sum_{k=1}^{\hat{j}} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} < p_d < \widetilde{p}_d(p_a)$, the cost parameters (p_a, p_d) are in the set $\Lambda_i^{\hat{j}}$, and $p_d < \widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a)$. Again from Lemma 8, SPE is in (38).
Therefore, we can conclude that in regime $\widetilde{\Lambda}^i$, SPE is in (38).

(b) Type $\widetilde{\text{II}}$ regimes $\widetilde{\Lambda}_j$, where $j = 1, \ldots, K$:
Since $\widetilde{p}_d(p_a)$ is strictly increasing in p_a and $\lim_{p_a \to C_{(1)}-C_\emptyset} \widetilde{p}_d(p_a) = +\infty$, we know that for any $p_d > 0$, any $p_a < \widetilde{p}_d^{-1}(p_d) < C_{(1)} - C_\emptyset$. Therefore, we can re-express $\widetilde{\Lambda}^1$ as follows:

$$\widetilde{\Lambda}^1 \overset{35}{=} \left\{(p_a, p_d) \,\middle|\, p_a < \widetilde{p}_d^{-1}(p_d),\ p_d > \left(\frac{E_{(1)}}{C_{(1)} - C_\emptyset}\right)^{-1}\right\}$$

$$= \left\{(p_a, p_d) \,\middle|\, p_d > \widetilde{p}_d(p_a),\ p_d > \left(\frac{E_{(1)}}{C_{(1)} - C_\emptyset}\right)^{-1}, 0 \le p_a \le C_{(1)} - C_\emptyset\right\}$$

$$\overset{52}{=} \bigcup_{i=1}^{K} \left(\Lambda_1^i \bigcap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\}\right). \tag{57}$$

For any $j = 2, \ldots, K$, if $p_a > C_{(j)} - C_\emptyset$, then from Lemma 6, we have:

$$\widetilde{p}_d(p_a) > \bar{p}_d(p_a) \overset{17}{\ge} \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)} - C_\emptyset}\right)^{-1}. \tag{58}$$

Therefore, for any $p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1}$, we know that any $p_a < \widetilde{p}_d^{-1}(p_d) < C_{(j)} - C_\emptyset$. Analogous to (57), we re-express the set $\widetilde{\Lambda}_j$ as follows:

$$\begin{aligned}
\widetilde{\Lambda}_j &\overset{(36)}{=} \left\{ (p_a, p_d) \,\middle|\, p_a < \widetilde{p}_d^{-1}(p_d),\ \left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \le p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \right\} \\
&\overset{(58)}{=} \left\{ (p_a, p_d) \,\middle|\, p_d > \widetilde{p}_d(p_a),\ \left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \le p_d < \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1},\ 0 \le p_a \le C_{(j)} - C_\emptyset \right\} \\
&\overset{(52)}{=} \bigcup_{i=j}^{K} \left(\Lambda_j^i \bigcap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\} \right).
\end{aligned}$$

We next show that for any $j = 1, \ldots, K$, and any $i = j, \ldots, K$, the set $\Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\} \subseteq \Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > p_d^{ij}(p_a)\}$. Considering any cost parameters (p_a, p_d) in the set $\Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\}$, from (31), we can find $\hat{j}$ such that $\frac{\sum_{k=\hat{j}+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}} \le p_a < \frac{\sum_{k=\hat{j}}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}$, and $\widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a)$. We discuss the following three cases separately:

- If $\hat{j} > j$, then we must have $p_a < \frac{\sum_{k=\hat{j}}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}} \le \frac{\sum_{k=j+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}$. Hence, from (30), $p_d^{ij}(p_a) < \left(\sum_{k=1}^{j} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1}$. From the definition of the set Λ_j^i in (52), we know that $p_d > p_d^{ij}(p_a)$ in this set, and thus $(p_a, p_d) \in \Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > p_d^{ij}(p_a)\}$.
- If $\hat{j} = j$, then we directly obtain that $(p_a, p_d) \in \Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > p_d^{ij}(p_a)\}$.
- If $\hat{j} < j$, then since $p_a \ge \frac{\sum_{k=\hat{j}+1}^{i} E_{(k)}}{\sum_{k=1}^{i} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}}$, from (30), we have $\widetilde{p}_d(p_a) = p_d^{i\hat{j}}(p_a) \ge \left(\sum_{k=1}^{\hat{j}} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1} \ge \left(\sum_{k=1}^{j-1} \frac{E_{(k)}}{C_{(k)}-C_\emptyset}\right)^{-1}$. From the definition of the set Λ_j^i in (52), the set $\Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\}$ is empty and thus can be omitted.

We can conclude from all three cases that $\Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d > \widetilde{p}_d(p_a)\} \subseteq \Lambda_j^i \cap \{(p_a, p_d) \,|\, p_d \ge p_d^{ij}(p_a)\}$. Therefore, from Lemma 8, SPE is in (40) (or (39) if $j = 1$) in the regime $\widetilde{\Lambda}_j$.

□

References

1. Acemoglu D, Malekian A, Ozdaglar A (2016) Network security and contagion. J Econ Theory 166:536–585
2. Alderson DL, Brown GG, Carlyle WM, Wood RK (2011) Solving defender-attacker-defender models for infrastructure defense. Technical report, Naval Postgraduate School, Monterey, CA, Department of Operations Research

3. Alderson DL, Brown GG, Carlyle WM, Wood RK (2017) Assessing and improving the operational resilience of a large highway infrastructure system to worst-case losses. Transp Sci 52(4):739–1034
4. Alpcan T, Basar T (2003) A game theoretic approach to decision and analysis in network intrusion detection. In: Proceedings of the 42nd IEEE conference on decision and control, 2003, IEEE. 3:2595–2600
5. Alpcan T, Başar T (2010) Network security: a decision and game-theoretic approach. Cambridge University Press, Cambridge
6. Amin S, Schwartz GA, Sastry SS (2013) Security of interdependent and identical networked control systems. Automatica 49(1):186–192
7. Bagwell K (1995) Commitment and observability in games. Games Econ Behav 8(2):271–280
8. Başar T, Olsder GJ (1998) Dynamic noncooperative game theory. SIAM, Philadelphia
9. Battigalli P, Gilli M, Molinari MC (1992) Learning and convergence to equilibrium in repeated strategic interactions: an introductory survey. Università Commerciale" L. Bocconi", Istituto di Economia Politica
10. Beggs AW (2005) On the convergence of reinforcement learning. J Econ Theory 122(1):1–36
11. Bell MGH, Kanturska U, Schmöcker J-D, Fonzone A (2008) Attacker-defender models and road network vulnerability. Philos Trans R Soc Lond A: Math Phys Eng Sci 366(1872):1893–1906
12. Bier V, Oliveros S, Samuelson L (2007) Choosing what to protect: strategic defensive allocation against an unknown attacker. J Public Econ Theory 9(4):563–587
13. Bier VM, Hausken K (2013) Defending and attacking a network of two arcs subject to traffic congestion. Reliabil Eng Syst Saf 112:214–224
14. Blum A, Even-Dar E, Ligett K (2006) Routing without regret: on convergence to Nash equilibria of regret-minimizing algorithms in routing games. In : Proceedings of the 25th annual ACM symposium on principles of distributed computing. ACM, pp 45–52
15. Brown GW (1951) Iterative solution of games by fictitious play. Act Anal Prod Alloc 13(1):374–376
16. Brown G, Carlyle M, Salmerón J, Wood K (2006) Defending critical infrastructure. Interfaces 36(6):530–544
17. Cárdenas AA, Amin S, Lin Z-S, Huang Y-L, Huang C-Y, Sastry S (2011) Attacks against process control systems: risk assessment, detection, and response. In: Proceedings of the 6th ACM symposium on information, computer and communications security. ACM, pp 355–366
18. Chen L, Leneutre J (2009) A game theoretical framework on intrusion detection in heterogeneous networks. IEEE Trans Inf Forensics Secur 4(2):165–178
19. Cominetti R, Melo E, Sorin S (2010) A payoff-based learning procedure and its application to traffic games. Games Econ Behav 70(1):71–83
20. Cominetti R, Facchinei F, Lasserre JB (2012) Adaptive dynamics in traffic games. In: Modern optimization modelling techniques. Springer, pp 239–257
21. Correa JR, Stier-Moses NE (2011) Wardrop equilibria. Wiley Encyclopedia of Operations Research and Management Science, Hoboken
22. Dahan M, Amin S (2015) Network flow routing under strategic link disruptions. In: 53rd Annual Allerton conference on communication, control, and computing (Allerton), 2015. IEEE, pp 353–360
23. Dritsoula L, Loiseau P, Musacchio J (2012) A game-theoretical approach for finding optimal strategies in an intruder classification game. In: IEEE 51st Annual conference on decision and control (CDC), 2012. IEEE, pp 7744–7751
24. Dritsoula L, Loiseau P, Musacchio J (2017) A game-theoretic analysis of adversarial classification. IEEE Trans Inf Forensics Secur 12(12):3094–3109
25. Dziubiński M, Goyal S (2013) Network design and defence. Games Econ Behav 79:30–43
26. Dziubiński M, Goyal S (2017) How do you defend a network? Theor Econ 12(1):331–376
27. Fudenberg D, Kreps DM (1995) Learning in extensive-form games i. self-confirming equilibria. Games Econ Behav 8(1):20–55
28. Fudenberg D, Levine DK (1995) Consistency and cautious fictitious play. J Econ Dyn Control 19(5–7):1065–1089
29. Goyal S, Vigier A (2014) Attack, defence, and contagion in networks. Rev Econ Stud 81(4):1518–1542
30. Hofbauer J, Sandholm WH (2002) On the global convergence of stochastic fictitious play. Econometrica 70(6):2265–2294
31. Kalai E, Lehrer E (1993) Rational learning leads to Nash equilibrium. Econometrica 61(5):1019–1045
32. Kalai E, Lehrer E (1993) Subjective equilibrium in repeated games. Econometrica 61(5):1231–1240
33. Kalai E, Shmaya E (2015) Learning and stability in big uncertain games. Technical report
34. Laporte G, Mesa JA, Perea F (2010) A game theoretic framework for the robust railway transit network design problem. Transp Res Part B Methodol 44(4):447–459
35. Manshaei MH, Zhu Q, Alpcan T, Bacşar T, Hubaux J-P (2013) Game theory meets network security and privacy. ACM Comput Surv (CSUR) 45(3):25

36. Marden JR, Arslan G, Shamma JS (2007) Regret based dynamics: convergence in weakly acyclic games. In: Proceedings of the 6th international joint conference on Autonomous agents and multiagent systems. ACM, p 42
37. Maskin E, Tirole J (2001) Markov perfect equilibrium: I. observable actions. J Econ Theory 100(2):191–219
38. Moteff J, Parfomak P (2004) Critical infrastructure and key assets: definition and identification. Library of Congress Washington, DC, Congressional Research Service
39. Nguyen KC, Alpcan T, Basar T (2009) Stochastic games for security in networks with interdependent nodes. In: International conference on game theory for networks, GameNets' 09. IEEE, pp 697–703
40. Powell R (2007) Defending against terrorist attacks with limited resources. Am Polit Sci Rev 101(3):527–541
41. Rinaldi SM, Peerenboom JP, Kelly TK (2001) Identifying, understanding, and analyzing critical infrastructure interdependencies. IEEE Control Syst 21(6):11–25
42. Rosenthal RW (1974) Correlated equilibria in some classes of two-person games. Int J Game Theory 3(3):119–128
43. Sandberg H, Amin S, Johansson KH (2015) Cyberphysical security in networked control systems: an introduction to the issue. IEEE Control Syst 35(1):20–23
44. Sandholm WH (2001) Potential games with continuous player sets. J Econ Theory 97(1):81–108
45. Schwartz GA, Amin S, Gueye A, Walrand J (2011) Network design game with both reliability and security failures. In: 49th Annual Allerton conference on communication, control, and computing (Allerton), 2011. IEEE, pp 675–681
46. Sethi AR, Amin S, Schwartz G (2017) Value of intrusion detection systems for countering energy fraud. In: American Control Conference (ACC), 2017. IEEE, pp 2739–2746
47. Sridhar S, Hahn A, Govindarasu M (2012) Cyber-physical system security for the electric power grid. Proc IEEE 100(1):210–224
48. Von Stengel B, Zamir S (2004) Leadership with commitment to mixed strategies
49. Washburn A, Wood K (1995) Two-person zero-sum games for network interdiction. Oper Res 43(2):243–251