Inter-Temporal Incentives in Security Information Sharing Agreements

Parinaz Naghizadeh and Mingyan Liu
Department of Electrical Engineering and Computer Science
University of Michigan, Ann Arbor, MI 48105
Email: {naghizad, mingyyan}@umich.edu

***Abstract*—Sharing of security information among organizations has often been proposed as a method for improving the state of cyber-security for all. However, such disclosure entails costs for the reporting entity, including a drop in market value, loss of customer confidence, and bureaucratic burdens. By adopting a game-theoretic approach for understanding firms' incentives, we first show that in one shot interactions, disclosure costs act as disincentives for sharing, leading to no information sharing at equilibrium. We then consider a repeated game formulation that enables the use of inter-temporal incentives (i.e., the conditioning of future decisions on the history of past interactions) to support firms' cooperation on information sharing. We show that the nature of the monitoring available to firms greatly affects the possibility of sustaining nearly efficient outcomes through repeated interactions. Specifically, we first illustrate the limitations arising when firms' have independent and imperfect private monitoring technologies. We then consider the role of a public monitoring/assessment system, and show that despite using the same imperfect technology as the individual firms, the monitor can facilitate cooperation among participants. Our results therefore illustrate the impact of a public rating/reputation system on the viability of security information sharing agreements.**

I. Introduction

Improving the ability of analyzing cyber-threats and cyber-incidents, and ensuring that the results of the analysis are shared among organizations and authorities, have received increased attention in the recent years by governments and policy makers. This growing attention to security information sharing policies is on one hand due to governments' interest in disseminating the attained information across federal agencies, to better protect the national infrastructure against potential cyber-attacks. On the other hand, the availability of security information benefits non-federal organizations, by allowing them to leverage other companies' experience to prevent similar attacks, and to invest in the most effective preventive and protective measures. Finally, information sharing laws can protect consumer rights; e.g., by requiring the disclosure of breaches involving personally identifiable information.

In general, depending on the breach notification law or the information sharing agreement, a firm may be required to either publicly announce an incident, to report it to other firms participating in the agreement or within its industry sector, to notify affected individuals, and/or to notify the appropriate authorities; see [1] for a summary of prominent US and EU laws. In this paper, motivated by the latest initiatives in the US (in particular, Executive Order 13691 [2]), we are primarily interested in information sharing agreements *among firms*, both with and without facilitation by an authority. Currently, joining and reporting in all such existing agreements is voluntary.

A. Problem motivation

Despite the existence of studies showing that laws (even indirectly) encouraging higher focus on reporting of security-related information (e.g. the Sarbanes-Oxley Act of 2002) can have a positive effect on disclosure of information security by organizations [3], there exist anecdotal and empirical evidence that security breaches remain vastly under-reported. For example, a survey of 300 attendees at the 2007 RSA Conference concluded that 89% of that year's incidents had gone unreported [4]. The same year, Bryan Sartin, VP of investigative response at Verizon, estimated that only 5% of over 500 forensic investigations conducted by Verizon Business Security Solutions had been disclosed [5]. More recently, in a 2013 survey of 200 security professionals in US enterprises, nearly 6 in 10 reported investigating a breach that was never disclosed by their employers [6].

These observed disincentives by companies for sharing security information can be primarily explained by analyzing the associated economic impacts. [7], [8] conduct event-study analyses of market reaction to breach disclosures, both demonstrating a drop in market values following the announcement of a security breach. In addition to an initial drop in stock prices, an exposed breach or security flaw can result in loss of consumer/partner confidence in a company, leading to a further decrease of revenues in the future [9]. Finally, documenting and announcing security breaches impose a bureaucratic burden on the company, as an agreement may require the reports to comply with a certain incident reporting terminology; examples of such frameworks include the recently proposed categorization by DHS [10], and the Vocabulary for Event Recording and Incident Sharing (VERIS) proposed by the Verizon RISK team [11].

Given these potential disclosure costs, and the evidence of under-reporting of security information, it is clear that we need a better understanding of firms' incentives for participating in information sharing organizations, as well as the economic incentives that could lead to voluntary cooperation by firms in these agreements.

B. A game-theoretic approach to information sharing

In this paper, we present a game-theoretic study of information sharing agreements among firms. In addition to the disclosure costs, our model takes into account firms' gains from information sharing agreements, by assuming that each firm can benefit from having access to other firms' security information, as she can prevent similar attacks and invest in the best security measures by leveraging other firms' experience. As a result, an outcome in which firms fully disclose their security information yields higher payoffs to all participants. Nevertheless, we show that in a (one-shot) information sharing game among rational firms, the disclosure costs will act as a deterrent, leading firms' to exhibit free-riding behavior (i.e., attempt to benefit from other firms' information without disclosing their own security status). Consequently, there will be no information sharing at the state of equilibrium, as also predicted by similar game-theoretic models which consider one-shot information sharing games (see Section V). Existing research has further proposed audits and sanctions (e.g. by an authority or the government) or introducing additional economic incentives (e.g. taxes and rewards for members of ISACs) as remedies for encouraging information disclosure.

In this work, we propose a different approach for providing incentives for cooperation in information sharing agreements. We model the information sharing game in a *repeated game* framework, therefore allowing for firms' future disclosure decisions to be dependent on the history of their interactions with other firms in the agreement. It has been well-known in the economic literature that repetitions of an otherwise non-cooperative and inefficient game can lead economically rational agents to coordinate on efficient equilibria [12], [13], [14]. A well-known example of this phenomenon is that of a prisoner's dilemma game: while two rational players should always defect in a one shot (or for finite repetitions) of the game, cooperation can be supported in an infinitely repeated game, as conditioning of future behavior on the history of past actions can prevent players from behaving opportunistically. Similarly, we are interested in understanding whether inter-temporal incentives can be used to sustain information disclosure in sharing agreements.

The possibility of achieving efficient outcomes in a repeated game depends heavily on whether the monitoring of other participants' actions is perfect or imperfect, and private or public. In particular, for information sharing games, each firm can (at best) only imperfectly assess the honesty and comprehensiveness of others' shared information. We therefore first cast information sharing games as repeated prisoner's dilemma game with *imperfect private monitoring*. We start by illustrating the complications arising from the private nature of firms' beliefs about one another, by studying the level of cooperation attainable in a two-stage game. We then discuss the infinitely repeated information sharing game, and briefly discuss the possibility of supporting cooperation when additional measures such as communication, public correlating devices, or public actions, are introduced in the game. Alternatively, we consider the role of a public rating/reputation system in facilitating cooperation on information disclosure. Although the monitoring provided by such system is also in general imperfect, it is available to both firms, leading to a repeated game with *imperfect public monitoring*. We show that eventhough this public assessment system employs the same monitoring technology as the individual firms, the public availability of the assessments can allow the firms to coordinate on cooperative outcomes.

Therefore, the main contributions of this work are twofold. First, to the best of our knowledge, this work is the first to study security information sharing agreements in a repeated game framework. Second, our work identifies the crucial role that the availability of a public, external assessment system plays in sustaining cooperative behavior in information sharing agreements.

The paper is organized as follows. We present the model and preliminaries of information sharing games in Section II, followed by the study of games with private monitoring in Section III. We illustrate the role of public monitoring in facilitating cooperation in Section IV. In Section V, we review the literature most related to the current paper. Section VI concludes the paper.

II. The information sharing game

A. The stage game

Consider two (symmetric) firms who have an information sharing agreement; e.g., due to joining an ISAC. Each firm chooses a level of investment in security measures to protect her network. Examples include implementing an intrusion detection system, introducing employee education initiatives, and installing and maintaining up-to-date software. We assume these investments are determined exogenously; i.e., are not affected by either user's information sharing decision.[1]

As part of the information sharing agreement, firm i is assumed to share her security information with firm j. This information may include a summary of implemented security measures, and a history of both successful and unsuccessful attacks/breaches. Nevertheless, firm i can decide whether to comply with this agreement; i.e., whether to (fully and honestly) disclose this information. We denote the decision of firm i by $r_i \in \{0, 1\}$, denoting (partially) concealing and (fully) disclosing, respectively. A choice of $r_i = 1$ by firm i will benefit firm j by improving her state of security. This is because firm j can leverage i's experience to prevent similar attacks and ongoing threats, and to invest in the most effective protective measures. We denote this *information gain* by $G > 0$.

[1]We make this assumption for two reasons. First, this allows us to focus only on firms' incentives for information sharing. More importantly, here the information shared by firm i is assumed to be a *substitute* to firm j's investment; i.e., firm j can decrease her security expenditure when she receives information from firm i. This possible reduction in the positive externality of j's investments on firm i's security state may therefore result in further disincentives for firm i for sharing security information. We therefore remove this disincentive by assuming fixed security expenditures.

TABLE I
FIRMS' PAYOFFS IN AN INFORMATION SHARING GAME

	1	0
1	$G-C, G-C$	$-C, G$
0	$G, -C$	0, 0

The disincentive of firm i in adopting a choice of $r_i = 1$ is due to its associated *disclosure cost* $C > 0$. This cost includes the man-hours spent in documenting and reporting security information, as well as a potential loss in reputation leading to decreased business opportunities with potential collaborators, loss of current customers to the competitor firm j, lowered stock market prices, and the like, following the disclosure of a breach or existing security flaws.

Given these gains and losses, the payoff matrix of the *information sharing* game is given by Table I, where the first (second) payoff under each strategy profile represents the row (column) player's payoff. Assume $G > C$. This game is therefore an instance of the *prisoner's dilemma*: in a single shot, simultaneous move game with rational players, the only Nash equilibrium is for neither firm to disclose her security information, despite the fact that the outcome in which firms share security information would be more beneficial to both participants. This observation is consistent with similar studies of one-shot information sharing games in [15], [1], which also conclude that, in the absence of audit mechanisms or secondary incentives, firms will choose to share no information because of the associated disclosure costs.

B. Repeated interactions: the role of monitoring

A one shot formulation of the information sharing game entails that firms do not anticipate future interactions, and so expect no retribution or reward outside this game following their decisions. In contrast, a repeated game formulation can not only capture the notions of trust and reputation that exist in a real-world scenario, but can also leverage firms' interest in maintaining a good reputation to sustain cooperation among them [14]. In the following sections, we are interested in understanding firms' decisions regarding information sharing under repeated interactions. To this end, we need to specify the available monitoring capabilities; i.e., how and to what extent can a firm tell whether the other is following their agreement.

Private monitoring: first, assume each firm conducts her own monitoring and forms a belief about the other firm's disclosure decision. Specifically, by monitoring firm j's externally observed security posture, firm i forms a *belief* b_i about j's report. We let $b_i = 1$ indicate a belief by firm i that firm j has been honest and is fully disclosing all information, and $b_i = 0$ otherwise. In other words, $b_i = 0$ indicates that firm i's monitoring provides her with evidence that firm j has experienced an undisclosed breach or has an unreported security flaw. We assume that the monitoring technology and the inputs to it, and hence the belief b_i of firm i is imperfect, private, and independent of firm j's belief b_j about firm i. Formally, we assume the following distribution on firm i's belief given firm j's report:

$$\pi(b_i|r_j) = \begin{cases} \epsilon, & \text{for } b_i = 0, r_j = 1 \\ 1-\epsilon, & \text{for } b_i = 1, r_j = 1 \\ \alpha, & \text{for } b_i = 0, r_j = 0 \\ 1-\alpha, & \text{for } b_i = 1, r_j = 0 \end{cases} \tag{1}$$

with $\epsilon \in (0, 1/2)$ and $\alpha \in (1/2, 1)$. First, note that ϵ is in general assumed to be small; therefore, if firm j fully discloses all information ($r_j = 1$), firm i's belief will be almost consistent with the received information. Intuitively, this entails the assumption that with only a small probability ϵ, firm i will be observing flaws or breaches that have gone undetected by firm j herself, as internal monitoring is more accurate than externally available information. On the other hand, firm i has an accuracy α in detecting when firm j conceals security information ($r_j = 0$). Note that ($\epsilon = 0, \alpha = 1$) is equivalent to the special case of perfect monitoring.

Public monitoring: alternatively, consider an independent entity (the government, a white hat, or a research group), referred to as *the monitor*, who assesses the comprehensiveness of firms' disclosure decisions, and publicly reveals the results. We assume the distribution of the beliefs (b_i, b_j) formed by the monitor is:

$$\hat{\pi}((b_i, b_j)|(r_i, r_j)) := \pi(b_i|r_j)\pi(b_j|r_i) \ . \tag{2}$$

where the distributions $\pi(b_i|r_j)$ and $\pi(b_j|r_i)$ follow (1), with ϵ and α interpreted similarly. Note that the monitoring technology of the monitor, i.e. (α, ϵ), may in general be more accurate than that available to the firms.

We next analyze the possibility of sustaining cooperation among firms over repeated interactions using the described private and public monitoring technologies.

III. IMPERFECT PRIVATE MONITORING

In this section, we consider the role of private monitoring in providing inter-temporal incentives for information sharing. Unlike repeated games with imperfect public monitoring, relatively less is known about games with private monitoring [16]. To illustrate why, we first study a two stage game: the first stage is an information sharing game, followed by coordination on a business partnership that is contingent on the first period outcome. We show that cooperation in the information sharing agreement can only be partially (i.e., inefficiently) sustained due to the private nature of monitoring outcomes. We then discuss the implications of this observation on the possibility of attaining efficiency when the game is played infinitely often.

A. A two stage game

We begin our analysis of repeated information sharing games by considering a two-stage interaction among the firms. In the first stage, firms decide whether to fully disclose their security information, with payoffs following those of the single stage game of Table I. Similar to [14, Ch. 12.1], we assume that the second stage game is described by a coordination game, with a payoff matrix given in Table II. This game

TABLE II
COORDINATION ON A PARTNERSHIP

	H	L
H	h,h	0,0
L	0,0	ℓ,ℓ

captures decisions on a subsequent business agreement among firms, with H (L) denoting a high (low) profit partnership. High profit partnerships yield a better payoff to both firms, with $h > \ell > 0$, and a partnership will be established only if both firms agree on its type. We assume no discounting on future payoffs. In addition, all payoffs are observed at the end of the second stage, so that first-stage payoffs are uninformative about firms' disclosure decisions. We are interested in strategies which, through leveraging the second stage outcome, can support information disclosure (i.e., $r_i = r_j = 1$) in the first stage. For example, a potential candidate is a *trigger* strategy: firms coordinate on a high profit partnership in the second stage if and only if both have followed $(r_i, r_j) = (1, 1)$ in the first stage. If firms could perfectly observe first stage decisions, a trigger strategy would be an equilibrium for $C < h - \ell$. Nevertheless, in practice each firm can (at best) only imperfectly assess the honesty and comprehensiveness of the other's report, as given by the monitoring in (1).

1) Pure strategy equilibria: We first attempt to identify a pure strategy equilibrium that supports $(r_i, r_j) = (1, 1)$ in the first period, by conditioning the second stage partnership on the first stage decisions. We start by considering the second stage coordination game. First note that it is optimal for a firm i to play H in this game if and only if she assigns probability of at least $\frac{\ell}{h+\ell}$ to firm j also playing H.

For concreteness, we assume that firms follow trigger strategies, requiring a firm i to play H in the second period if and only if she observes $b_i = 1$ following the first stage. Then, if firm i follows $r_i = 1$ in the first period, she knows (with high probability $1 - \epsilon$) that firm j is going to play H in the second stage. As a result, firm i will always play H, *regardless of her belief* b_i *about* r_j, and will therefore not be following the prescription of the trigger strategy. We conclude that the trigger strategy is not an equilibrium.

Following a similar argument, we conclude that in general, firm i's belief about firm j's action in the second period is independent of i's observed signal. In other words, it not sequentially rational for firm i to consider her signal in the second period. Therefore, with pure strategies, inter-temporal incentives can not be used to coordinate on $(r_i, r_j) = (1, 1)$.

2) First period randomization: Next, consider allowing firms' to randomize their actions in the first period. This in turn allows the observed signals to carry information that is helpful for continuation plays. Formally, suppose in the first period, firm i plays $r_i = 1$ with probability β, and $r_i = 0$ otherwise. In the second period, this firm will play H if and only if she has played $r_i = 1$ in the first period, and she has a belief $b_i = 1$ about firm j. We are interested in identifying a mixing probability β for the first period decisions that would lead to this strategy profile being an equilibrium.

We start by analyzing the second period optimality of this strategy. Denote the probability of firm j playing H by p. Then it is optimal for firm i to play H if and only if $hp > \ell(1-p)$. Therefore, it is optimal for firm i to play H if and only if firm j plays H with probability at least $\frac{\ell}{\ell+h}$, and to play L if and only if firm j plays L with probability at least $\frac{h}{\ell+h}$. We therefore need the following to hold:

$$P(j \text{ plays } H \mid r_i = 1, b_i = 1) = \frac{P(r_j=1,b_j=1,r_i=1,b_i=1)}{P(r_i=1,b_i=1|r_j=1)P(r_j=1)+P(r_i=1,b_i=1|r_j=0)P(r_j=0)} = \frac{\beta(1-\epsilon)^2}{\beta(1-\epsilon)+(1-\beta)(1-\alpha)} \geq \frac{\ell}{\ell+h} \tag{3}$$

Similarly,

$$P(j \text{ plays } L \mid r_i = 1, b_i = 0) = \frac{\beta\epsilon^2 + (1-\beta)\alpha}{\beta\epsilon + (1-\beta)\alpha} \geq \frac{h}{\ell+h} , \tag{4}$$

$$P(j \text{ plays } L \mid r_i = 0, b_i = 1) = \frac{\beta\alpha(1-\epsilon) + (1-\beta)(1-\alpha)}{\beta(1-\epsilon) + (1-\beta)(1-\alpha)} \geq \frac{h}{\ell+h} , \tag{5}$$

$$P(j \text{ plays } L \mid r_i = 0, b_i = 0) = \frac{\beta\epsilon\alpha + (1-\beta)\alpha}{\beta\epsilon + (1-\beta)\alpha} \geq \frac{h}{\ell+h} . \tag{6}$$

All of the above inequalities will hold if ϵ is sufficiently small, α is sufficiently large, and β is sufficiently large and bounded away from 1.

We now consider the first period incentives. For the mixing probability β to be an equilibrium, we require the firm to be indifferent between the expected payoff following $r_i = 1$:

$$\begin{aligned} &P(r_j = 1)[u_{i,t=1}(1,1) + P(b_i = 1, b_j = 1)u_{i,t=2}(H,H) \\ &\qquad + P(b_i = 0, b_j = 0)u_{i,t=2}(L,L)] \\ &+ P(r_j = 0)[u_{i,t=1}(1,0) + P(b_j = 0)u_{i,t=2}(L,L)] \\ &= \beta(G - C + (1-\epsilon)^2 h + \epsilon^2 \ell) + (1-\beta)(-C + \alpha\ell) , \end{aligned}$$

and similarly, that following $r_i = 0$:

$$\beta(G + \alpha\ell) + (1-\beta)(0 + \ell) .$$

Solving the above for β as a function of the monitoring parameters leads to:

$$\beta(\alpha, \epsilon) = \frac{C + (1-\alpha)\ell}{(1-\epsilon)^2 h + (1 + \epsilon^2 - 2\alpha)\ell} . \tag{7}$$

In particular, when $\epsilon \to 0$ and $\alpha \to 1$, i.e., as monitoring technologies become almost perfectly accurate, we get:

$$\beta \to \frac{C}{h - \ell} .$$

Note that by cooperation in the first period (conditional on the other firm playing H if and only if first period play is $(1, 1)$), a player forgoes a gain of C in the first period in return for

a gain of $h - \ell$ in the subsequent business agreement. The above limit on the mixing probability therefore implies that if monitoring is accurate enough, as the gap between first stage loss and second stage gain decreases, both firms will be playing $r_i = 1$ with a higher probability. In other words, if deviations are only of limited benefit, firms will be truthfully disclosing their security information more frequently. Also interestingly, (7) illustrates that, if the high partnership is made more lucrative (by increasing the high and low profit partnership gap), cooperation becomes less frequent. This is because firms will receive higher gains if they happen to end up cooperating in the second stage, thus deciding to shift the mixing probability β to harness gain from defection instead.[2] In conclusion, to support more frequent collaboration, it is better to maintain a small gap between first stage loss C and second stage gain $h - \ell$.

We conclude by noting that, despite the fact that the constructed equilibrium has led to $(r_i = 1, r_j = 1)$ being played in the first period with positive probability, outcomes in which either player conceals security information, or in which they fail to coordinate in the second stage despite cooperation in the first period, can still emerge with a positive probability as well, making the equilibrium inefficient. These are a result on the private nature of monitoring signals. There exist alternative equilibrium construction methods, particularly by introducing a public correlating device, which can lead to more efficient equilibria (see [14]). Nevertheless, it will still be impossible to guarantee $(r_i = 1, r_j = 1)$ with probability one.

B. The infinitely repeated game

In light of our observations in the two stage game, in this section, we ask whether it is possible to attain better results by considering longer lasting interactions, and in particular, infinitely repeated information sharing games. A longer history of play can allow for more elaborate strategies; e.g., punishment (non-cooperation) periods that only start after a certain number of observed deviations, or that last only for a certain number of rounds. Therefore, one may expect the possibility of supporting cooperation efficiently. We consider the stage game of Table I repeated infinitely, with the conditionally independent private monitoring given in (1).

In particular, we are interested in a folk theorem for this repeated game. A folk theorem provides a full characterization of payoffs (of which efficient payoffs in terms of social optimality are of particular interest) that can be achieved in a repeated game if players are sufficiently patient (i.e., their future payoffs are sufficiently important to them). With imperfect public monitoring, [12], [13] present a folk theorem under relatively general conditions. The possibility of this result hinges heavily on that players share common information on each others actions (i.e., the public monitoring outcome), as a result of which it is possible to recover a recursive structure for the game, upon which the folk theorem is based; see Section IV-A. However, a similar folk theorem with private monitoring remains an open problem, mainly due to the lack of a common public signal. Nevertheless, the possibility of cooperation, and in particular folk theorems, have been shown to exist for some particular classes of games. These include:

- Games in which firms are allowed to communicate (cheap talk) after each period. This approach has been proposed in [17], [18], and in essence, allows communication to serve as a public signal, allowing players to achieve cooperation.
- Games in which firms have public actions (e.g., announcement of sanctions) in addition to private decisions (here, disclosure decisions), as proposed by [19] for the study of international trade agreements. Intuitively, public actions serve a similar purpose as communication, allowing players to signal the initiation of punishment phases.
- Games with almost public monitoring, i.e., private monitoring with signals that are sufficiently correlated. With such signals, [20] proves a folk theorem for almost-perfect and almost-public monitoring.

We leave an analysis of the possibility of using similar ideas in information sharing agreements as a direction of future work. Alternatively, we next analyze the role of a monitor in providing imperfect public monitoring to make cooperation among the firms possible.

IV. Imperfect Public Monitoring

The possibility of public monitoring (either perfect or imperfect) simplifies the provision of inter-temporal incentives to a great extent. With perfect public monitoring, deviations from the intended equilibrium path are perfectly observable by all players, and can be accordingly punished. As a result, it is possible to design appropriate punishment phases (i.e., a finite or infinite set of stage games in which deviators receive a lower payoff) that keep sufficiently patient players from deviating to their myopic (stage game) best responses. With imperfect public monitoring on the other hand, deviations can not be detected with complete certainty. Nevertheless, the publicly observable signals can be distributed so that some are more indicative that a deviation has occurred. In that case, as players can all act based on their observations of the same signal to decide whether to start punishment or cooperation phases, despite the fact that punishment phases may still occur on the equilibrium path, it is possible for the players to cooperate to attain higher payoffs than those of the stage game.

In the remainder of this section, we first formalize the above intuition by providing some preliminaries on infinitely repeated games with imperfect public monitoring, and in particular, by discussing a folk theorem for these games. We then discuss how this folk theorem applies to information sharing games with monitoring given by (2), and the implications of this application.

[2]Note that this observation is a consequence of firms' risk neutrality in this model. In contrast, a risk averse firm would increase cooperation frequency in response to a more lucrative second stage payoff.

A. The folk theorem

In this section, we present the folk theorem due to Fudenberg, Levine, and Maskin [13]. Consider n rational players. At the stage game, each player i chooses an action $r_i \in R_i$. Let $\mathbf{r} \in R := \prod_{i=1}^{n} R_i$ denote a profile of actions. At the end of each stage, a public outcome $b \in B$ is observed by all players, where B is a finite set of possible signals. The realization of the public outcome b depends on the profile of actions $\mathbf{r}$. Formally, assume the probability of observing b following $\mathbf{r}$ is given by $\pi(b|\mathbf{r})$. Let $u_i^*(r_i, b)$ be the utility of player i when she plays r_i and observes the signal b. Note that i's utility depends on others' actions only through b, and thus the stage payoffs are not informative about others' actions. The ex-ante stage game payoff for user i when $\mathbf{r}$ is played is given by:

$$u_i(\mathbf{r}) = \sum_{b \in B} \pi(b|\mathbf{r}) u_i^*(r_i, b).$$

Let $\mathcal{F}^\dagger$ denote the set of convex combinations of players' payoffs for outcomes in R, i.e., the convex hull of $\{(u_1(\mathbf{r}), \ldots, u_n(\mathbf{r}))|\mathbf{r} \in R\}$. We refer to $\mathcal{F}^\dagger$ as the set of *feasible* payoffs. Of this set of payoffs, we are particularly interested in those that are *individually rational*: an individually rational payoff profile $\mathbf{v}$ is one that gives each player i at least her minmax payoff $\underline{v}_i := \min_{\boldsymbol{\rho}_{-i}} \max_{r_i} u_i(r_i, \boldsymbol{\rho}_{-i})$ (where $\boldsymbol{\rho}_{-i}$ denotes a mixed strategy profile by players other than i). Let $\boldsymbol{\rho}^i := \arg\min_{\boldsymbol{\rho}_{-i}} \max_{r_i} u_i(r_i, \boldsymbol{\rho}_{-i})$ denote the minmax profile of player i, and $\mathcal{F}^* := \{\mathbf{v} \in \mathcal{F}^\dagger | v_i > \underline{v}_i, \forall i\}$ denote the set of feasible and strictly individually rational payoffs. The main purpose of a folk theorem is to specify which of the payoffs in $\mathcal{F}^*$ (of which Pareto efficient payoffs are of particular interest) can be supported (as average payoffs) by some equilibrium of the repeated game.

Let us now discuss the repeated game. When the stage game is played repeatedly, at time t, each player has a private history containing her own past actions, $h_i^{t-1} := \{r_i^0, \ldots, r_i^{t-1}\}$, as well as a public history of the public signals observed so far, $h^{t-1} := \{b^0, \ldots, b^{t-1}\}$. Player i then uses a mapping σ_i^t from (h_i^{t-1}, h^{t-1}) to (a probability distribution over) R_i to decide her next play. We refer to $\sigma_i = \{\sigma_i^t\}_{t=0}^{\infty}$ as player i's strategy. Each player discounts her future payoffs by a discount factor δ. Hence, if player i has a sequence of stage game payoffs $\{u_i^t\}_{t=0}^{\infty}$, her average payoff throughout the repeated game is given by $(1-\delta)\sum_{t=0}^{\infty} \delta^t u_i^t$. Player is choosing her strategy σ_i to maximize this expression.

Among the set of all possible strategies σ_i, we will consider *public strategies*: these consist of decisions σ_i^t that depend only on the public history h^{t-1}, and not on player i's private information h_i^{t-1}. Whenever other players are playing public strategies, then player i will also have a public strategy best-response. A *perfect public equilibrium (PPE)* is a profile of public strategies that, starting at any time t and given any public history h^{t-1}, form a Nash equilibrium of the game from that point on. PPEs facilitate the study of repeated games to a great extent, as they are "recursive". This means that when a PPE is being played, the continuation game at each time point is strategically isomorphic to the original game, and therefore the same PPE is induced in the continuation game as well. Note that such recursive structure can not be recovered using private strategies, leading to the comparatively limited results in private monitoring games, as discussed in Section III. Let $\mathcal{E}(\delta)$ be the set of all payoff profiles that can be attained using public strategies as PPE average payoffs when the discount factor is δ. We know that $\mathcal{E}(\delta) \subseteq \mathcal{F}^*$. The main question is under what conditions does the reverse hold, i.e., when is it possible to attain any point in the interior of $\mathcal{F}^*$ (particularly nearly efficient payoffs) as PPE payoffs?

In order to attain nearly efficient payoffs, players need to be able to support cooperation by detecting and appropriately punishing deviations. In PPEs, where strategies are public, all such punishment should occur solely based on the public signals. As a result, the public signals should be distributed such that they allow players to statistically distinguish between deviations by two different players, as well as different deviations by the same player. We now formally specify these conditions. The first condition, referred to as *individual full rank*, gives a sufficient condition under which deviations by a single player are statistically distinguishable; i.e., the distribution over signals induced by some profile $\boldsymbol{\rho}$ are different from that induced by any $(\rho_i', \boldsymbol{\rho}_{-i})$ for $\rho_i' \neq \rho_i$. Formally,

Definition 1: The profile $\boldsymbol{\rho}$ has individual full rank for player i if given the strategies of the other players, $\boldsymbol{\rho_{-i}}$, the $|R_i| \times |B|$ matrix $A_i(\boldsymbol{\rho_{-i}})$ with entires $[A_i(\boldsymbol{\rho_{-i}})]_{r_i,b} = \pi(b|r_i, \boldsymbol{\rho_{-i}})$ has full row rank. That is, the $|R_i|$ vectors $\{\pi(\cdot|r_i, \boldsymbol{\rho}_{-i})\}_{r_i \in R_i}$ are linearly independent.

The second general condition, *pairwise full rank*, is a strengthening of individual full rank to pairs of players. In essence, it ensures that deviations by players i and j are distinct, as they introduce different distributions over public outcomes. Formally,

Definition 2: The profile $\boldsymbol{\rho}$ has pairwise full rank for players i and j if the $(|R_i| + |R_j|) \times |B|$ matrix $A_{ij}(\boldsymbol{\rho}) := [A_i(\boldsymbol{\rho_{-i}}); A_j(\boldsymbol{\rho_{-j}})]$ has rank $|R_i| + |R_j| - 1$.

Therefore, given an adequate public monitoring signal, we have the following (minmax-threat, full) folk theorem under imperfect pubic monitoring.

The imperfect public monitoring folk theorem: Assume R is finite, the set of feasible payoffs $\mathcal{F}^\dagger \subset \mathbb{R}^n$ has non-empty interior, and all the pure action equilibria leading the extreme points of $\mathcal{F}^\dagger$ have pairwise full rank for all pairs of players. If the minmax payoff profile $\underline{\mathbf{v}} = (\underline{v}_1, \ldots, \underline{v}_n)$ is inefficient, and the minmax profile $\hat{\boldsymbol{\rho}}^i$ has individual full rank for each player i, then for any profile of payoffs $\mathbf{v} \in \text{int}\mathcal{F}^*$, there exists a discount factor $\underline{\delta} < 1$, such that for all $\delta \in (\underline{\delta}, 1)$, $\mathbf{v} \in \mathcal{E}(\delta)$.

B. Supporting cooperation in information sharing

We can now verify that the above folk theorem applies to information sharing games with imperfect public monitoring strategies given by (2). That is, when the firms are sufficiently patient, they can sustain cooperation on full security information sharing in a repeated setting, by making their disclosure decisions based only on the imperfect, publicly announced

observations of the monitor about their past actions. To this end, we need to verify that the conditions of the folk theorem, in particular those on the public monitoring signal, hold for (2).

We first verify that the minmax profile of the repeated information sharing game has individual full rank for either firm. The minmax action profile for player i, $\mathbf{r}^i$, is both firms concealing their information, i.e., $(r_i^i, r_j^i) = (0,0)$. Then, $A_i(\mathbf{r}^i)$ is given by:

$$\begin{array}{c} \mathbf{b} = \\ r_i = 0 \\ r_i = 1 \end{array} \begin{array}{c} \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \end{array} \\ \begin{pmatrix} \alpha^2 & (1-\alpha)\alpha & \alpha(1-\alpha) & (1-\alpha)^2 \\ \epsilon\alpha & (1-\epsilon)\alpha & \epsilon(1-\alpha) & (1-\epsilon)(1-\alpha) \end{pmatrix} \end{array}$$

The rows of the above matrix are linearly independent (given $\alpha \neq \epsilon$), and hence the minmax profiles have individual full rank for both players.

We also need to verify that all pure strategy action profiles have pairwise full rank. We do so for $\mathbf{r} = (1,0)$, the remaining can be shown similarly. For $\mathbf{r} = (1,0)$, the matrix $A_{ij}((1,0)) := [A_i(r_j = 1); A_j(r_i = 0)]$ is given by:

$$\begin{array}{c} \mathbf{b} = \\ r_i = 0 \\ r_i = 1 \\ r_j = 0 \\ r_j = 1 \end{array} \begin{array}{c} \begin{array}{cccc} (0,0) & (1,0) & (0,1) & (1,1) \end{array} \\ \begin{pmatrix} \alpha\epsilon & (1-\alpha)\epsilon & \alpha(1-\epsilon) & (1-\alpha)(1-\epsilon) \\ \epsilon^2 & \epsilon(1-\epsilon) & (1-\epsilon)\epsilon & (1-\epsilon)^2 \\ \alpha^2 & (1-\alpha)\alpha & \alpha(1-\alpha) & (1-\alpha)^2 \\ \alpha\epsilon & (1-\alpha)\epsilon & \alpha(1-\epsilon) & (1-\alpha)(1-\epsilon) \end{pmatrix} \end{array}$$

It is straightforward to verify that the above has row rank 3; i.e., that the first three rows are linearly independent. As a result, $\mathbf{r} = (1,0)$ has pairwise full rank. A similar procedure shows that the remaining pure action profiles also have pairwise full rank.

We therefore conclude that the conditions of the folk theorem of Section IV-A hold with the public signals distributed according to (2), and as a result, when the firms are sufficiently patient, i.e., they value the future outcomes of their information sharing agreement, it is possible for them to nearly efficiently cooperate on full information disclosure through repeated interactions.

Note that the above observation holds despite the fact that the monitoring technology employed by the monitor has the same accuracy (α, ϵ) as that of the individual firms. That is, the folk theorem holds regardless of the accuracy of signals once they satisfy the appropriate full rank conditions. Nevertheless, the accuracy of the monitoring will indirectly affect the achievability of nearly efficient payoffs through $\underline{\delta}$. A lower α and/or a higher ϵ make the requirements on $\underline{\delta}$ more severe; i.e., firms should be more patient for the folk theorem to hold. Determining the dependence of $\underline{\delta}$ on (α, ϵ) is a direction of future work.

Finally, it is worth mentioning that given the statement of the folk theorem in Section IV-A, our conclusions are equally applicable if the current setup is extended to games with non-binary but finite disclosure decisions and monitoring outputs; that is, information sharing agreements in which firms can select a level of disclosure from a finite set, and in which the assessment system assigns a (discrete) rating to each firm.

V. Related work

A number of research papers have analyzed the welfare implications of information sharing agreements, as well as firms' incentives for adhering to these agreements, using game-theoretic models of one shot information sharing games.

The work by [21] and [1] consider the effects of security breach reporting between firms and an authority. [21] show that if the availability of shared information[3] can reduce either attack probabilities or firms' interdependency, it will benefit social welfare by inducing firms to improve investments in self-protection and cyber-insurance. On the other hand, [1] studies the effectiveness of mandatory breach reporting, and shows that enforcing breach disclosure to an authority (through the introduction of audits and sanctions) is effective in increasing social welfare only under certain conditions, including high interdependence among firms and low disclosure costs.

[15] and [9] propose game-theoretic models of information sharing among firms, arriving at seemingly contradictory results. [15] studies the welfare implications of information sharing among two firms. The authors show that, if security information from a partner firm is a *substitute* to a firm's own security expenditures, then (mandatory) information sharing laws reduce expenditure in security measures, but can nevertheless increase social welfare. However, firms will not voluntarily comply with sharing agreements, requiring additional economic incentives to be in place (e.g., a charge on a member of the ISAC for losses on the other member). A similar study is conducted by [9], where two firms in a competitive market environment select security expenditure and information sharing levels, as well as product pricing. Here, although information sharing entails some disclosure costs for firms (similar to [15]), it may also increase consumer confidence in the firm, as it is believed that the firm is taking steps towards securing her system. As a result, sharing by the partner firm acts as a *complement* to the firm's own security expenditures; i.e., increased sharing by the partner increases the firm's expenditure in security, so that the firm can maintain her share of the competitive market. Using this model, the authors show that when the positive demand effects of information sharing are high enough, added expenditure and/or sharing by one firm can incentivize the other firm to also increase her expenditure and/or sharing levels. Therefore, these studies illustrate how firms' incentives for voluntarily sharing security information is highly dependent on direct disclosure costs, as well as its demand-side implications.

In this work, similar to [15], [21], [1], we assume disclosure costs are higher than potential demand-side benefits, therefore similarly predicting a lack of voluntary information sharing at equilibrium. Our proposed approach of considering repeated interactions as an incentive solution is however different from those proposed in aforementioned literature. To the best of our knowledge, our work is the first to study security information sharing games in a repeated game framework.

[3]Firms' incentives for information disclosure or the mechanisms for ensuring breach disclosure have not been modeled in [21].

VI. Conclusion

We studied firms' incentives for information disclosure in security information sharing agreements. By formulating a single stage information sharing game as a prisoner's dilemma scenario, we observe that disclosure costs lead firms to exhibit free-riding behavior, despite the fact that an outcome in which firms fully disclose their security information would be preferred by both firms. We proposed a repeated-game approach to this problem, and discussed the role of monitoring (private vs. public) on determining whether inter-temporal incentives can lead to the support of cooperation (i.e., full disclosure).

Specifically, we illustrated the limitations arising due to the private nature of firms' beliefs when monitoring is carried out independently by firms, and briefly discussed the possibility of modifying the game to allow for communication, almost public monitoring, and possibility of public actions, in order to support cooperation in infinitely repeated interactions. A detailed analysis of these approaches, and their implications on the role of authorities as facilitators of public monitoring or communication, is a direction of future work.

Alternatively, we considered the role of a public monitoring/assessment system in facilitating coordination among firms. We showed that even thought the system uses the same monitoring technology as the individual firms, the resulting inter-temporal incentives can support cooperation among sufficiently patient firms. As part of our future work, we are interested in identifying similar inter-temporal incentives under extensions of the current model. First, it would be interesting to study firms' optimal disclosure decisions if they are allowed to choose partial sharing levels (i.e., from a continuous decisions set). Similarly, the output of the rating system could be chosen from a continuous set, taking the form of a "reputation" score for each firm. By assuming that such reputation affects firms payoffs not only in the information sharing agreement, but also in other business partnerships/interactions, it is of interest to study firms' optimal disclosure decisions, as well as the optimal level of reputation each firm is willing to maintain.

Acknowledgment

This material is based on research sponsored by the Department of Homeland Security (DHS) Science and Technology Directorate, Homeland Security Advanced Research Projects Agency (HSARPA), Cyber Security Division (DHS S&T/HSARPA/CSD), BAA 11-02 via contract number HSHQDC-13-C-B0015.

References

[1] S. Laube and R. Böhme, "The economics of mandatory security breach reporting to authorities," in *Workshop on the economics of information security (WEIS)*, 2015.

[2] "Executive order 13691: Promoting private sector cybersecurity information sharing," https://www.whitehouse.gov/the-press-office/2015/02/13/executive-order-promoting-private-sector-cybersecurity-information-shari, 2015.

[3] L. A. Gordon, M. P. Loeb, W. Lucyshyn, and T. Sohail, "The impact of the sarbanes-oxley act on the corporate disclosures of information security activities," *Journal of Accounting and Public Policy*, vol. 25, no. 5, pp. 503–530, 2006.

[4] T. Claburn, "Most security breaches go unreported," http://www.darkreading.com/attacks-and-breaches/most-security-breaches-go-unreported/d/d-id/1070576?, 2008.

[5] ——, "Data breaches made possible by incompetence, carelessness," http://www.darkreading.com/risk-management/data-breaches-made-possible-by-incompetence-carelessness/d/d-id/1068741?page_number=1, 2008.

[6] ThreatTrack, "Majority of malware analysts aware of data breaches not disclosed by their employers," http://www.threattracksecurity.com/press-release/majority-of-malware-analysts-aware-of-data-breaches-not-disclosed-by-their-employers.aspx, 2013.

[7] K. Campbell, L. A. Gordon, M. P. Loeb, and L. Zhou, "The economic cost of publicly announced information security breaches: empirical evidence from the stock market," *Journal of Computer Security*, vol. 11, no. 3, pp. 431–448, 2003.

[8] H. Cavusoglu, B. Mishra, and S. Raghunathan, "The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers," *International Journal of Electronic Commerce*, vol. 9, no. 1, pp. 70–104, 2004.

[9] E. Gal-Or and A. Ghose, "The economic incentives for sharing security information," *Information Systems Research*, vol. 16, no. 2, pp. 186–208, 2005.

[10] DHS, "Enhancing resilience through cyber incident data sharing and analysis," http://www.dhs.gov/sites/default/files/publications/Data%20Categories%20White%20Paper%20-%20508%20compliant.pdf, 2015.

[11] "Vocabulary for event recording and incident sharing," http://veriscommunity.net/index.html, 2015.

[12] D. Abreu, D. Pearce, and E. Stacchetti, "Toward a theory of discounted repeated games with imperfect monitoring," *Econometrica: Journal of the Econometric Society*, pp. 1041–1063, 1990.

[13] D. Fudenberg, D. Levine, and E. Maskin, "The folk theorem with imperfect public information," *Econometrica*, vol. 62, no. 5, pp. 997–1039, 1994.

[14] G. J. Mailath and L. Samuelson, *Repeated games and reputations*. Oxford university press Oxford, 2006, vol. 2.

[15] L. A. Gordon, M. P. Loeb, and W. Lucyshyn, "Sharing information on computer systems security: An economic analysis," *Journal of Accounting and Public Policy*, vol. 22, no. 6, pp. 461–485, 2003.

[16] M. Kandori, "Introduction to repeated games with private monitoring," *Journal of Economic Theory*, vol. 102, no. 1, pp. 1–15, 2002.

[17] O. Compte, "Communication in repeated games with imperfect private monitoring," *Econometrica*, pp. 597–626, 1998.

[18] M. Kandori and H. Matsushima, "Private observation, communication and collusion," *Econometrica*, pp. 627–652, 1998.

[19] J.-H. Park, "Enforcing international trade agreements with imperfect private monitoring," *The Review of Economic Studies*, vol. 78, no. 3, pp. 1102–1134, 2011.

[20] G. J. Mailath and S. Morris, "Repeated games with almost-public monitoring," *Journal of Economic Theory*, vol. 102, no. 1, pp. 189–228, 2002.

[21] H. Ogut, N. Menon, and S. Raghunathan, "Cyber insurance and it security investment: Impact of interdependence risk." in *WEIS*, 2005.