Dynamic Games and Applications manuscript No.
(will be inserted by the editor)

An Efficient Dynamic Allocation Mechanism for Security in Networks of Interdependent Strategic Agents

Farzaneh Farhadi[1,2] · **Hamidreza Tavafoghi**[1,3] · **Demosthenis Teneketzis**[1] · **and S. Jamaloddin Golestani**[2]

Received: date / Accepted: date

Abstract Motivated by security issues in networks, we study the problem of incentive mechanism design for dynamic resource allocation in a multi-agent networked system. Each strategic agent has a private security state which can be safe or unsafe and is only known to him. At every time, each agent faces security threats from outside as well as from his unsafe neighbors. Therefore, the agents' states are correlated and have interdependent stochastic dynamics. Agents have interdependent valuations, as each agent's instantaneous utility depends on his own security state as well as his neighbors' security states. There is a network manager that can allocate a security resource to one agent at each time so as to protect the network against attacks and maximize the overall social welfare. We propose a dynamic incentive mechanism that implements the efficient allocation and is ex-ante (in expectation) individually rational and budget balanced. We present a reputation-based payment that mitigates any risk that the agents or the network manager may face to get a negative utility or to run a budget deficit, respectively, for some realizations of the network stochastic evolution. Therefore, our results provide a dynamic incentive mechanism that implements efficient allocations in networked systems with strategic agents that have correlated types and interdependent valuations, and is approximate ex-post individually rational and budget balanced.

Keywords Security games · dynamic mechanism design · epidemics over networks · strategic agents.

A preliminary version of this paper appeared in the Proceeding of the 7th EAI International Conference on Game Theory for Networks (GameNets), May 2017 (See [14]). This work was supported in part by the NSF grants CNS-1238962, CCF-1111061, ARO-MURI grant W911NF-13-1-0421, and ARO grant W911NF-17-1-0232.

[1]University of Michigan, Ann Arbor, USA
[2]Sharif University of Technology, Tehran, Iran
[3]University of California, Berkeley, USA
E-mail: ffarhadi, tavaf, teneket @ umich.edu, golestani@ieee.org

1 Introduction

Studying the dynamic behavior of strategic agents in networks, where each agent's state and utility are influenced by his interactions with his neighbors, is a topic that has recently drawn the attention of many researchers in many areas of applications. Examples of such applications include opinion dynamics in social networks [1], epidemics spreading over a population [40], dynamic adoption of new technologies and products over networks [24], and network security [34]. In this paper, we formulate and study a problem of incentive design in a dynamic networked system motivated by a network security application.

We consider a network of strategic agents who are exposed to cyber-threats and attacks from both outside and their neighbors over time. In this network, evolutions of the agents' security states are inter-temporally correlated as an agent that is occupied by an attacker at time t can be used to launch attacks to his neighbors at time $t+1$. We assume that the agents have interdependent valuations as the utility each agent gets in the network depends not only on his own security state (type), but also on the security states of his neighbors with whom he is connected over the network.

The process of spreading attacks through a network can lead to costly cascades unless controlled by an external intervention. For this purpose, there is a network manager (she) in the system who is responsible for protecting the network against attacks by dynamically allocating certain/limited resources to the agents who are most crucial for security enhancement over time. To determine the crucial agents, the manager needs to know all agents' security states. However, at every time, each agent's security state is his own private information. Therefore, a strategic agent may have an incentive to misrepresent his private security state to strategically manipulate the network manager's allocation decisions so as to increase his own utility. As a result, to elicit the agents' true security states the network manager needs to provide additional incentives to each agent so as to align his individual utility with the network overall security (i.e. social welfare) she is attempting to maximize, considering the agents' correlated types and interdependent valuations.

One of the standard approaches to incentivize strategic agents is to provide monetary incentives to them. Monetary incentives give agents rewards or penalties based on the information they share with the network manager. To provide monetary incentives the network manager must design incentive mechanisms that consist of two components: (1) a message space, that is, a communication alphabet through which the agents can send information to the network manager; and (2) an outcome function that maps vector of messages into allocation decisions and monetary payments/taxes. These components must be designed so as to align the selfish objective of each agent with the social objective and hence motivate agents to reveal their information to the network manager who is maximizing the social welfare, truthfully.

In addition to social optimality, incentive mechanisms are required to satisfy two additional constraints, namely maintaining a *balanced budget* and ensuring *individual rationality*. Budget balance is a requirement on the monetary

incentives the mechanism offers to the agents; it ensures that at equilibrium the network manager redistributes agents' payments entirely and exclusively as rewards to other agents, and on the average she neither retains a surplus as profit nor sustains losses, but merely uses monetary payments as a regulatory tool. Individual rationality ensures that on the average each agent prefers the outcome attained from participating in the mechanism, to what he could attain by unilaterally opting out.

In this paper, we propose a dynamic incentive mechanism for agents with correlated types and interdependent valuations that is individually rational and budget balanced, and enables the network manager to achieve the socially efficient outcome. Our result is in contrast with the existing impossibility results for incentive mechanisms that are socially efficient, individually rational, and budget balanced in static settings [39]. We show that in dynamic settings we can exploit the inter-temporal correlation among the agents' security states so as to design a dynamic incentive mechanism that is social welfare maximizing, individually rational and budget balanced. Specifically, we determine a set of *inference signal*s for the agents' security states over time. Utilizing the proposed set of inference signals, we characterize set of monetary payments for each agent that aligns his objective with the designers' objective, and ensures individual rationality and budget balance.

Moreover, by using a collection of past inference signals instead of just the most recent ones, we present a *reputation-based* payment for every agent that achieves approximate ex-post individual rationality and budget balance of the proposed mechanism. Ex-post individual rationality guarantees that for every realization of the stochastic network's operation (and not only in expectation) agent do not regret participating in the mechanism. Ex-post budget balance ensures that for every realization of the stochastic network's operation the network manager does not incur any budget deficit/surplus.

1.1 Review of Related Works

There is a growing body of literature on network security games (see [34, 36] and references therein). A strand of literature study the interactions between network agents and the attacker as a two-player attacker-defender game where all the agents are treated as one player [7, 27, 33]. This class of papers neglects the strategic interactions among the agents within the networks and treat them as one agent which tries to defend/protect the network against an external strategic attacker. Another class of papers [18, 21, 29, 31, 41, 42], study the interactions among interdependent strategic agents with misaligned objectives within the network as a network game (see [20, 30] and references therein), assuming that the attacker's behavior is exogenously fixed. For instance, the work of [21] studies a network security game and shows that performance at equilibrium compared to the social optimum can be very poor and tends to decrease with increases in network size and the agents' interdependency. In our work, we also study the dynamic interactions among strategic agents within a

network. However, we study the design of an incentive mechanism to improve the overall security in a network rather than analyzing the resulting security game for a given environment.

The existing literature on mechanism design for network security considers mainly static environments where agents make a decision once [26, 32, 39]. The works of [32] and [26] consider static settings and investigate the role of cyber-insurance as an incentive instrument for agents to increase their security investment in self-protection. The authors in [39] studies a general form of mechanism design problems for increasing security in static networks and shows that there exists no incentive mechanism which can implement socially efficient outcome and simultaneously ensures individual rationality and budget balance. In contrast to [26, 32, 39], we study the problem of designing incentive mechanisms in dynamic environments. Our paper contributes to this set of literature by showing that the impossibility result shown in [39] does not apply to dynamic settings.

In dynamic environments, an agent's strategy choice at each time depends on his strategy at other times. Therefore, on one hand, the problem of incentive mechanism design is more challenging in dynamic environments as an agent has more opportunities to deviate by coordinating his strategies over time. But on the other hand, the long-term interactions between the network manager and the agents present in dynamic environments create new opportunities to design invective mechanisms, and thus, offer a richer family of incentive mechanisms to choose from compared to those in static environments. In this paper, utilizing history-dependent monetary payments that exploits the inter-temporal interdependence between agents' security states, we design an incentive mechanism for dynamic environments that implements the socially efficient outcome, ensures the agents' incentive compatibility and individual rationality, and is budget balanced. The fact that the long-term interactions between the agents and the network manager could be beneficial in designing incentive mechanisms has been previously shown within the context of repeated games in [16] in a general but simple setting and in [21, 38] for specific settings motivated by network security applications. Our work is different from those in [16,21,38] that consider repeated game settings in two aspects: (1) We take a mechanism design approach rather than analyzing a fixed game setting. (2) In repeated game settings there is no system dynamics, and the potential improvements in repeated settings compared to static settings demonstrated in [16, 21, 38] is just due to the reputation that agents try to form over time to avoid punishment. Our work provides another insight for potential advantages of dynamic incentive mechanisms by capitalizing on the coupling among the agents' security states over time.

The model we consider in this paper is also related the literature on Susceptible-Infected-Susceptible (SIS) epidemic models over networks when the agents are strategic (see [40] and references therein). The existing literature on SIS epidemic models have mostly studied settings with non-strategic agents. A set of papers assume a homogeneous continuum of agents connected over a random graph model and study the SIS epidemic process using differen-

tial equations (see [2] and references therein). Epidemic SIS models over fixed networks with non-strategic agents have been studied in [11,12,17]. The work of [12] provides a dynamic policy for allocating a fixed amount of security measures to a set of non-strategic agents in the network, at each time instant, so as to minimize the expected extinction time of the epidemic. In [11], the authors provide a lower bound on the expected time for extinction under any such dynamic allocation policy. The impact of network topologies in the persistence of epidemics is studied in [17]. The authors in [15,43] study variations of SIS epidemic models over networks with strategic agents. In [43], the authors investigate a static game where strategic agents make one-time investment decisions in their security which then affect the epidemic process. The work of [15] studies a dynamic marketing problem on networks using a SIS epidemic model, and investigates a game problem between two firms which compete for market shares over the network.

The mechanism design problem we consider in this paper can be viewed as a dynamic resource allocation mechanism with strategic agents. The existing literature on resource allocation mechanisms have mostly focused on static problems. The work of [25] studies the resource allocation problem in a static network with non-strategic agents. For settings with strategic agents, the works in [8,19,44] use the well-known Vickrey-Clark-Groves (VCG) mechanism and achieve the socially efficient outcome when the agents have *private valuations*, meaning that each agent's utility depends only on his own type and the network manager's action, not on the other agents' types; however, the VCG mechanism utilized in [8, 19, 44] is not budget-balanced. The works by Arrow [3] and d'Aspremont and Gerard-Varet [10] propose a static efficient mechanism, called AGV mechanism, that is budget-balanced, but is not individually rational. The authors in [13, 22] take an *implementation theory* approach and propose resource allocation mechanisms for static settings that are simultaneously social welfare maximizing, budget balanced and individually rational. In dynamic settings, the work of [5] presents an extension of the VCG mechanism to a network of agents with dynamic *independent* types and private valuations, which is socially efficient but not budget-balanced. In [4] the authors propose a generalization of the AGV mechanism to a dynamic environment; this mechanism achieves social efficiency and budget balance when the agents have private valuations, but is not individually rational. Liu [35] considers a dynamic mechanism design problem for agents with correlated types and interdependent valuations, and utilizes the correlation among agents' types to provide a mechanism that is socially optimal and budget balanced.

The mechanism we propose in this paper is inspired by [35]. Nevertheless, our results are distinctly different from those of [35] in the following three aspects. First, in contrast to Liu ([35]) who proves only the existence of an efficient dynamic mechanism, we characterize explicitly an efficient dynamic mechanism for the specific model considered in this paper. Second, Liu's mechanism is budget balanced but not individually rational; our proposed mechanism is both individually rational and budget balanced. Third, and most importantly, a drawback of the mechanism proposed in [35] is that

it exposes the agents and the network manager to the risk of negative utility and money deficit, respectively, for some realizations of the network's stochastic evolution. Our proposed mechanism alleviates this drawback by ensuring the approximate ex-post individual rationality and budget balance, and this effectively eliminates any risk for the agents and the network manager.

1.2 Outline of the Paper

The rest of the paper is organized as follows. We present our model for the dynamic network security problem with strategic agents in Section 2. We formulate the dynamic incentive design problem in Section 3 and describe the mechanism we propose as the solution in Section 4. In Section 5, we show that the proposed mechanism can incentivize agents to reveal their information truthfully, while ensuring budget balanced and individual rationality. In Section 6, presenting reputation-based payments, we enable our incentive-compatible mechanism to achieve both approximate ex-post individual rationality and budget balance. In Section 7 we study the problem the network manager must solve to find a socially efficient outcome after the agents disclose their information. We conclude our paper in Section 8. The proofs of all the results that appear in this paper can be found in the appendix.

2 Model

We consider an environment with n strategic agents who are living in distinct nodes of an interconnected network interacting over time $t \in \mathcal{T} := \{0, 1, 2, \ldots\}$. There is a network manager in the system who is responsible for the security of the network and for protecting it against attacks. We represent the network by a directed graph $G = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N} = \{1, ..., n\}$ denotes the set of strategic agents and $\mathcal{E}$ denotes the set of directed edges. Agent j is said to be a neighbor of agent i if there is a directed edge from j to i in the graph, i.e. $(j, i) \in \mathcal{E}$. Let $\mathcal{N}^i := \{j : (j, i) \in \mathcal{E}\}$ denote the set of agent i's neighbors. The security of agent i *depends* on the security of each of his neighbors $j \in \mathcal{N}^i$, as attacks could be propagated from j to i through the edge (j, i). The strength of this dependence is determined by the probability of spreading attacks from j to i which is denoted by $l_{ji} \in (0, 1]$. Let $L := (l_{ji})$ denote the *dependence* matrix of the graph. The dependence matrix L contains non-zero elements l_{ji} if and only if $j \in \mathcal{N}^i$.

At each time $t \in \mathcal{T}$, each agent $i \in \mathcal{N}$ can be in one of two security states: safe or unsafe/attacked. We denote the security state of agent i at time t by $\theta_t^i \in \Theta := \{0, 1\}$, where $\theta_t^i = 1$ indicates that agent i is safe and $\theta_t^i = 0$ indicates that he is unsafe. At any time $t \in \mathcal{T}$, if agent i is safe, *i.e.* $\theta_t^i = 1$, he may be attacked directly from outside with probability d_i, or indirectly from any of his unsafe neighbors $j \in \mathcal{N}^i$ with probability l_{ji}. The topology of the network G, the dependence matrix L, and the probabilities of external attacks

d_i, $i \in \mathcal{N}$, remain fixed over time. At each time t, the security state θ_t^i of agent i, which is also referred to as agent i's *type*, is his own private information and cannot be observed by either the network manager or any other agent. We denote the security state of the network at time t by $\boldsymbol{\theta}_t = (\theta_t^1, \ldots, \theta_t^n) \in \Theta^n$.

The network manager's goal is to maximize the overall security of the network, which is measured by the social-welfare. At each time t, the network manager has one security measure and can selectively choose one of the agents $a_t \in \mathcal{N}$ to whom she applies the security measure. Applying the security measure enhances the security of the chosen agent a_t in two ways: (1) It helps a_t to restore his safety, if he is in the unsafe state, i.e. $\theta_t^{a_t} = 0$; (2) It builds a firewall around agent a_t that may protect him from external attacks. These two objectives are met independently, each with probability of success $h \in (0, 1]$.

During each time $t \in \mathcal{T}$, as a result of network manager's action a_t as well as new direct attacks from outside, and the propagation of internal attacks in the network, the network's state evolves from $\boldsymbol{\theta}_t$ to $\boldsymbol{\theta}_{t+1}$ according to the following Markovian dynamics:

$$\mathbb{P}\{\boldsymbol{\theta}_{t+1} = \boldsymbol{b} | \boldsymbol{\theta}_t, a_t\} = \prod_{i=1}^{n} \mathbb{P}\{\theta_{t+1}^i = b_i | \boldsymbol{\theta}_t, a_t\}, \ \forall \boldsymbol{b} = (b_1, \ldots, b_n) \in \Theta^n, \quad (1)$$

where,

$$\mathbb{P}\{\theta_{t+1}^i = 1 | \boldsymbol{\theta}_t, a_t\} = \begin{cases} 0, & \theta_t^i = 0, i \neq a_t \\ h(1 - d_i(1-h)) \prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji}), & \theta_t^i = 0, i = a_t \\ (1 - d_i) \prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji}), & \theta_t^i = 1, i \neq a_t \\ (1 - d_i(1-h)) \prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji}), & \theta_t^i = 1, i = a_t \end{cases}, \quad (2)$$

and $\mathbb{P}\{\theta_{t+1}^i = 0 | \boldsymbol{\theta}_t, a_t\} = 1 - \mathbb{P}\{\theta_{t+1}^i = 1 | \boldsymbol{\theta}_t, a_t\}$. We assume that the external attacks and spreading of internal attacks within the network are independent across different agents. Therefore, given the previous state $\boldsymbol{\theta}_t$ and the network manager's action a_t, the state transitions of each agent occur independently as in (1). Equation (2) describes these transitions: (i) if agent i is unsafe and he does not receive any security measure from the network manager at time t, he remains to be unsafe at the next time $t + 1$; (ii) if agent i is unsafe, but he receives the security measure, he restores his safety if the security measure is successful in its first objective (prob h). Agent i will keep this safety until the next time $t + 1$, if he is not subject to any external attacks from outside (prob $1 - d_i(1 - h)$) or any internal attacks from his unsafe neighbors (prob $\prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji})$); (iii) if agent i is safe and he does not receive the security measure, he will remain in the safe state if he is not attacked from outside (prob $(1 - d_i)$) and he is not attacked from his neighbors (prob $\prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji})$); (iv) if agent i is in the safe state and he is receiving the security measure, he will remain safe if he is not attacked from outside (prob $(1 - d_i(1 - h))$) and he is not attacked from his unsafe neighbors (prob $\prod_{j \in N^i : \theta_t^j = 0} (1 - l_{ji})$). It is clear from (2) that in this problem, the agents' types are *correlated* with each

other, as the evolution of each agent's type depends on his neighbors' security types (whether they are safe of unsafe).

Moreover, agents have *interdependent valuations*, meaning that each agent's utility is not only affected by his own security state, but also depends on the security states of his neighbors. At each time $t \in \mathcal{T}$, each agent $i \in \mathcal{N}$ has an interdependent valuation that depends on his own security state θ_t^i, the *security index* of his neighborhood s_t^i, and the security measure he may or may not receive from the network manager. The security index of agent i's neighborhood at time t is defined as the weighted average of his neighbors' security states $\left\{\theta_t^j : j \in \mathcal{N}^i\right\}$, where the weights are the dependencies l_{ji} of agent i on each of his neighbors $j \in \mathcal{N}^i$. That is,

$$s_t^i(\boldsymbol{\theta}_t) = \frac{\sum_{j\in\mathcal{N}^i} l_{ji}\theta_t^j}{\sum_{j\in\mathcal{N}^i} l_{ji}}.^{1} \tag{3}$$

Specifically, the valuation of agent i at time t is given by

$$v^i(\boldsymbol{\theta}_t, a_t) = \theta_t^i + \alpha \mathbf{1}_{\left\{\theta_t^i=1 \text{ or } a_t=i\right\}} s_t^i(\boldsymbol{\theta}_t), \tag{4}$$

where $\mathbf{1}_{\{A\}}$ is the indicator function of an event A and $0 < \alpha < 1$ captures the value of a safe neighborhood to an agent i. According to (4), agent i cares about the security index of his neighborhood only if he is in the safe state or he is receiving a security measure at time t, *i.e.* $\{\theta_t^i = 1 \text{ or } a_t = i\}$. This is because when agent i is unsafe and does not receive any security measure at time t, irrespective of his neighbors' security states, he has no chance to become safe at the next time $t+1$.

At each time $t \in \mathcal{T}$, every agent $i \in \mathcal{N}$ may receive or make a monetary payment p_t^i to the network manager. This payment can be any positive or negative real number; if $p_t^i > 0$, then agent i pays money (tax), whereas $p_t^i < 0$ implies that agent i receives money (subsidy). Therefore, the total instantaneous utility of agent i at time t is given by,

$$u_t^i(\boldsymbol{\theta}_t, a_t, p_t^i) = v^i(\boldsymbol{\theta}_t, a_t) - p_t^i. \tag{5}$$

All agents discount the future with a common discount factor $\delta \in (0,1)$ which measures how much the agents value a future utility over the current one. Therefore, the total utility of each agent i from time $t = 0$ to ∞ is given by

$$U^i = (1-\delta)\sum_{t=0}^{\infty} \delta^t u_t^i(\boldsymbol{\theta}_t, a_t, p_t^i) = (1-\delta)\sum_{t=0}^{\infty} \delta^t [v^i(\boldsymbol{\theta}_t, a_t) - p_t^i], \tag{6}$$

[1] Notice that from the technical point of view, the weights in this weighted average do not need to be necessarily the same as l_{ji}. We can define the security index of agent i's neighborhood based on any arbitrary set of weights and all the results continue to hold. However, the dependencies l_{ji}s are the most natural choice according to the model.

where $(1-\delta)$ is a normalization factor. Each strategic agent makes strategic decisions to maximize his own utility function U^i. However, the network manager's objective is to maximize the overall expected security of the network, expressed by the social welfare W defined by

$$W = \mathbb{E}\{(1-\delta)\sum_{t=0}^{\infty}\delta^t\sum_{i=1}^{n}v^i(\boldsymbol{\theta}_t, a_t)\}. \tag{7}$$

The network manager's problem would be a standard stochastic control problem (Markov decision problem) if she knew the agents' security states $\boldsymbol{\theta}_t$ at each time t. However, this information is unavailable to her, as for each $i \in \mathcal{N}$, θ_t^i is agent i's private information. Thus, in order to make an optimal decision at any time t, i.e. to choose the agent to whom she applies the security measure, the network manager has to elicit information about each agent's security state. Since agents are strategic and are only working towards their own objective, they do not voluntarily reveal their information to the network manager whose objective is different from theirs. As it can be seen in (2), if it had not been for the (incentive) payments p_t^i, each agent would always prefer to receive the security measure at every time. Therefore, the manager needs to design an incentive mechanism so as to align the agents' objectives with her own objective. In the next section, we formulate the network manager's problem as a dynamic mechanism design problem. Then, in Section 4, we present a dynamic incentive mechanism to solve the problem.

3 Dynamic Incentive Design Problem

A dynamic incentive mechanism specifies the set of messages $\mathcal{M}_t^i$ that each agent $i \in \mathcal{N}$ can use at each time $t \in \mathcal{T}$ to transmit information to the network manager, along with the decision rule the manager *commits* to use for decision making based on the transmitted messages. A mechanism could be very complicated since there is no restriction on the set of messages $\mathcal{M}_t^i$. However, the revelation principle for dynamic games [37] states that without loss of generality the network manager can restrict her attention to dynamic *direct mechanisms* satisfying *incentive compatibility*.

A direct mechanism is a special class of mechanisms where each agent's set of messages at each time t is the same as his type space, i.e. $\mathcal{M}_t^i = \Theta$, $i \in \mathcal{N}$, $t \in \mathcal{T}$. The term *direct* is due to the fact that at each time agents are asked to directly report their private information. In response to this request, at every time t, each agent i publicly reports a type r_t^i which is not necessarily the same as his actual security type θ_t^i. Then, a public allocation decision a_t and a transfer p_t^i to each agent i are made as functions of the current report profile $\boldsymbol{r}_t = (r_t^i, i \in \mathcal{N}) \in \Theta^n$ and the time-t public history h_t. The time-t public history contains all reports and allocations up to time $t-1$, i.e.,

$$h_t := \{\boldsymbol{r}_s, a_s, s \leq t-1\}. \tag{8}$$

Let $\mathcal{H}_t$ denote the set of all possible public histories at time t. Formally, a dynamic direct mechanism $(a(\cdot), p(\cdot)) = \{a_t(\cdot), p_t^i(\cdot), i \in \mathcal{N}, t \in \mathcal{T}\}$ consists of two components: (i) an allocation policy $a(\cdot) = (a_t(\cdot), t \in \mathcal{T})$, where $a_t : \Theta^n \times \mathcal{H}_t \to \mathcal{N}$ determines the agent who receives the security measure at time t, and (ii) a tax function $p(\cdot) = (p_t^i(\cdot), i \in \mathcal{N}, t \in \mathcal{T})$, where $p_t^i : \Theta^n \times \mathcal{H}_t \to \mathbb{R}$ determines the monetary payment (or the negative of the monetary incentive) that agent i makes (receives) at time t based on the current report profile $\boldsymbol{r}_t$ and the time-t public history h_t.

In this mechanism, a reporting strategy for agent i at time t is a function $\sigma_t^i : \Theta \times \mathcal{H}_t^i \to \Delta(\Theta)$ that maps each pair of his time-t security type $\theta_t^i \in \Theta$ and time-t private history $h_t^i \in \mathcal{H}_t^i$ to a probability distribution on his type space according to which agent i chooses his report. Here $\Delta(\Theta)$ denotes the set of all probability distributions on Θ. The private history h_t^i of agent i at time t consists of the time-t public history h_t and the sequence of agent i's private information $\{\theta_s^i, s \leq t-1\}$ up to time $t-1$, i.e.,

$$h_t^i = \{\boldsymbol{r}_s, a_s, \theta_s^i, s \leq t-1\}. \tag{9}$$

A dynamic direct mechanism is *incentive compatible* (*IC*) if at every time t, every agent i prefers truth-telling strategy to all future reporting strategies $\{\sigma_\tau^i, \tau \geq t\}$ he can adopt, given that the other agents report truthfully, i.e.

$$\begin{aligned}
&\mathbb{E}\{(1-\delta)\sum_{\tau=t}^{\infty}\delta^{\tau-t}\Big[v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)) - p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)\Big]\} \geq \\
&\mathbb{E}\{(1-\delta)\sum_{\tau=t}^{\infty}\delta^{\tau-t}\Big[v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \sigma_\tau(\theta_\tau^i, h_\tau^i), h_\tau)) - p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \sigma_\tau(\theta_\tau^i, h_\tau^i), h_\tau)\Big]\},
\end{aligned} \tag{10}$$

for all i, t, $\{\boldsymbol{\theta}_\tau, \tau \leq t\}$ and $\{\sigma_\tau^i, \tau \geq t\}$, where the expectation is taken with respect to the dynamics of the network given by (1) and (2). In other words, a direct mechanism is incentive compatible if and only if truth-telling is a Bayesian Nash equilibrium.

Since the network manager cannot force strategic agents to join the mechanism, she needs to ensure that the agents voluntarily participate in the mechanism. Agent i voluntarily participates in the mechanism $(a(.), p(.))$ if on the average, the utility he gains at the truth-telling equilibrium, is greater than or equal to the *reservation utility/outside option* $U_0^i(a)$ he obtains when he does not participate in the mechanism. We note that the outside option is endogenous, i.e. it depends on the specification of the mechanism. This is because, when agent i unilaterally opts out of the mechanism, his utility still depends on the allocation rule a through the externality he receives from the other agents' participation. Therefore, agents' voluntary participation in the mechanism is ensured by the following *individual rationality* (*IR*) constraints

as follows,

$$\mathbb{E}\{(1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}\Big[v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i},\theta_\tau^i,h_\tau))-p_\tau^i(\boldsymbol{\theta}_\tau^{-i},\theta_\tau^i,h_\tau)\Big]\}\geq\mathbb{E}\{U_0^i(a)\},\forall i\in\mathcal{N}. \tag{11}$$

Furthermore, it is desired that, at truth-telling equilibrium, the expected value of all money collected from the agents (i.e. taxes) equals the expected value of all money paid out to them (i.e. subsidies). This means that the entity running the mechanism neither profits nor subsidizes but merely uses monetary payments as a regulatory tool. This requirement called *Budget Balance* (BB), can be written as

$$\mathbb{E}\{(1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}\sum_{i\in\mathcal{N}}p_\tau^i(\boldsymbol{\theta}_\tau^{-i},\theta_\tau^i,h_\tau)\}=0. \tag{12}$$

Therefore, we can formulate the dynamic incentive design problem for the network manager as follows:

$$\max_{a(\cdot),p(\cdot)}\ \mathbb{E}\{(1-\delta)\sum_{t=0}^{\infty}\delta^t\sum_{i=1}^{n}v^i(\boldsymbol{\theta}_t,a_t)\}, \tag{13}$$

subject to IC constraints (10), IR constraints (11) and BB constraints (12).

The incentive design problem formulated in (13) is a *dynamic* mechanism design problem with agents with *correlated types* and *interdependent valuations*. It is dynamic in the strategic sense since the incentive constraints of the agent at each time t depend on their decision strategies at other times (see 10). Moreover, according to (1)-(2), the evolution of each agent's security type depends on the other agents' security types; therefore, the types of different agents are inter-temporally correlated. Furthermore, because of (3)-(4), each agent's valuation is not only affected by his own security state, but also influenced by the security states of his neighbors; hence the agents have interdependent valuations. As a result of the correlation among agents' types and the interdependency among agents' valuations, the dynamic extensions of the classical Vickrey-Clarke-Groves (VCG) and d'Aspremont and Gerard-Varet (AGV) mechanisms [4,5] cannot be used to solve the network manager's problem (13).

In this paper, we present an alternative approach to the dynamic incentive design problem which is suitable for the model described in Section 2 with coupled dynamics and interdependent valuations. We utilize the inter-temporal correlation among agents' types to construct a *cross inference signal* about the security type of agent i at each time which is independent of his own reports. We use this signal to internalize the effect of each agent's security state on the overall network security through a set of incentive payments. The idea of utilizing the correlation among agents' types to extract their private information was first exploited by Cremer and McLean in a static setting [9].

They formed a cross inference signal for each agent by utilizing the correlation among the realization of agents' types, determined appropriate incentive payments that depend on the cross inference signals, and extracted the agents' private information. However in dynamic settings, a time-by-time extension of Cremer and McLean's mechanism is not incentive compatible, because in dynamic environments agents have more opportunities to deviate and each agent can coordinate his strategies over time. We describe our approach for designing an incentive compatible mechanism for dynamic settings below.

4 Specification of the Mechanism

We present a 'Dynamic Cross Inference' (DCI) mechanism that maximizes the social welfare and satisfies the IC, IR and BB constraints; hence it solves the network manager's problem (13). The description of our mechanism is divided into two parts: the allocation policy $(a_t(\cdot), t \in \mathcal{T})$, and the monetary transfers $(p_t^i(\cdot), i \in \mathcal{N}, t \in \mathcal{T})$.

Allocation Policy: The specification of the allocation policy is based on the premise that the mechanism is incentive compatible. In an incentive compatible mechanism, at each time t, the agents report their true security state to the network manager, i.e. $r_t^i = \theta_t^i$, $i \in \mathcal{N}, t \in \mathcal{T}$. Therefore, after receiving the reports the network manager is confronted with a stochastic control problem with complete information. In Section 7, we show that this control problem always has an optimal *stationary* solution $a^* : \Theta^n \to \mathcal{N}$ that specifies an agent that must receive the security measure for each state $\boldsymbol{\theta}_t$ of the network, independent of the time t and the public history h_t. We choose this optimal stationary solution as the allocation policy of our mechanism, i.e. $a_t(\boldsymbol{r}_t, h_t) = a^*(\boldsymbol{r}_t), \forall t \in \mathcal{T}, \forall h_t \in \mathcal{H}_t$. In Section 7, we discuss how the network manager can find such an optimal policy.

Monetary Transfers: To obtain an incentive compatible mechanism, we design monetary transfers so that they align each agent's objective with the social welfare. By doing so, there is no conflict of interests between the agents and the network manager; thus, the agents have no objection to revealing their private information to the network manager. The most common approach in the literature for reconciling the conflict of interests between the agents and the manager is to convert the individual utility of each agent into the social welfare function W (7) by simply paying each agent i the total valuations of other agents. This approach, which is the central idea behind Groves' mechanisms, works only if the agents have private valuation and hence the total valuation of all agents except i depends on the report of agent i only through the determination of the social allocation. However, this condition is not satisfied in our problem because the agents' valuations are interdependent. Indeed, from (3)-(4) it is clear that the valuations of all agents except i depend directly on agent i's report.

To resolve this issue, we utilize the correlation between agent i's security state θ_t^i at time t and other agents' security states θ_{t+1}^j, $j \neq i$, at time $t+1$,

and form a cross inference signal about the security state of agent i that is independent of his own reports. We use this cross inference signal to construct a set of payments for time $t+1$ such that, in expectation, every agent $i \in \mathcal{N}$ receives the sum of all other agents' valuation flow at t. By doing so, agent i's continuation payoff at time t is equal to the social surplus from time t onward. Hence the agents are willing to reveal their true security types to the network manager with whom they have no conflict of interest.

Specifically, let $\boldsymbol{r}_t^{-i}$ denote the report profile of all agents except i at time t. We define the cross inference signal for agent i at time t as follows:

$$m_t^i = \begin{cases} 0, & \text{if } r_{t+1}^j = 0, \forall j \in O^i, \\ 1, & \text{otherwise,} \end{cases} \tag{14}$$

where $O^i := \{j \in \mathcal{N} : i \in \mathcal{N}^j\}$ is the set of agents whose securities are influenced by agent i's security. We call these agents *output neighbors* of agent i. If at time $t+1$, all output neighbors of agent i report to be unsafe, the manager interprets this as a signal that agent i was unsafe at time t. Otherwise, she assesses agent i as a safe agent at time t.

By using the cross inference signal m_t^i, we define the tax $p_{t+1}^i(m_t^i, \boldsymbol{r}_t^{-i}, a_t)$ to be paid by agent i, $i \in \mathcal{N}$, at time $t+1$, as follows:

$$p_{t+1}^i(1, \boldsymbol{r}_t^{-i}, a_t) = \frac{1}{\delta}\Big[-\sum_{j \neq i} v^j(\theta_t^i = 1, \boldsymbol{\theta}_t^{-i} = \boldsymbol{r}_t^{-i}, a_t) - \frac{\mathcal{P}(m_t^i = 0 | \theta_t^i = 1, \boldsymbol{r}_t^{-i}, a_t)}{\mathcal{P}(m_t^i = 0 | \theta_t^i = 0, \boldsymbol{r}_t^{-i}, a_t) - \mathcal{P}(m_t^i = 0 | \theta_t^i = 1, \boldsymbol{r}_t^{-i}, a_t)} \; \alpha \sum_{\substack{j \in O^i: \\ r_t^j = 1 \text{ or } a_t = j}} \frac{l_{ij}}{\sum_{k \in \mathcal{N}^j} l_{kj}} \Big], \tag{15}$$

and

$$p_{t+1}^i(0, \boldsymbol{r}_t^{-i}, a_t) = \frac{1}{\delta}\Big[-\sum_{j \neq i} v^j(\theta_t^i = 0, \boldsymbol{\theta}_t^{-i} = \boldsymbol{r}_t^{-i}, a_t) + \frac{\mathcal{P}(m_t^i = 1 | \theta_t^i = 0, \boldsymbol{r}_t^{-i}, a_t)}{\mathcal{P}(m_t^i = 0 | \theta_t^i = 0, \boldsymbol{r}_t^{-i}, a_t) - \mathcal{P}(m_t^i = 0 | \theta_t^i = 1, \boldsymbol{r}_t^{-i}, a_t)} \; \alpha \sum_{\substack{j \in O^i: \\ r_t^j = 1 \text{ or } a_t = j}} \frac{l_{ij}}{\sum_{k \in \mathcal{N}^j} l_{kj}} \Big], \tag{16}$$

where $\mathcal{P}(m_t^i | \theta_t^i, \boldsymbol{r}_t^{-i}, a_t)$ is the probability of m_t^i given θ_t^i, $\boldsymbol{r}_t^{-i}$ and a_t, assuming truthful reports of all agents except i. The tax incentives (15) and (16) of our mechanism consist of two components. The first component in (15) and (16) is an approximation of the sum of valuations of all agents except i; this approximation is derived by assuming that the inference signal m_t^i reveals the correct type of agent i at time t, i.e. $\theta_t^i = m_t^i$. The second component of the tax in (15) and (16) is a reward or a punishment based on the network manager's inference m_t^i about the agent i's type at time t. If the network

manager interprets i as a safe agent at time t, i.e. $m_t^i = 1$, she rewards him by paying a fraction of the positive effect $\alpha \sum_{j \in O^i : r_t^j = 1 \text{ or } a_t = j} \frac{l_{ij}}{\sum_{k \in \mathcal{N}^j} l_{kj}}$ he had on the total valuations of his neighbors at time t (see 15). Otherwise, she charges agent i a fraction of the negative effect he imposed on the network (see 16). The sum of these components is multiplied by a factor of $1/\delta$ as this tax actually accounts for the performance of the network at time t, but it is charged with one time step delay at time $t+1$.

In the next section we show that the tax function $p_{t+1}^i(m_t^i, \boldsymbol{r}_t^{-i}, a_t)$ incentivizes agents to tell the truth, however, when the agents adopt truthful strategies, the total amount of monetary transfers the network manager receives from the agents is negative. This means that the mechanism runs a budget deficit subsidizing agents. Therefore, to balance the budget, the network manager charges each agent i a participation fee $\tilde{p}_0^i$ equal to the discounted value of the subsidies he will get in the future. That is,

$$\tilde{p}_0^i = -\mathbb{E}\{\sum_{t=0}^{\infty} \delta^{t+1} p_{t+1}^i(m_t^i, \boldsymbol{\theta}_t^{-i}, a_t^*)\}, \tag{17}$$

where the expectation is taken with respect to the dynamics of the security network and the initial distribution of the security states, which is assumed to be common knowledge among the network manager and the agents, assuming truthful reports of the agents. At time $t = 0$, after the mechanism is announced to the agents and before realizing the first period's security state θ_0^i, each agent i decides whether or not to participate in the mechanism. If he decides to participate, he should pay the participation fee $\tilde{p}_0^i$ given by (17). The manager runs the mechanism only if all the agents decide to participate.

5 Properties of the DCI Mechanism

We prove that the DCI mechanism proposed in Section 4 solves the dynamic incentive design problem (13) for the network manager. We establish this result by proceeding as follows. We first show that using the DCI mechanism the network manager is able to align each agent's interests with the social welfare objective function (Lemma 1). We use this property to show that the mechanism is incentive compatible (Proposition 1) and can implement the socially efficient outcome (Proposition 2). Then, we show that at the truth-telling equilibrium the expected sum of the monetary incentives provided to all strategic agents is equal zero. This result establishes that the mechanism is budget balanced (Proposition 3). Finally, we prove that the positive benefit created for each agent by the mechanism is high enough that strategic agents are willing to pay the participation fees and participate in the mechanism. This result establishes individual rationality of the DCI mechanism (Proposition 4). We combine all these results in Theorem 1 to show that the DCI mechanism solves the network manager's problem (13).

We present the proofs of the following propositions, theorems and lemmas in the Appendix.

Lemma 1 The discounted expected value of the incentive payment every agent $i \in \mathcal{N}$ gets at time $t+1$ is equal to the sum of all other agents' instantaneous valuation at time t; i.e.

$$\delta \, \mathbb{E}\{-p_{t+1}^i(m_t^i, \boldsymbol{\theta}_t^{-i}, a_t)\} = \sum_{j \neq i} v^j(\boldsymbol{\theta}_t, a_t), \forall \boldsymbol{\theta}_t, a_t, \tag{18}$$

where the expectation is taken with respect to the inference signal m_t^i, assuming truthful reports of all agents except i.

Lemma 1 shows that even though the incentive payment $p_{t+1}^i(m_t^i, \boldsymbol{\theta}_t^{-i}, a_t)$ is independent of both agent i's type and report, in expectation, it can convert agent i's individual utility into the social objective. Therefore, the DCI mechanism eliminates the conflict of interests between the network manager and the agents, so agents are willing to reveal their true security states to the network manager.

Proposition 1 The DCI mechanism with monetary payments (15)-(16) and participation fees (17) is incentive compatible.

Since the mechanism is incentive compatible, agents report their true types to the network manager. Therefore, the outcome of allocating resources based on the policy a^*, which is optimal under the complete information assumption, coincides with the socially efficient outcome. These arguments establish the following result.

Proposition 2 The DCI mechanism maximizes the social welfare function W at its truth-telling equilibrium.

At truth-telling equilibrium, the expected value of the monetary incentives each agent receives is equal to the participation fee (17) he has already paid. Therefore, the manager neither profits nor subsidizes from the payments, but solely pays the participation fee of each agent back to him as an incentive across time, so as to incentivize him to tell the truth. This argument establishes the following result.

Proposition 3 The DCI mechanism is budget balanced.

The next proposition shows that even though on the average, the agents earn no 'monetary' benefit from participating in the mechanism, the non-monetary benefits they get from receiving the security measures is high enough that the agents are willing to participate in the DCI mechanism.

Proposition 4 The DCI mechanism satisfies the voluntarily participation requirement (11), and hence is individually rational.

We are now ready to state the main result of this section.

Theorem 1 The DCI mechanism is budget balanced, satisfies the IC and IR constraints, and maximizes the social welfare W. Therefore, it is an optimal solution to the network manager's dynamic incentive design problem (13).

6 Ex-post Individual Rationality and Ex-post Budget Balance

The form of individual rationality (IR) we have considered so far, is ex-ante IR, where agents choose to participate in the mechanism at the beginning of the time and before they even know their own initial security type. In an ex-ante individually rational mechanism, each agent finds participating in the mechanism, *in expectation*, more beneficial than opting out. However, this form of IR does not guarantee the satisfaction of an agent from participating in the mechanism after all information has been revealed and decisions and transfers have been fully specified. In other words, although on the average each agent gains from participating in the mechanism, since the allocations and payments are uncertain and depend on the stochastic events that occur in the network, there is always a risk for the agent to join. A mechanism is risk-free for the agents if it satisfies the *ex-post individual rationality* constraint, meaning that all agents gain from participating in the mechanism along any realization of the information, payments and transfers.

The uncertainty of the payments is also a risk for the network manager. According to Theorem 3, in the DCI mechanism the budget is balanced in expectation. However, there is no guarantee that the mechanism does not run large deficits in any single run. To assert budget balance along any realization of the stochastic events, we need to satisfy an *ex-post budget balance* constraint.

In this section, by introducing a *reputation*-based payment we make the above-mentioned risks for both the network manager and the agents arbitrarily small. By doing so, the mechanism becomes approximate ex-post individually rational and budget balanced.

Definition 1 We call a mechanism approximate ex-post IR and BB, if for every given $\eta, \zeta > 0$, the mechanism can adjust its payments so that

$$\mathbb{P}\{(1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}\Big[v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)) - p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)\Big] \leq U_0^i(a) - \eta\} \leq \zeta, \tag{19}$$

for all $i \in \mathcal{N}$, and

$$\mathbb{P}\{|(1-\delta)\sum_{\tau=0}^{\infty}\delta^{\tau}\sum_{i\in\mathcal{N}} p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)| \geq \eta\} \leq \zeta. \tag{20}$$

That is, for any given $\eta, \zeta > 0$, the mechanism can make both the probabilities that an agent losses more than η by participating in the mechanism, and the network manager runs a deficit or surplus larger than η, below ζ. The following lemma provides a sufficient condition for a mechanism to be approximate ex-post IR and BB.

Lemma 2 A mechanism is approximate ex-post IR and BB, if 1) each agent's expected payment in the mechanism is zero; and 2) it can make the variance of the agents' payments arbitrarily small.

The first condition of Lemma 2 is already satisfied in the DCI mechanism. Therefore, to make the DCI mechanism approximate ex-post IR and BB, we only need to make the variance of the agents' payments arbitrarily small. For that matter, we need the following technical assumption:

Assumption 1 The evolution of the system under the optimal allocation policy a^* is described by an *ergodic uni-chain*.

An ergodic uni-chain is a generalization of an ergodic Markov chain that allows for some transient states. The purpose of Assumption 1 is to ensure that the probability distribution over the states of the Markov chain converges as time goes to infinity, i.e. the limiting distribution exists. The following theorem provides an intuitive condition sufficient to guarantee that Assumption 1 is satisfied.

Theorem 2 Assumption 1 is satisfied if the optimal policy always applies the security measure to one of the *unsafe* agents of the network (if such an agent exists) provided that the probabilities of both external attacks d_i, $i \in \mathcal{N}$, and internal attacks l_{ij}, $i \in \mathcal{N}, j \in \mathcal{N}^i$, are strictly between 0 and 1.[2]

Under Assumption 1, we present a modified payment function that depends on the history of the inference signals $\{m_s^i, s \leq t\}$ rather the current one m_t^i, and satisfies the second condition of Lemma 2 maintaining the properties stated in Theorem 1. In the original form of the DCI mechanism, the tax $p_{t+1}^i(m_t^i, \boldsymbol{r}_t^{-i}, a_t)$ each agent i pays at time $t+1$ is determined only based on the network manager's assessment m_t^i about his safety at time t. This dependency on only one assessment makes the variance of agent i's payment high, since any change in the assessment m_t^i would imply a corresponding change in agent i's payment. Now consider a tax policy with finite memory K where agent i's tax at each time $t+1$ is determined based on the network manager's assessments of agent i's security types during the last K periods, i.e. $m_{t-K+1}^i, \ldots, m_t^i$. Under this policy, the history of previous assessments builds a reputation for agent i that makes his payments less dependent on one single assessment, and hence reduces the variance.

With this idea in mind, in contrast to the original memoryless tax policy, the network manager provides a history-dependent tax policy by splitting the tax p_t^i of each agent i at each time t defined by (15-16), over the K next periods through an installment agreement, such that the discounted sum of his payments is equal to p_t^i. The simplest installment agreement is to pay p_t^i in K equal-valued payments $\hat{p}_{t,t+k}^i(\delta, K)$, $k = 0, \ldots, K-1$, so that the discounted value of the payment $\hat{p}_{t,t+k}^i(\delta, K)$ at time t equals p_t^i/K. That is,

$$\hat{p}_{t,t+k}^i(\delta, K) = \frac{p_t^i}{K\delta^k}, \qquad \forall k = 0, \ldots, K-1, \tag{21}$$

[2] In Section 7, we show that the sufficient condition of Theorem 2 is not always satisfied, that is there exist some network instances where the optimal policy chooses a safe agent to apply the security measure.

where $\hat{p}^i_{t,t+k}(\delta, K)$ is the part of p^i_t that must be paid by agent i at time $t+k$, when discount factor is δ and the agreement period is K. Based on (21), we can derive the total payment of agent i at time t as the summation of all installments agent i pays at time t:

$$\hat{p}^i_t(\delta, K) = \sum_{k=0}^{\min(K-1,t-1)} \hat{p}^i_{t-k,t}(\delta, K) = \sum_{k=0}^{\min(K-1,t-1)} p^i_{t-k}/K\delta^k. \tag{22}$$

Then the total discounted payment of agent $i \in \mathcal{N}$ over time is

$$\hat{p}^i(\delta, K) = (1-\delta)\sum_{t=0}^{\infty} \delta^t \hat{p}^i_t(\delta, K). \tag{23}$$

In the next theorem, we show that if agents are patient enough (use sufficiently large discount factor δ), the network manager can choose an appropriate K so as to make the variance of each agent's payment $\hat{p}^i(\delta, K)$ arbitrarily small.

Theorem 3 For each $\epsilon > 0$, there exists a threshold $\delta_0 < 1$, such that for all $\delta > \delta_0$, the network manager can make the variance of each agent's payment smaller than ϵ.

Theorem 3 proves that the mechanism satisfies both conditions of Lemma 2 for sufficiently patient agents. Therefore, the mechanism is both approximate ex-post individually rational and budget balanced. This argument establishes the main result of this section.

Theorem 4 For δ sufficiently close to 1, the DCI mechanism with reputation-based payments maximizes the social welfare W, and ensures approximate ex-post individual rationality and budget balance.

7 On Dynamic Optimal Policy for the Network Manager

In this section, we study the problem the network manager must solve to find an optimal allocation policy a^* when the agents reveal their security states $\{\boldsymbol{\theta}_t\}$ truthfully. In this case, the network manager is faced with an infinite horizon Markov decision process (MDP) with perfect observations and time-independent dynamics and rewards, where the transition probabilities are given by (1)-(2) and the instantaneous reward is $r(\boldsymbol{\theta}_t, a_t) := \sum_{i=1}^{n} v^i(\boldsymbol{\theta}_t, a_t)$. For this class of problems it is known that there exists *stationary Markov* policies that are optimal [28]. An optimal stationary Markov policy can be determined by the solution of a linear program that can be solved in polynomial time [23].

An interesting feature of the solution is that although in most cases the network manager prefers to apply the security measure to an unsafe agent, there are situations where it is optimal to allocate the security measure to a safe

agent. In these cases, the level of threats the network is facing is so high that the manager gives up on restoring the safety of the already attacked agents and focuses on keeping the rest of the network safe. Below, we provide conditions sufficient to ensure that it is optimal to allocate the security measure to an unsafe agent (if it exists). Furthermore, we present an instances where this sufficient condition is not satisfied, and it is optimal for the network manager to allocate the security measure to a safe agent.

Theorem 5 Consider a fully connected network with n agents where $d_i = d$, and $l_{ij} = l$, for all $i, j \in \mathcal{N}$. If $d(1-h) \leq l$, that is, the probability of direct attack to a protected agent is less than or equal to the probability of spreading attack through a link, it is optimal to apply the security measure to an unsafe agent at every state with at least one unsafe agent.

Theorem 5 provides a condition sufficient to satisfy the condition of Theorem 2, and thus Assumption 1. When this condition is not satisfied, we can find instances where it is always optimal to apply the security measure to a safe agent. One such instance is a network with four agents and the following parameters: $d = 0.5$, $h = 0.2$ and $l = 0.1$. In this network, the probability of external attacks d compared to the success probability h of the security measure is so high that it is not optimal for the network manager to try to eliminate the existing insecurities, and therefore, she focuses on preventing the existing safe agents from getting attacked.

8 Conclusion

We studied a dynamic mechanism design problem for a network of interdependent strategic agents with coupled dynamics. In contrast to the existing negative results for static settings, we presented a dynamic mechanism that maximizes the social-welfare, and ensures ex-ante individual rationality and ex-ante budget balance. The mechanism uses the inter-temporal correlation among agents' states to determine at each time, a set of inference signals for all agents that are independent of their own reports. Utilizing this set of inference signals we internalized each agent's effect on the overall network dynamic status, and thus, aligned each agent's objective with the social welfare. Moreover, by using a collection of past inference signals instead of just the most recent ones, we presented reputation-based payments that achieve approximate ex-post individual rationality and budget balance.

The results we presented in this paper are under the assumptions that: (i) the agents have no financial limitation and can pay any participation fee upfront as long as their expected discounted utility is positive, (ii) the agents have transferable and quasi-linear utility, and thus, can be incentivized using monetary payments, and (iii) the network manager's objective is to maximize the social welfare. For future work, it would be interesting to investigate the generalizations of our results by relaxing the above-mentioned assumptions

namely as follows: (1) study the problem of dynamic incentive mechanism design when the agents have limited liability (*i.e.* an agent's discounted utility up to any time t cannot be more negative than an exogenously given finite value), (2) investigate the problem of dynamic non-monetary incentive mechanism design where it is not possible to use monetary payments as incentives to the agents, and (3) study the problem of dynamic incentive mechanism design when the network manager has an arbitrary objective that is different from social welfare.

References

1. Acemoglu, D., Ozdaglar, A.: Opinion dynamics and learning in social networks. Dynamic Games and Applications **1**(1), 3–49 (2011)
2. Allen, L.J., Brauer, F., Van den Driessche, P., Wu, J.: Mathematical epidemiology (2008)
3. Arrow, K.J.: The property rights doctrine and demand revelation under incomplete information. In: M.J. BOSKIN (ed.) Economics and Human Welfare, pp. 23 – 39. Academic Press (1979)
4. Athey, S., Segal, I.: An efficient dynamic mechanism. Econometrica **81**(6) (2013)
5. Bergemann, D., Välimäki, J.: The dynamic pivot mechanism. Econometrica **78**(2), 771–789 (2010)
6. Börgers, T., Krahmer, D., Strausz, R.: An Introduction to the Theory of Mechanism Design. Oxford University Press (2015)
7. Chen, L., Leneutre, J.: A game theoretical framework on intrusion detection in heterogeneous networks. IEEE Trans. on Inf. Forensics and Security (2009)
8. Clarke, E.: Multipart pricing of public goods. Public Choice **2**, 19–33 (1971)
9. Cremer, J., McLean, R.P.: Full extraction of the surplus in bayesian and dominant strategy auctions. Econometrica (1988)
10. d'Aspremont, C., Gérard-Varet, L.A.: Incentives and incomplete information. Journal of Public Economics **11**(1), 25 – 45 (1979)
11. Drakopoulos, K., Ozdaglar, A., Tsitsiklis, J.: When is a network epidemic hard to eliminate? Mathematics of Operations Research **42** (2015)
12. Drakopoulos, K., Ozdaglar, A., Tsitsiklis, J.N.: An efficient curing policy for epidemics on graphs. IEEE Transactions on Network Science and Engineering **1**(2), 67–75 (2014)
13. Farhadi, F., Golestani, S.J.: Mechanism design for network resource allocation: A surrogate optimization approach. available at `http://ee.sharif.edu/~farhadi/Surrogate-Optimization-Approach.pdf` (2016)
14. Farhadi, F., Tavafoghi, H., Teneketzis, D., Golestani, J.: A dynamic incentive mechanism for security in networks of interdependent agents. In: Game Theory for Networks: 7th International EAI Conference, GameNets 2017 Proceedings, pp. 86–96. Springer International Publishing (2017)
15. Fazeli, A., Ajorlou, A., Jadbabaie, A.: Competitive diffusion in social networks: Quality or seeding? IEEE Transactions on Control of Network Systems **PP**(99), 1–1 (2016). DOI 10.1109/TCNS.2016.2553364
16. Friedman, J.W.: A non-cooperative equilibrium for supergames. The Review of Economic Studies **38**(1), 1–12 (1971)
17. Ganesh, A., Massoulie, L., Towsley, D.: The effect of network topology on the spread of epidemics. In: Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies., vol. 2 (2005)
18. Grossklags, J., Christin, N., Chuang, J.: Secure or insure?: A game-theoretic analysis of information security games. In: Proceedings of the 17th International Conference on World Wide Web, WWW '08, pp. 209–218. ACM (2008)
19. Groves, T.: Incentives in teams. Econometrica **41**, 617–631 (1973)
20. Jackson, M., Zenou, Y.: Games on networks. In: Handbook of Game Theory with Economic Applications, vol. 4, chap. 3. Elsevier (2015)

21. Jiang, L., Anantharam, V., Walrand, J.: How bad are selfish investments in network security? IEEE/ACM Transactions on Networking **19**(2) (2011)
22. Kakhbod, A., Teneketzis, D.: An efficient game form for unicast service provisioning. IEEE Trans. on Automatic Control **57**(2) (2012)
23. Karmarkar, N.: A new polynomial-time algorithm for linear programming. Combinatorica **4**, 373–395 (1984)
24. Katz, M.L., Shapiro, C.: Technology adoption in the presence of network externalities. Journal of Political Economy **94**(4), 822–841 (1986)
25. Kelly, F., Maulloo, A., Tan, D.: Rate control for communication networks: shadow prices, proportional fairness and stability. Journal of the Oper. Res. society (1998)
26. Khalili, M.M., Naghizadeh, P., Liu, M.: Designing cyber insurance policies in the presence of security interdependence. In: The 12th Workshop on the Economics of Networks, Systems and Computation (NetEcon 2017) (2017)
27. Khouzani, M.H.R., Sarkar, S., Altman, E.: A dynamic game solution to malware attack. In: IEEE INFOCOM (2011)
28. Kumar, P.R., Varaiya, P.: Stochastic systems: Estimation, identification, and adaptive control, vol. 75. SIAM (2015)
29. Kunreuther, H., Heal, G.: Interdependent security. Journal of Risk and Uncertainty **26**(2), 231–249 (2003)
30. Laszka, A., Felegyhazi, M., Buttyan, L.: A survey of interdependent information security games. ACM Comput. Surv. **47**(2), 1–38 (2014)
31. Lelarge, M.: Economics of malware: Epidemic risks model, network externalities and incentives. In: 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton), pp. 1353–1360 (2009)
32. Lelarge, M., Bolot, J.: Economic incentives to increase security in the internet: The case for insurance. In: IEEE INFOCOM 2009, pp. 1494–1502 (2009)
33. Li, M., Koutsopoulos, I., Poovendran, R.: Optimal jamming attacks and network defense policies in wireless sensor networks. In: IEEE INFOCOM (2007)
34. Liang, X., Xiao, Y.: Game theory for network security. IEEE Communications Surveys Tutorials **15**(1) (2013)
35. Liu, H.: Efficient dynamic mechanisms in environments with interdependent valuations. Available at SSRN 2504731 (2014)
36. Manshaei, M.H., Zhu, Q., Alpcan, T., Basar, T., Hubaux, J.P.: Game theory meets network security and privacy. ACM Comput. Surv. **45**(3), 1–39 (2013)
37. Myerson, R.B.: Multistage games with communication. Econometrica: Journal of the Econometric Society (1986)
38. Naghizadeh, P., Liu, M.: On the role of public and private assessments in security information sharing agreements. arXiv preprint arXiv:1604.04871 (2016)
39. Naghizadeh, P., Liu, M.: Opting out of incentive mechanisms: A study of security as a non-excludable public good. IEEE Trans. on Inf. Forensics and Security (2016)
40. Nowzari, C., Preciado, V.M., Pappas, G.J.: Analysis and control of epidemics: A survey of spreading processes on complex networks. arXiv:1505.00768 (2015)
41. Ogut, H., Menon, N., Raghunathan, S.: Cyber insurance and it security investment: Impact of interdependence risk. In: WEIS (2005)
42. Theodorakopoulos, G., Boudec, J.Y.L., Baras, J.S.: Selfish response to epidemic propagation. IEEE Transactions on Automatic Control **58**(2), 363–376 (2013)
43. Trajanovski, S., Hayel, Y., Altman, E., Wang, H., Mieghem, P.V.: Decentralized protection strategies against sis epidemics in networks. IEEE Transactions on Control of Network Systems (2015)
44. Vickrey, W.: Counterspeculation, auctions and competitive sealed tenders. Journal of Finance pp. 8–37 (1961)

A Proof of Lemma 1:

Consider an arbitrary agent $i \in \mathcal{N}$ and time $t \in \mathcal{T}$. The discounted value of the expected incentive agent i gets at time $t+1$ is

$$\begin{aligned}&\delta\ \mathbb{E}\{-p^i_{t+1}(m^i_t, \boldsymbol{\theta}_t^{-i}, a_t)\} = \\ &-\delta\ [\mathcal{P}(m^i_t = 0|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t) p^i_{t+1}(0, \boldsymbol{r}_t^{-i}, a_t) + \mathcal{P}(m^i_t = 1|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t) p^i_{t+1}(1, \boldsymbol{r}_t^{-i}, a_t)]. \quad (24)\end{aligned}$$

Substituting the tax incentives (15)-(16) in (24) we have

$$\begin{aligned}\delta\ \mathbb{E}\{-p^i_{t+1}(m^i_t, \boldsymbol{\theta}_t^{-i}, a_t)\} &= \mathcal{P}(m^i_t = 0|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t)\Big[\sum_{j\neq i} v^j(0, \boldsymbol{\theta}_t^{-i}, a_t) - \\ &\frac{\mathcal{P}(m^i_t = 1|0, \boldsymbol{\theta}_t^{-i}, a_t)}{\mathcal{P}(m^i_t = 0|0, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)}\ \alpha \sum_{\substack{j\in O^i_:\\ r^j_t=1 \text{ or } a_t=j}} \frac{l_{ij}}{\sum_{k\in\mathcal{N}^j} l_{kj}}\Big] + \\ &\mathcal{P}(m^i_t = 1|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t)\Big[\sum_{j\neq i} v^j(1, \boldsymbol{\theta}_t^{-i}, a_t) + \\ &\frac{\mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)}{\mathcal{P}(m^i_t = 0|0, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)}\ \alpha \sum_{\substack{j\in O^i_:\\ r^j_t=1 \text{ or } a_t=j}} \frac{l_{ij}}{\sum_{k\in\mathcal{N}^j} l_{kj}}\Big]. \quad (25)\end{aligned}$$

Using (3)-(4), we can show that

$$\alpha \sum_{\substack{j\in O^i_:\\ r^j_t=1 \text{ or } a_t=j}} \frac{l_{ij}}{\sum_{k\in\mathcal{N}^j} l_{kj}} = \sum_{j\neq i} v^j(1, \boldsymbol{\theta}_t^{-i}, a_t) - \sum_{j\neq i} v^j(0, \boldsymbol{\theta}_t^{-i}, a_t). \quad (26)$$

Substituting (26) in (25) and rearranging, we obtain

$$\begin{aligned}&\delta\ \mathbb{E}\{-p^i_{t+1}(m^i_t, \boldsymbol{\theta}_t^{-i}, a_t)\} = \\ &\frac{\mathcal{P}(m^i_t = 0|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)}{\mathcal{P}(m^i_t = 0|0, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)} \sum_{j\neq i} v^j(0, \boldsymbol{\theta}_t^{-i}, a_t) + \\ &\frac{\mathcal{P}(m^i_t = 0|0, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t)}{\mathcal{P}(m^i_t = 0|0, \boldsymbol{\theta}_t^{-i}, a_t) - \mathcal{P}(m^i_t = 0|1, \boldsymbol{\theta}_t^{-i}, a_t)} \sum_{j\neq i} v^j(1, \boldsymbol{\theta}_t^{-i}, a_t) = \\ &\mathbf{1}_{\{\theta^i_t=0\}} \sum_{j\neq i} v^j(0, \boldsymbol{\theta}_t^{-i}, a_t) + \mathbf{1}_{\{\theta^i_t=1\}} \sum_{j\neq i} v^j(1, \boldsymbol{\theta}_t^{-i}, a_t) = \sum_{j\neq i} v^j(\theta^i_t, \boldsymbol{\theta}_t^{-i}, a_t), \quad (27)\end{aligned}$$

which completes the proof. ■

B Proof of Proposition 1:

We want to show that the DCI mechanism is incentive compatible. To prove this, we can disregard participation fees (17) taken from the agents at the beginning of the mechanism, since these participation fees are independent of the agents' future reports, and hence do not affect the agents' strategies when they join the mechanism. To prove that the truth-telling is the best strategy that each agent i can choose from each time t onward, using the one shot deviation principle [6], we only need to show that for each i, t and $\boldsymbol{\theta}_t$, telling the truth

by agent i at time t maximizes his expected continuation payoff, i.e. $r_t^i = \theta_t^i$ is a solution to the following optimization problem

$$\max_{r_t^i} \Big[v^i(\boldsymbol{\theta}_t, a_t^*(r_t^i, \boldsymbol{\theta}_t^{-i})) - p_t^i(m_{t-1}^i, \boldsymbol{r}_{t-1}^{-i}, a_{t-1}) + \delta \mathbb{E}\{ v^i(\boldsymbol{\theta}_{t+1}, a_{t+1}^*(\boldsymbol{\theta}_{t+1})) - p_{t+1}^i(m_t^i, \boldsymbol{\theta}_t^{-i}, a_t^*(r_t^i, \boldsymbol{\theta}_t^{-i})) \} + \mathbb{E}\{ \sum_{\tau=t+2}^{\infty} \delta^{\tau-t} (v^i(\boldsymbol{\theta}_\tau, a_\tau^*(\boldsymbol{\theta}_\tau)) - p_\tau^i(m_{\tau-1}^i, \boldsymbol{\theta}_{\tau-1}^{-i}, a_{\tau-1}^*(\boldsymbol{\theta}_{\tau-1}))) \} \Big]. \quad (28)$$

Using Lemma 1 we can show that the problem (28) is equivalent to the following optimization problem:

$$\max_{r_t^i} \Big[\sum_{j=1}^{n} v^j(\boldsymbol{\theta}_t, a_t^*(r_t^i, \boldsymbol{\theta}_t^{-i})) + \mathbb{E}\{ \sum_{\tau=t+1}^{\infty} \delta^{\tau-t} \sum_{j=1}^{n} v^j(\boldsymbol{\theta}_\tau, a_\tau^*(\boldsymbol{\theta}_\tau)) \} - p_t^i(m_{t-1}^i, \boldsymbol{r}_{t-1}^{-i}, a_{t-1}) \Big]. \quad (29)$$

Since the third term in the objective function of (29), i.e. $p_t^i(m_{t-1}^i, \boldsymbol{r}_{t-1}^{-i}, a_{t-1})$, is independent of r_t^i, the optimal report of agent i is also a solution to the optimization problem below which maximizes the social surplus from time t onward, i.e.

$$\max_{r_t^i} \Big[\sum_{j=1}^{n} v^j(\boldsymbol{\theta}_t, a_t^*(r_t^i, \boldsymbol{\theta}_t^{-i})) + \mathbb{E}\{ \sum_{\tau=t+1}^{\infty} \delta^{\tau-t} \sum_{j=1}^{n} v^j(\boldsymbol{\theta}_\tau, a_\tau^*(\boldsymbol{\theta}_\tau)) \} \Big]. \quad (30)$$

Therefore, agent i's objective is aligned with the social welfare. The allocation policy a^* used by the network manager is the policy that maximizes the social welfare under the centralized information; therefore, the best strategy for agent i, whose objective is shown to be aligned with the social welfare, is to trust the network manager and to provide her with truthful information. Consequently, $r_t^i = \theta_t^i$ is a solution to problem (30) and this establishes the incentive compatibility of the DCI mechanism. ■

C Proof of Proposition 4:

We want to show that the DCI mechanism is individually rational, i.e. each agent prefers the outcome of the mechanism to that attained by opting out. According to (17), the participation fee each agent pays to join the mechanism is equal to the the expected value of the subsidies he gets at truth-telling equilibrium; therefore,

$$\mathbb{E}\{(1-\delta) \sum_{\tau=0}^{\infty} \delta^\tau p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)\} = 0, \forall i \in \mathcal{N}. \quad (31)$$

However, each agent's valuation when he participates in the mechanism is greater than or equal to the utility he gets when he opts out. This is because when agent i does not participate in the mechanism the network manager allocates no security measure to any agent which makes the security situation worse for all agents. Therefore,

$$\mathbb{E}\{(1-\delta) \sum_{\tau=0}^{\infty} \delta^\tau \Big[v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)) - p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau) \Big] \} =$$
$$\mathbb{E}\{(1-\delta) \sum_{\tau=0}^{\infty} \delta^\tau v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau))\} \geq \mathbb{E}\{U_0^i(a)\}, \forall i \in \mathcal{N}. \quad (32)$$

This proves individual rationality of the DCI mechanism. ■

D Proof of Theorem 1:

In Propositions 1-4 we prove that the DCI mechanism maximizes the social welfare function W and is incentive-compatible, budget balanced, and individually rational. These results proves that the DCI mechanism solves the network manager's dynamic incentive design problem (13). ■

E Proof of Lemma 2:

Let $V^i = (1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau v^i(\boldsymbol{\theta}_\tau, a_\tau(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau))$ and $P^i = (1-\delta)\sum_{\tau=0}^{\infty}\delta^\tau p_\tau^i(\boldsymbol{\theta}_\tau^{-i}, \theta_\tau^i, h_\tau)$ denote agent i's total valuation and total payment, respectively. Using these notations, ex-post requirements (19)-(20) can be rewritten as

$$\mathbb{P}\{V^i - P^i \leq U_0^i(a) - \eta\} \leq \zeta, \forall i \in \mathcal{N}, \tag{33}$$

and

$$\mathbb{P}\{|\sum_{i\in\mathcal{N}} P^i| \geq \eta\} \leq \zeta. \tag{34}$$

In DCI mechanism and for any realization of the stochastic events, the valuation of the agent when he participates in the mechanism is at least equal to his utility when opts out, i.e. $V^i \geq U_0^i(a)$. This is because, if agent i unilaterally opts out, the network manager stops allocating security measures to the network agents, therefore, the epidemics spread uncontrollably and all of the agents lose. Therefore, a condition sufficient to satisfy (33) is

$$\mathbb{P}\{|P^i| \geq \eta\} \leq \zeta, \forall i \in \mathcal{N}. \tag{35}$$

Suppose that the conditions of Lemma 2 are satisfied. Due to condition (1), each P^i, $i \in \mathcal{N}$, and hence $\sum_{i\in\mathcal{N}} P^i$ are zero-mean random variables. Therefore, using the Chebyshev inequality, we have

$$\mathbb{P}\{|P^i| \geq \eta\} \leq \frac{Var(P^i)}{\eta^2}, \forall i \in \mathcal{N}. \tag{36}$$

and

$$\mathbb{P}\{|\sum_{i\in\mathcal{N}} P^i| \geq \eta\} \leq \frac{Var(\sum_{i\in\mathcal{N}} P^i)}{\eta^2} \leq \frac{(\sum_{i\in\mathcal{N}}\sqrt{Var(P^i)})^2}{\eta^2}, \tag{37}$$

where the last inequality in (37) follows from Cauchy-Schwarz inequality.

According to condition (2), the mechanism can make the variance of the agents' payments smaller than any bound $\epsilon > 0$. Let $\epsilon = \zeta\eta^2/n^2$. Substituting $Var(P^i) \leq \zeta\eta^2/n^2$, $i \in \mathcal{N}$, in (36)-(37), we conclude inequalities (35) and (34), respectively, and establish that the mechanism is approximate ex-post IR and BB.

■

F Proof of Theorem 2:

We divide the proof into two cases:

Case 1: $h < 1$. In this case we show that Markov chain M^* derived from the evolution of the system under the optimal allocation policy is *irreducible* and *aperiodic*, hence is *ergodic*. We prove this via several steps.

Step 1. We show that in the Markov chain M^* every state is *accessible* from the clean state $\boldsymbol{\theta} = \mathbf{1}$ where all agents are safe. A state $\hat{\boldsymbol{\theta}}$ is said to be accessible from a state $\boldsymbol{\theta}$ if a system started in state $\boldsymbol{\theta}$ has a non-zero probability of transitioning into state $\hat{\boldsymbol{\theta}}$ in one or more moves. If the system is in the clean state, since all the probabilities d_i, $i \in \mathcal{N}$, l_{ij}, $i \in \mathcal{N}$, $j \in \mathcal{N}^i$, and h are strictly between 0 and 1, for each agent i both the events that i gets attacked at the next time or remains safe have non-zero probabilities. Therefore, any state $\hat{\boldsymbol{\theta}}$ is accessible from $\boldsymbol{\theta} = \mathbf{1}$ in just one move.

Step 2. We show that the clean state $\boldsymbol{\theta} = \mathbf{1}$ is accessible from any state $\hat{\boldsymbol{\theta}} \in \Theta^n$. We prove this by induction on the number of unsafe agents in the original/starting state $\hat{\boldsymbol{\theta}}$, i.e. $K(\hat{\boldsymbol{\theta}}) = n - \sum_{i \in \mathcal{N}} \hat{\theta}^i$.

Base Case. Let $K(\hat{\boldsymbol{\theta}}) = 0$. The only state in the network that has no unsafe agent is the clean state; i.e. $\hat{\boldsymbol{\theta}} = \mathbf{1}$. Starting from the clean state the claim directly follows from Step 1.

Induction step. Suppose that the clean state is accessible from any state with at most k unsafe agents. Now, let $K(\hat{\boldsymbol{\theta}}) = k + 1$. In a state $\hat{\boldsymbol{\theta}}$ with $k + 1$ unsafe agents, by the assumption made in the statement of the theorem the network manager applies the security measure to one of the unsafe agents, say agent i. Since the success probability h of the security measure is non-zero, and the probabilities of new external or internal attacks to agents are strictly less than one, there is a positive probability that the system transitions to a state $\tilde{\boldsymbol{\theta}}$ where agent i is safe and no new agent gets attacked. In state $\tilde{\boldsymbol{\theta}}$ the number of unsafe agents is k; therefore, by the induction hypothesis, the clean state is accessible from $\tilde{\boldsymbol{\theta}}$. Since the accessibility relation is transitive, the clean state $\boldsymbol{\theta} = \mathbf{1}$ is accessible from the starting state $\hat{\boldsymbol{\theta}}$.

Conclusion. By the principle of induction, we have proved that the clean state is accessible from every state.

Step 3. We show that the Markov chain M^* is irreducible. Since the accessibility relation is transitive, it follows from Steps 1 and 2 that in M^* every state is accessible from every other state. Therefore, the Markov chain M^* is irreducible.

Step 4. We show that the Markov chain M^* is aperiodic. For irreducible Markov chains, either all states are periodic or all are aperiodic. Therefore, since in Step 3 we proved that the Markov chain M^* is irreducible, it is enough to show that one of its states is aperiodic. A sufficient condition for a state to be aperiodic is that it has a self-loop, i.e. starting from that state the probability that the next state is the same as the current state is non-zero. According to Step 1, in the Markov chain M^* the clean state has a self-loop. therefore, it is an aperiodic state, and this proves the aperiodicity of the Markov chain M^*.

It follows from Steps 3 and 4 that the Markov chain M^* is both irreducible and aperiodic; furthermore, M^* is finite-state, every state in it is positive recurrent. Therefore, M^* is ergodic.

Case 2: $h = 1$. In this case we show that the Markov chain M^* consists of one class of ergodic states and possibly one transient state, hence it is an ergodic uni-chain. We prove this in two steps.

Step I. We show that in the Markov chain M^* every state $\hat{\boldsymbol{\theta}} \neq \mathbf{0}$ is *accessible* from the clean state $\boldsymbol{\theta} = \mathbf{1}$. Let $i = a^*(\mathbf{1})$ denote the agent who receives the security measure when the network is in the clean state. Since the probability of success when applying the security measure is $h = 1$, starting from the clean state, agent i will keep his safety until the next time slot. However, since the probabilities of external and internal attacks are strictly between 0 and 1, other agents except i could be either safe or unsafe in the next time. Therefore, all states with $\hat{\theta}^i = 1$ are accessible from the clean state in just one move. Now consider a state $\hat{\boldsymbol{\theta}} \neq \mathbf{0}$ with $\hat{\theta}^i = 0$. Since $\hat{\boldsymbol{\theta}} \neq \mathbf{0}$ there is at least one agent $j \neq i$ such that $\hat{\theta}^j = 1$. We show that $\hat{\boldsymbol{\theta}}$ is accessible from the clean state in two moves. Consider a state $\bar{\boldsymbol{\theta}}$ where $\bar{\theta}^j = 0$, and $\bar{\theta}^k = 1$, for all $k \neq j$, that is agent j is unsafe and all other agents are safe. As a result of the argument of step 1, $\bar{\boldsymbol{\theta}}$ is accessible from the clean state $\boldsymbol{\theta} = \mathbf{1}$ in one move. By the assumption made in the statement of theorem, in state $\bar{\boldsymbol{\theta}}$ the optimal policy

applies the security measure to agent j, i.e. $a^*(\bar{\boldsymbol{\theta}}) = j$. As a result, agent j switches to the safe state at the next time; all other agents may or may not get attacked, and each event occurs with a positive probability. Thus, all states where agent j is safe, which includes $\hat{\boldsymbol{\theta}}$, are accessible from $\bar{\boldsymbol{\theta}}$ in one move. Since $\bar{\boldsymbol{\theta}}$ is itself accessible from the clean state in one move, we can conclude that $\hat{\boldsymbol{\theta}}$ is accessible from the clean state in two moves. Therefore, every state $\hat{\boldsymbol{\theta}} \neq \mathbf{0}$ is accessible from the clean state in one or two moves.

Step II. We show that the Markov chain M^* is an ergodic uni-chain. Using the same proof as in Step 2 of Case 1, we can show that the claim of Step 2 is still true when $h = 1$, i.e. the clean state $\boldsymbol{\theta} = \mathbf{1}$ is accessible from every state. Therefore, due to the transitivity of the accessibility relation, all states except $\mathbf{0}$ are accessible from each other. If state $\mathbf{0}$ is also accessible from the clean state, using the same proof as in Steps 3 and 4 of Case 1, we can show that the Markov chain is ergodic. However, if state $\mathbf{0}$ is not accessible from the clean state, it is a transient state. For finite-state Markov chains, a state is transient if it is not accessible from at least one of the states to which it has access. In this case, following the proofs of Steps 3 and 4 of Case 1, we can show that the set of states except $\mathbf{0}$ is an ergodic class. Therefore, the Markov chain is an ergodic uni-chain. ■

G Proof of Theorem 3:

Using the Cauchy-Schwarz inequality we have,

$$Var(\hat{p}^i(\delta, K)) = (1-\delta)^2 Var(\sum_{t=0}^{\infty} \delta^t \hat{p}_t^i(\delta, K)) \leq (1-\delta)^2 \, (\sum_{t=0}^{\infty} \delta^t \sqrt{Var(\hat{p}_t^i(\delta, K))}\,)^2. \quad (38)$$

Equation (38) gives an upper bound for the variance of agent i's total payment based on the variances of his payments $\hat{p}_t^i(\delta, K)$ at each time t. Using (22), we can derive time-t variance of agent i's payment as

$$Var(\hat{p}_t^i(\delta, K)) = \frac{1}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \frac{1}{\delta^{2k}} \, \sigma_i^2 + \frac{2}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \sum_{k'=0}^{k-1} \frac{\rho_{k-k'}^i}{\delta^{k+k'}} \, \sigma_i^2, \quad (39)$$

where $\sigma_i^2 = Var(p_t^i)$, $t \in \mathcal{T}$, denotes the variance of agent i's tax at each time t, and $\rho_k^i = Corr(p_t^i, p_{t+k}^i)$, $t \in \mathcal{T}$, $k > 0$, is the correlation coefficient between agent i's tax at any time t and $t + k$. Note that since the allocation policy and payment function of the DCI mechanism are time-invariant, for each $i \in \mathcal{N}$, the tax time-series $\{p_t^i, t \in \mathcal{T}\}$ is a *stationary* random process. Therefore, parameters such as, the variance σ_i^2 and the correlation coefficients ρ_k^i, $k > 0$, are independent of t.

We want to show that when the discount factor δ is large enough, the network manager can choose an appropriate K so as to make $Var(\hat{p}_t^i(\delta, K))$ and hence $Var(\hat{p}^i(\delta, K))$, sufficiently small. First we consider the extreme case where $\delta = 1$. In this case, we show that the manager can achieve his goal by choosing the agreement period K large enough, since $Var(\hat{p}^i(\delta, K))$ goes to zero when K goes to infinity. Setting $\delta = 1$ in (39), we obtain

$$Var(\hat{p}_t^i(1, K)) = \frac{\min(K, t)}{K^2} \, \sigma_i^2 + \frac{2}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \sum_{k'=0}^{k-1} \rho_{k-k'}^i \, \sigma_i^2. \quad (40)$$

When K goes to infinity, the first term of (40) goes to zero. We upper bound the second term by taking the absolute value and using the triangle inequality as follows,

$$\begin{aligned} \frac{2}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \sum_{k'=0}^{k-1} \rho_{k-k'}^i \, \sigma_i^2 \;&\leq\; \left| \frac{2\sigma_i^2}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \sum_{k'=0}^{k-1} \rho_{k-k'}^i \right| \\ &\leq \frac{2\sigma_i^2}{K^2} \sum_{k=0}^{\min(K-1, t-1)} \sum_{k'=0}^{k-1} \left| \rho_{k-k'}^i \right| \;\leq\; \frac{2\sigma_i^2}{K^2} \sum_{k=0}^{K-1} \sum_{k'=0}^{k-1} \left| \rho_{k-k'}^i \right| \end{aligned}$$

$$= \frac{2\sigma_i^2}{K^2} \sum_{l=1}^{K-1} (K-l) \left|\rho_l^i\right|, \tag{41}$$

where the last inequality holds because all the terms in the summation are non-negative, hence raising the upper bound of summation increases the total value. The last equality follows from the change of variable $l := k - k'$. We show that the upper bound derived in (41) approaches zero when K goes to infinity. To prove this we use the following lemma.

Lemma 3 For the fixed allocation policy π^*, we have

$$\lim_{l \to \infty} \rho_l^i = 0, \qquad \forall i \in \mathcal{N}. \tag{42}$$

Proof By definition, we have $\rho_l^i = Corr(p_t^i, p_{t+l}^i)$, for all $t \in \mathcal{T}$. Having fixed the allocation policy, p_t^i is a function of the network states at time t and $t-1$, i.e.

$$p_t^i = f(\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_t). \tag{43}$$

Therefore, considering $t = 1$, we have

$$\rho_l^i = Corr(f(\boldsymbol{\theta}_0, \boldsymbol{\theta}_1), f(\boldsymbol{\theta}_l, \boldsymbol{\theta}_{l+1})). \tag{44}$$

According to Assumption 1, the allocation policy π^* leads to an ergodic uni-chain. Therefore, independently of the initial states $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$, when l goes to infinity, the distribution of the state of the network approaches the limiting distribution of the Markov chain. Consequently, when l goes to infinity, $\boldsymbol{\theta}_l$ and $\boldsymbol{\theta}_{l+1}$ become independent of $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_1$. Functions of independent random variables are also independent and hence are uncorrelated; therefore,

$$\lim_{l \to \infty} \rho_l^i = 0. \tag{45}$$

■

Using the definition of limit, Lemma 3 says that for any $\epsilon > 0$, there is an $l_0(\epsilon)$ such that $l > l_0(\epsilon)$ implies $\left|\rho_l^i\right| < \epsilon$. Since we want to find the limit of the upper bound derived in (41) when K approaches infinity, for each $\epsilon > 0$ we can assume that $K > l_0(\epsilon) + 1$. Then, we have

$$\frac{2\sigma_i^2}{K^2} \sum_{l=1}^{K-1} (K-l) \left|\rho_l^i\right| = \frac{2\sigma_i^2}{K^2} \left[\sum_{l=1}^{l_0(\epsilon)} (K-l) \left|\rho_l^i\right| + \sum_{l=l_0(\epsilon)+1}^{K-1} (K-l) \left|\rho_l^i\right| \right]. \tag{46}$$

According to the definition of $l_0(\epsilon)$, $\left|\rho_l^i\right| < \epsilon$, for all $l > l_0(\epsilon)$. Moreover, due to the definition of correlation coefficient, $\left|\rho_l^i\right| \leq 1$ for all $l \leq l_0(\epsilon)$. Using these bounds in (46) and computing the series, we obtain

$$\frac{2\sigma_i^2}{K^2} \sum_{l=1}^{K-1} (K-l) \left|\rho_l^i\right| \leq \sigma_i^2 \left[\frac{l_0(\epsilon)(2K - l_0(\epsilon) - 1)}{K^2} + \frac{\epsilon(K - l_0(\epsilon))(K - l_0(\epsilon) - 2)}{K^2} \right]. \tag{47}$$

The right-hand side of (47) goes to $\epsilon\,\sigma_i^2$ when K goes to infinity. Since ϵ could be any arbitrary positive number, it implies that

$$\lim_{K \to \infty} \frac{2\sigma_i^2}{K^2} \sum_{l=1}^{K-1} (K-l) \left|\rho_l^i\right| = 0. \tag{48}$$

Since a variance is always non-negative, from (40), (41), (47), and (48) it follows that

$$\lim_{K \to \infty} Var(\hat{p}_t^i(1, K)) = 0. \tag{49}$$

Using (49), we conclude from (38) that

$$\lim_{K\to\infty} Var(\hat{p}^i(1,K)) = 0. \tag{50}$$

Summing (50) over all $i \in \mathcal{N}$, we get

$$\lim_{K\to\infty} \sum_{i\in\mathcal{N}} Var(\hat{p}^i(1,K)) = 0, \tag{51}$$

which, due to the definition of limit, is equivalent to

$$\forall \epsilon_1 > 0,\ \exists K_0(\epsilon_1) > 0, \quad s.t. \quad K \geq K_0(\epsilon_1) \ \Rightarrow\ \sum_{i\in\mathcal{N}} Var(\hat{p}^i(1,K)) < \epsilon_1. \tag{52}$$

Therefore, when δ equals one, for each $\epsilon_1 > 0$, the network manager can find appropriate Ks so as to make the total variance of all agents' payments smaller than ϵ_1. Such Ks obviously make the individual variance of each agent's payment less than ϵ_1.

Now, we consider the case where $\delta < 1$. In this case, the first term on the right hand side of (39) is a geometric series with common ratio $1/\delta^2 > 1$, hence it diverges as K goes to infinity. This is expected because when the value of money is discounted over time agents must pay larger installments in the future to compensate a tax in the past. Thus, if the agreement period is too long, the installments of the agents become large and the variance of the payments tends to infinity. Therefore, in this case in order to make the risk of agents arbitrarily small, the network manager must choose an appropriate agreement period K, far away from zero and infinity, so as to strike a balance between the positive effects of reputation and the negative effects of large installments. This goal is achievable due to the continuity of the function $\sum_{i\in\mathcal{N}} Var(\hat{p}^i(\delta,K))$ with respect to δ.

For each $K > 0$, $\sum_{i\in\mathcal{N}} Var(\hat{p}^i(\delta,K))$ is a continuous function of δ. Therefore, by the definition of continuity, we have

$$\forall \epsilon_2 > 0,\ \exists \delta_0(K,\epsilon_2) < 1,\ s.t.\ \delta > \delta_0(K,\epsilon_2) \Rightarrow \left| \sum_{i\in\mathcal{N}} Var(\hat{p}^i(\delta,K)) - \sum_{i\in\mathcal{N}} Var(\hat{p}^i(1,K)) \right| < \epsilon_2. \tag{53}$$

For each $\epsilon > 0$, consider $\epsilon_1 = \epsilon_2 = \epsilon/2$. Then, from (52) we get

$$\sum_{i\in\mathcal{N}} Var(\hat{p}^i(1,K_0(\epsilon/2))) < \epsilon/2. \tag{54}$$

Considering $K = K_0(\epsilon/2)$, it follows from (53) that

$$\delta > \delta_0(K_0(\epsilon/2),\epsilon/2) \ \Rightarrow\ \sum_{i\in\mathcal{N}} Var(\hat{p}^i(\delta,K_0(\epsilon/2))) < \epsilon \ \Rightarrow\ Var(\hat{p}^i(\delta,K_0(\epsilon/2))) < \epsilon, \forall i \in \mathcal{N}, \tag{55}$$

where the last step follows because of the non-negativity of variance. Therefore, for each $\epsilon > 0$, there exists a threshold $\delta_0(K_0(\epsilon/2),\epsilon/2)$, such that for all $\delta > \delta_0(K_0(\epsilon/2),\epsilon/2)$, the network manager can make the variance of each agent's payment smaller than ϵ by choosing $K = K_0(\epsilon/2)$. This completes the proof of Theorem 3. ■

H Proof of Theorem 5:

Due to symmetry, we can capture the state of the system by a new state variable $s = \sum_{i=1}^{n} \theta^i$ which only counts the number of safe agents. We can also define a new action variable $b \in \{0,1\}$, where $b = 0$ ($b = 1$) represents the act of healing an unsafe (safe) agent. It is easy to show that the optimal value function $V^*(s)$ is increasing in s, meaning that a state with

a greater number of safe agents is strictly preferred to a state with smaller number of safe agents.

To prove the theorem, we should show that if the condition holds, $\pi^*(s) = 0$, for all $s < n$, which is trivial when $s = 0$, and is equivalent to showing

$$r(s,0) + \delta \sum_{s'} P(s'|s,0)V^*(s') \geq r(s,1) + \delta \sum_{s'} P(s'|s,1)V^*(s'), \tag{56}$$

when $s \in [1, n-1]$. We prove (56) in two steps.

Step 1: We show that $r(s,0) \geq r(s,1)$, for all $s \in [1, n-1]$. Due to (4), we have

$$r(s,0) = s(1 + \alpha\frac{s-1}{n-1}) + \alpha\frac{s}{n-1} \geq s(1 + \alpha\frac{s-1}{n-1}) = r(s,1), \tag{57}$$

where the inequality holds because α is positive.

Step 2: We prove that $P(.|s,0)$ has first order stochastic dominance (FSD) over $P(.|s,1)$, for all $s \in [1, n-1]$. To this end, we show that $P(s' \geq k|s,0) \geq P(s' \geq k|s,1)$, for all k. Let n_0 and n_1 denote one of the unsafe and one of the safe agents, respectively. Therefore, due to symmetry, choosing each action variable $b \in \{0,1\}$ is equivalent to applying security measure to agent n_b. Now we have

$$\begin{aligned} P(s' \geq k|s,b) = P(\sum_i \theta'^i \geq k|s, a = n_b) = \\ P(\sum_{i \neq n_0,n_1} \theta'^i \geq k|s) + P(\sum_{i \neq n_0,n_1} \theta'^i = k-1|s)P(\theta'^{n_0} + \theta'^{n_1} \geq 1|s, a = n_b) + \\ P(\sum_{i \neq n_0,n_1} \theta'^i = k-2|s)P(\theta'^{n_0} + \theta'^{n_1} = 2|s, a = n_b), \end{aligned} \tag{58}$$

for $b = 0, 1$. The terms that depend on b can be computed as follows:

$$P(\theta'^{n_0} + \theta'^{n_1} = 2|s, a = n_1) = 0, \tag{59a}$$

$$P(\theta'^{n_0} + \theta'^{n_1} = 2|s, a = n_0) = h(1-d)(1-d(1-h))(1-l)^{2(n-s)-1}, \tag{59b}$$

$$P(\theta'^{n_0} + \theta'^{n_1} \geq 1|s, a = n_1) = (1-d(1-h))(1-l)^{n-s}, \tag{59c}$$

$$P(\theta'^{n_0} + \theta'^{n_1} \geq 1|s, a = n_0) = 1-(1-h(1-d(1-h))(1-l)^{n-s-1})(1-(1-d)(1-l)^{n-s}). \tag{59d}$$

These probabilities (59a)-(59d) are derived from (2) by setting $\mathcal{N}^i = \mathcal{N} - i$ which is true for completely connected networks. It can be shown by some simple algebra that, if $d(1-h) \leq l$, (59b) and (59d) are greater than or equal to (59a) and (59c), respectively, and these inequalities complete the proof of Step 2.

Since $V^*(s)$ is increasing in s, we conclude from the above steps that (56) holds for all $s \in [1, n-1]$, therefore, the assertion of Theorem 5 is established. ■