On the Provision of Public Goods on Networks: Incentives, Exit Equilibrium, and Applications to Cyber Security

by

Parinaz Naghizadeh Ardabili

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in the University of Michigan
2016

Doctoral Committee:

Professor Mingyan Liu, Chair
Professor Martin Loeb, University of Maryland
Associate Professor Vijay Subramanian
Professor Demosthenis Teneketzis
Professor Michael Wellman

To Maman, Baba, and Parisa

Acknowledgements

I am very fortunate to have had the opportunity to complete this dissertation with the help and support of fantastic mentors, colleagues, and friends. First and foremost, I would like to express my deepest gratitude to my advisor, Professor Mingyan Liu, for her continued guidance and support. Thank you for all your advice, encouragement, and kindness. It has been a true privilege to have you as my mentor.

I am very grateful to Professor Demosthenis Teneketzis for being an excellent teacher and a great mentor to me. I would also like to express my gratitude to my committee members, Professor Martin Loeb, Professor Vijay Subramanian, and Professor Michael Wellman, for the helpful discussions and insightful feedback.

I feel very lucky to have been surrounded by amazing friends in Ann Arbor. Hamidreza, thank you for being a great colleague and the best friend. Azadeh, what are the odds? May they continue in our favor. Mai, thanks for fueling me through all my deadlines. Thanks also to Maryam, Arash, Amir, Katayoon, Parisa, Mehrzad, Elnaz, Armin, Payam, Diane, Berk, Yi-Chin, Shang-Pin, Qingsi, Yang, Armin, and the many more friends and colleagues I've been lucky to meet in Ann Arbor. You are what made Ann Arbor wonderful, even when buried deep, deep in snow.

Last but not least, none of this would have been possible without the endless love and support of my parents, Roya and Issa, and my sister, Parisa. One of the proudest moments on my Ph.D. journey was when Parisa found my research interesting and asked me to tell her more! The other was seeing my mom and dad smile as they sat through my defense over Skype. Maman, Baba, Parisa, I love you.

Table of Contents

List of Figures

Figure

List of Tables

Table

List of Appendices

Appendix

Abstract

On the Provision of Public Goods on Networks:
Incentives, Exit Equilibrium, and Applications to Cyber Security
by
Parinaz Naghizadeh Ardabili

Chair: Mingyan Liu

Attempts to improve the state of cyber security have been on the rise over the past years. In addition to improving prevention and protection methods, recent efforts have emphasized the need for ensuring that organizations and individuals have appropriate incentives for adopting better security practices. The importance of incentivizing better security decisions by users in the current landscape is two-fold: it not only helps users protect themselves against attacks, but also provides positive externalities to others interacting with them, as a protected user is less likely to become compromised and be used to propagate attacks against other entities. As a result, security can be viewed as a public good, the optimal provision of which requires the introduction of appropriate regulations or incentive mechanisms.

This thesis takes a game-theoretic approach to understanding the theoretical underpinnings of users' incentives in the provision of public goods, and in particular, cyber security. We analyze the strategic interactions of users in the provision of security as a non-excludable public good. We propose the notion of exit equilibrium to describe users' outside options from mechanisms for incentivizing the adoption of better security decisions, and use it to highlight the crucial effect of outside options on the design of incentive mechanisms for improving the state of cyber security.

We further focus on the general problem of public good provision games on networks, which include a class of security games played on networks as a special case. We identify necessary and sufficient conditions on the structure of the network for the existence and uniqueness of the Nash equilibrium in these games. We show that previous results in the literature can be recovered as special cases of our result. We provide a graph-theoretical interpretation of users' efforts at the Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria of these games, by linking users' effort decisions to their centralities in the interaction network. Using this characterization, we separate the effects of users' dependencies and influences (outgoing and incoming edges in the network, respectively) on their effort levels, and uncover an alternating effect over walks of different length in the network.

We also propose the design of inter-temporal incentives in a particular type of security games, namely, security information sharing agreement. We show that either public or private assessments can be used in designing incentives for participants to disclose their information in these agreements. In the case of private assessments, this is possible if participants are provided with a communication platform. Finally, we present a method for crowdsourcing reputation that can be useful in attaining assessments of users' efforts in security games.

Chapter 1

Introduction

1.1 Motivation: The current state of cyber security

The increased adoption of cyber technology in all aspects of our lives over the past few decades has proven to be a double-edged sword: while it has led to improved connection and ease of interaction for users across the globe, it has also increased the exposure of users to cyber risks, particularly systemic risks. This upward trend, as well as the evolution of types of cyber attacks, have been documented in Verizon's annual Data Breach Investigation Reports (DBIR) [32]. The latest DBIR report [33] uses data from insurance claims to estimate the average cost of a cyber-attack for the compromised entity, forecasting it to be from \$25k to \$8.8M, depending on the number of compromised records. The Ponemon institute conducts similar "Cost of Cyber Crime" studies, with its 2015 report estimating an average cost of cyber crime per company of \$7.7M [61]. Several high-profile security incidents have been covered extensively by the media in the past couple of years, including the Target, JP Morgan, Home Depot, Ashley Madison, and Sony Pictures hacks [29, 63].

This increase in the number of cyber attacks, as well as the considerable financial losses they impose on the affected organizations and individuals, have led to increased attention to methods for improving the state of cyber security. To this end, on one hand there is a need for improved cyber defense technologies as attackers grow ever more sophisticated. On the other hand, there is a need to provide appropriate in-

centives for users to adopt these enhanced technologies. In fact, several studies show that a vast number of attacks could have been prevented had the victims adopted better cyber security practices. For example, experts have argued that the Sony Pictures hack could have been easily prevented by encrypting email communications and having better employee education [19, 108]. Another example is a 2013 study by the Australian Signals Directorate which showed that at least 85% of cyber intrusions they responded to could have been prevented by only 4 mitigating strategies, including limiting administrative privileges and maintaining up-to-date software [6]. Such studies highlight the need for ensuring that organizations/individuals have the incentives to properly adopt existing software and best practices to improve their state of security.

Moreover, the importance of increasing cyber security investments by users in this landscape is two-fold: it not only helps users protect themselves against potential attacks, but it also provides *positive externalities* to other users interacting with them. This is because a protected user is less likely to become compromised and subsequently used for generating or propagating attacks against other entities. As a result, users' investments in security measures can be viewed as a *public good*. Formally, public goods are defined as those that are *non-rivalrous* [84], i.e., goods for which consumption by a user does not reduce its availability to others. In the case of cyber security, the investment by a user is viewed as a non-rivalrous good which, due to its (positive) externality, affects entities other than the investor. Anecdotal evidence also points to the public good nature of security. In a letter to shareholders announcing plans to increase expenditure in improving the company's cyber security posture, JP Morgan CEO Jamie Dimon refers to the complex and interdependent nature of cyber attacks [36]:

> "In our existing environment and at our company, cyber security attacks are becoming increasingly complex and more dangerous. The threats are coming in not just from computer hackers trying to take over our systems and steal our data but also from highly coordinated external attacks both directly and via third-party systems (e.g., suppliers, vendors, partners, exchanges, etc.)"

Another commonly adopted viewpoint reflecting a belief in the interdependent nature of security is that "a system is only as strong as its weakest link" [23], with human factors often regarded as the weakest link [4], and a starting point for improving the security of the entire system. According to Matt Moynahan, the chief executive of the security company Veracode, "the law of the weakest link always seems to prevail" [114].

Consequently, the provision of security in an interconnected system can be viewed as a public good provision problem. It is well known in the economic literature that, in the absence of regulation or incentive mechanisms, the provision of public goods by rational users is in general inefficient [84]. This further motivates the need for designing appropriate mechanisms to incentivize users' improved efforts towards providing the public good (here, improved security investment decisions).

Together, the rising number and cost of cyber incidents, lack of incentives for users to adopt better security practices, and the systemic and interdependent nature of cyber risks, motivate a collective effort towards increasing security investments to their optimal levels. A main focus of this dissertation is to work towards this goal by analyzing the theoretical underpinnings of users' incentives for provision of public goods in interconnected and interdependent systems.

While we will focus on cyber security as the main motivation, the framework and findings of this thesis are applicable to the general problem of provision of public goods over networks. Other applications and illustrative examples are presented throughout the thesis. These include the spread of data, innovation, or research, among firms in an industry, creation of new parks or libraries at neighborhood level in cities, and implementing measures for reducing pollution by neighboring towns.

1.2 Problem formulation and key challenges

1.2.1 Economics of security

The approach of this thesis is to look at security from an economic perspective, and to use tools from mathematical economics, namely game theory and mecha-

nism design, to gain a better understanding of users' incentives towards investing in security. Specifically, users' effort decisions depend on their cognitive abilities. We adopt a game-theoretic point of view, and assume that all users are rational decision-makers. Consequently, they account for other users' actions when optimizing their own security investment decisions. The interactions of such strategic users when providing the public good (security) will constitute a game, henceforth referred to as a *security game*.

From a social welfare perspective, the ideal outcome attainable in a security game is known as the *socially optimal* solution. These are the levels of effort at which the collective cost of security to all users is minimized. However, it is well known in the economics literature that the level of a public good provided by rational users at the status quo is generally far from its socially optimal level, and most often under-provided. This sub-optimality arises from the fact that users do not account for the externality of their actions when choosing effort levels, as they are optimizing only their own payoffs. Moreover, some users further decrease their effort levels, as they can free-ride on the externality of others' actions. Therefore, improving the provision of security to its socially optimal level requires the introduction of additional regulations or incentive schemes. These regulations/incentives result in a modified game structure, the equilibria of which lead to the designer's desired outcome. Proposing the regulations/incentives that lead to an appropriate game structure is the subject of mechanism design theory.

In the realm of cyber security, mechanisms that either dictate or incentivize better security efforts have been proposed in recent literature; see Section 1.3. This thesis will also adopt a mechanism design approach to the problem, with a focus on the design and performance of tax-based incentive mechanisms that guarantee the following desiderata. First, the objective of the proposed mechanism is to maximize social welfare. In addition, the mechanism should ensure *voluntary participation*, i.e., each user should prefer the outcome she attains through participation in the incentive mechanism, to that she could attain by opting out. The importance of ensuring voluntary participation lies in the fact that, in general, the mechanism designer either lacks the authority, or is reluctant, to mandate user cooperation. Finally, we will

be interested in mechanisms that maintain a *weakly balanced budget*. Weak budget balance is a requirement on the taxes collected/distributed by the mechanism; it ensures that the designer can redistribute users' payments as rewards, and ideally either retain a surplus as profit or at least not sustain losses. If the condition is not satisfied, the designer would need to spend (a potentially large amount of) external resources to run a proposed mechanism; such access to financing is not necessarily available to a designer or the users.

The three goals of social optimality, voluntary participation, and weak budget balance, are attainable using tax-based incentive mechanisms in several environments; examples include the financing of public projects (where users choose whether to do a given project or not), allocation of a single indivisible unit of a private good, and allocation of downlink power in cellular networks [115]. We are similarly interested in achieving these goals in security games, using either existing or new incentive mechanisms. One of the contributions of this thesis is to show that the non-excludable nature of security (as well as other non-excludable public goods[1]) introduces additional challenges in this mechanism design problem.

1.2.2 The issue of voluntary participation

A user's decision when contemplating participation in an incentive mechanism depends not only on the structure of the induced game, but also on the options available when staying out. The latter is what sets the study of incentive mechanisms for security games (as well as other non-excludable goods) apart from private or excludable goods.

To highlight this difference, note that due to the non-excludable nature of security,

[1]We note that according to an alternative definition, see e.g., [121, Chapter 23], all public goods are assumed non-excludable, with excludable non-rivalrous goods referred to as *club goods*. However, it is also common in the literature, especially in the engineering applications' literature, to make the coarser distinction of public vs. private goods based on rivalry alone. Examples include Samuelson's seminal work on public goods [113], where he considers private vs. collective consumption goods, the definition of Mas-Colell, Whinston, and Green [84, Chapter 11.C], where public goods are defined based on their *non-depletable* nature, and in the engineering applications' literature, the work of [115] on decentralized power allocation in cellular networks. We adopt this coarser categorization, and further distinguish based on excludability when needed.

a user who opts out from a proposed incentive mechanism can continue benefiting from the spill-over of improved investments by those participating in the mechanism. Similarly, the security decisions of this outlier continue to affect the security outcome for the participating users. This is in contrast to excludable goods, in which an outlier neither benefits from, nor influences, the public good produced by the participating users. For such excludable goods, tax-based mechanisms such as the Externality mechanism (e.g., [115]) and the Pivotal mechanism (e.g., [103]), can be designed so as to incentivize the socially optimal provision of an excludable good, guarantee voluntary participation, and maintain weak budget balance.

Given this distinction, to enable the design and study of similar incentive mechanisms for security games (and non-excludable public goods in general), this thesis proposes the notion of *exit equilibrium*. This new notion captures the different nature of outside options available to users given non-excludability, by accounting for how an outlier benefits from, and influences, the participants. In particular, at the exit equilibrium, a user unilaterally opts out of the proposed incentive mechanism, and best-responds to the remaining users who continue participating (these users are also best-responding to the outlier's action). A mechanism ensures voluntary participation if each user prefers the outcome attained in the socially optimal solution to that she can attain under her exit equilibrium.

Using this notion, this thesis shows a negative result: given that users can opt out to their exit equilibria, there exists no tax-based incentive mechanism that can simultaneously guarantee social optimality, voluntary participation, and weak budget balance, in all instances of security games. As discussed in Chapter 2, it is the non-excludability of the good that gives rise to this conflict. This requires the mechanism designer to consider restricted or modified problem environments, so as to work around this negative result. Accordingly, we proceed with the study of two alternative settings in this thesis. We begin by analyzing a subclass of security games played on networks. As the second alternative, we study the design of inter-temporal incentives by accounting for the repeated nature of users' interactions.

1.2.3 The effects of network structure

Public good provision games can be viewed as a subclass of games played on networks. Networks have been used extensively to model social and economic interactions. In settings where the n nodes represent rational players, games on networks are the n-person games in which network links represent the strengths and types of interactions among these players. Similar to other strategic settings, we may be interested in the existence, uniqueness, and comparative statics of equilibria of these games. For games on networks in particular, we are interested in answers in terms of the network structure and its graph-theoretical metrics. These characterizations can allow us to answer questions such as: how do players' positions in the network affect their effort decisions?, which player(s) should we target with tax/subsidy policies to improve the provision of the public good?, what network structures guarantee that equilibria exist and/or are stable?, and how will modifying the network structure, including adding or removing links, affect players' decisions? Answers to these questions advance the theory of n-person games on networks, and lead to important design and policy implications.

We will focus on the particular class of public good provision games on networks. These will be simultaneous move and complete information games, in which each player's payoff will depend on a linear weighted sum of her neighbors' efforts, with the weights determining the strength and type of interactions among players. Security games played on networks, as well as local public good games, coordination games, peer effect games, and the like, can be modeled in this framework. For these settings, we will consider Nash equilibria, Pareto efficient effort profiles (which include the socially optimal outcomes as a special case), and semi-cooperative equilibria emerging from interactions of coalitions of players (which include exit equilibria as a special case). We will study the existence and uniqueness of Nash equilibria, as well as graph-theoretical interpretations of players' efforts at the Nash equilibria, Pareto efficient profiles, and semi-cooperative equilibria. The findings will provide additional tools needed to address the theoretical challenges and policy design aspects of public good provision games, including security games.

1.2.4 Inter-temporal incentives

Another alternative when designing incentives for users to improve their security decisions is to account for the repeated nature of their interactions. This allows the designer to forgo monetary taxes/rewards, instead introducing *inter-temporal incentives* for users' cooperation. Conditioning users' future cooperation on the history of their past interactions can lead players to act cooperatively due to the prospect of higher future payoffs. This design is carried out in a *repeated game* framework. It has been well-known in the economic literature that repetitions of an otherwise non-cooperative and inefficient game can lead economically rational users to coordinate on efficient equilibria [1, 45, 80]. A prominent example of this phenomenon is that of a prisoner's dilemma game: while two rational players should always defect in 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 interactions can prevent players from behaving opportunistically. Similarly, we are interested in the design of inter-temporal incentives to sustain cooperation in security games.

Motivated by several recent policy initiatives (see Section 1.3), we will focus on a particular class of security games, namely, security information sharing agreements. The action of a user in these games is to decide whether to (fully and honestly) disclose security information, including but not limited to reports on recent successful and failed security breaches, to other users within her agreement. This disclosed information can help improve the state of cyber security, as participants can prevent similar attacks and invest in the best security measures by leveraging other users' experience.

Several studies have shown the positive effects of information sharing laws. Romanosky et al. [110] show that the introduction of breach disclosure laws has resulted in a reduction in identity theft incidents. Gordon et al. [52] argue that shared information can reduce the uncertainty in adopting a cyber security investment, thus leading firms to take a proactive rather than reactive approach to security, and consequently increasing the expected amount of investments in cyber security. Finally,

Gordon et al. [51] show that the Sarbanes-Oxley Act of 2002 (despite only indirectly encouraging higher focus on reporting of security-related information) has had a positive effect on disclosure of information security by organizations.

Nevertheless, anecdotal and empirical evidence suggest that security breaches remain under-reported; see e.g., [24, 120]. These observed disincentives for sharing security information can be primarily explained by analyzing the associated economic impacts. For example, [20, 21] 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 [47]. Finally, documenting and announcing security breaches impose a bureaucratic burden on the company, e.g., when an agreement requires the reports to comply with a certain incident reporting terminology. Examples include the recently proposed categorization by DHS [119], and the Vocabulary for Event Recording and Incident Sharing (VERIS) proposed by the Verizon RISK team [123].

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 can lead to voluntary cooperation by firms in these agreements. Our work in Chapter 4 focuses on this problem, and analyzes the use of public and private indicators of users' cooperation in the design of appropriate inter-temporal incentives in these agreements.

1.3 Related literature and policy initiatives

Related work will be presented throughout the dissertation to highlight the contributions of each chapter within the literature. This section discusses some of the theoretical work that is most closely related in terms of motivation and general methodologies, as well as some recent policy proposals, in order to situate this work within the broader context of cyber security.

1.3.1 Incentive mechanisms for improved security

Provision of security by interconnected users, both in general as well as in the context of cyber security, has been extensively studied in the framework of game theory, see e.g., [70, 122, 54, 55, 74, 64], and [71, 81] for surveys. Security games were first presented by Kunreuther and Heal [70] to study the incentive of airlines to invest in baggage checking systems, and by Varian [122] in the context of computer system reliability.

The majority of this literature focuses on under-investment in security, by comparing users' investments in the Nash equilibrium of the security game to the socially optimal levels of investments. Several methods for increasing users' investments, and thus the reliability of the interconnected system, have also been proposed in the literature. These mechanisms can be grouped into two main categories, based on whether they *incentivize* or *dictate* user cooperation [71]. Mechanisms that dictate user investment in security include regulations, audits, and third party inspections [71]. These methods leverage the power of an authority such as the government or an Internet service provider (ISP), and are therefore only effective if the authority has sufficient control over the users so as to accurately monitor their actions and establish a credible threat of punishment.

In the absence of authorities with regulation power, or in order to preserve users' autonomy and privacy, a designer can choose mechanisms that attempt to incentivize improved security behavior. A commonly proposed form of these incentives, also adopted in this thesis, is the introduction of monetary taxes/rewards or transfers; see e.g., [70, 122, 55]. Another increasingly popular incentive mechanism is *cyber insurance*; we further elaborate on this approach and its relation to this thesis in Section 1.3.2. The primary focus of this thesis in Chapter 2 is on the design of tax-based incentive mechanisms. It is the first in the literature to focus specifically on the issue of voluntary participation in these mechanisms, and to illustrate the complexities arising from the non-excludable nature of security.

1.3.2 Cyber insurance: Theory and practice

The study of cyber insurance, both as a risk transfer mechanism, i.e., as a means of managing residual security risks, as well as a potential solution to the problem of under-investment in security in interdependent systems, has been receiving considerable attention in both theory and practice. Using insurance, users transfer part of their security risks to an insurer, receiving coverage for damages due to certain pre-specified cyber incidents, in return for paying a premium fee.

There are currently over 30 insurance carriers offering cyber insurance contracts in the U.S. [109, 10]. Many insurers have reported growths of 10-25% in premiums in a 2012 survey of the market [10], with some carriers reporting even higher rates. For example, one carrier reports an increase of over 30% in the number of clients purchasing their contracts during both 2012 and 2014 [82, 83]. The total amount of premiums written are estimated to be between $500M and $1bn [109]. Typical premiums are estimated to start from $10k -$25k and go as high as $50M [109, 10]. The average limits of these contracts for one of the insurance brokers are reported as $16.8M, $11.1M, $12.8M from 2012 to 2014, respectively [82, 83]; other insurers have reported coverage limits of up to $200M-$300M [109]. We refer the interested reader to [10, 109, 2] for additional information on both the U.S. and the U.K. insurance markets, as well as common types of coverage offered through these policies, and the typical exclusions.

In the security literature, the study of cyber insurance both as a method for mitigating cyber security risks, and as an incentive mechanism for internalizing the externalities of security investments, has received considerable attention, see e.g., [53, 69, 70, 71, 116, 57, 76, 100, 75, 12]. This literature has mainly focused on one of the two market environments of competitive or monopolistic insurers when determining the power of cyber insurance in improving network security. On one hand, it can be shown that in competitive insurance markets, the introduction of insurance contracts not only fails to improve, but can further worsen network security relative to a no-insurance scenario [116, 100]. This is because contracts offered in such markets are optimal from the viewpoint of *individual* users, whereas socially

optimal contracts should be designed by keeping *social* welfare in mind. On the other hand, by engaging in premium discrimination, a monopolistic profit-neutral cyber insurer can induce socially optimal security investments in an interdependent system where security decisions are binary [57, 76, 100]. Therefore, a careful selection or regulation of the cyber insurance market is required to ensure its usability as an incentive mechanism for improved security. We further discuss these issues in Chapter 2.

Cyber insurance is affected by the classic insurance problems of adverse selection (higher risk users seek more protection) and moral hazard (users lower their investment in self-protection after being insured). Therefore, the insurance company needs to somehow mitigate the information asymmetry and calculate the premium fees with these considerations in mind. An example of such solutions is when an insurer chooses to monitor investments and/or inspect users' devices to prevent the moral hazard problem, specifying the terms of the contract accordingly to ensure appropriate levels of investment in self-protection [70]. Our work on the use of reputation mechanisms in security games in Chapter 5 can be viewed as a alternative solution to the problem of obtaining an assessment of users' security posture when designing cyber insurance contracts, without requiring a direct monitoring of users' investments.

In addition to the classic insurance problems, the design of cyber insurance contracts is further complicated by the risk interdependencies and the possibility of correlated damages in an interconnected system. Our findings on the design of incentive mechanisms for security games in Chapter 2 can be used as a guideline for devising premium discrimination schemes that consider the interdependent nature of the environment, and aim at incentivizing better security practices by the cyber insurance customers.

1.3.3 Security information sharing agreements and laws

In the U.S., improving information sharing is listed as one of President Obama's administration's priorities on cyber security, and is evidenced by its inclusion as one

of the key focus areas in the 2013 Executive Order 13636 on "Improving Critical Infrastructure Cybersecurity" [40], and as initiative #5 in the Comprehensive National Cybersecurity Initiative (CNCI) [26]. Most recently, during the first White House Summit on cybersecurity and consumer protection, President Obama signed Executive Order 13691 on "Promoting Private Sector Cybersecurity Information Sharing", encouraging companies to share cyber security information with one another and the federal government [41]. Following the executive order, the Department of Homeland Security (DHS) has started efforts to encourage the development of Information Sharing and Analysis Organizations (ISAOs) [35], as well as the Cyber Information Sharing and Collaboration Program (CISCP) in order to encourage Cooperative Research and Development Agreements. As of July 2015, 125 such agreements have been placed, with an additional 156 being negotiated [42].

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. Currently, most of the existing laws in the U.S. and the European Union require organizations only to report to an authority, with a few also mandating notification of the affected individuals, e.g., HIPAA for the health sector in the U.S. (see [73] for a summary of prominent U.S. and E.U. laws). However, motivated by the aforementioned trend in the newest initiatives in the U.S. (in particular, EO 13691), in this thesis we are primarily interested in information sharing agreements *among firms*, both with and without facilitation by an authority. Examples of existing agreements/organizations of this type include Information Sharing and Analysis Centers (ISACs), Information Sharing and Analysis Organizations (ISAOs), the United States Computer Emergency Readiness Team (US-CERT), and InfraGard.

From the theoretical perspective, a number of research papers have analyzed the welfare implications of information sharing agreements, as well as firms' incentives for adhering to these agreements, through the study of one-shot game-theoretic models; see e.g., [98, 73, 50, 47]. We will review this literature in more detail in Chapter 4, where we study these agreements in a repeated game framework.

1.4 Thesis outline

The main focus of this thesis is on using a game-theoretic approach to understand the theoretical underpinnings of users' incentives for exerting effort towards the provision of public goods, particularly cyber security.

Towards this end, Chapter 2 studies security games, with a focus on voluntary participation in incentive mechanisms. It proposes the notion of exit equilibrium to formalize the study of users' outside options in mechanisms for incentivizing the provision of non-excludable public goods. We use this notion to show a general negative result: there is no tax-based incentive mechanism that can implement the socially optimal solution, while guaranteeing voluntary participation and maintaining a weakly balanced budget, in all instances of security games, i.e., without prior knowledge of the network structure or users' preferences. This negative result is in sharp contrast with the performance guarantees of incentive mechanisms for provision of private and excludable public good provision problems, using which social optimality, voluntary participation, and weak budget balance, can be guaranteed simultaneously. It highlights the importance of accounting for users' outside options (in particular, their continued interaction with the system even after opting out) when incentivizing the provision of non-excludable goods. This negative result is then shown to extend to risk-averse users purchasing cyber insurance contracts, with important implications in the feasibility of using cyber insurance as a method for improving the state of cyber security. This chapter includes work that has appeared in [97, 93, 91, 89].

In light of this negative result, we pursue two possible alternatives. First, we can restrict attention to subclasses of the general model, for which the additional information on the network structure and/or users' preferences may aid the design of incentive mechanisms. Chapter 2.4.2 and Chapter 3 explore this alternative by restricting attention to a class of security games played on networks. Another possibility is to account for the repeated nature of users' interactions, in order to design inter-temporal incentives for cooperation. Chapter 4 pursues this direction for a particular type of security games, namely, security information sharing agreements.

Specifically, Chapter 3 studies the general framework of public good provision

games on networks, which include security games played on networks as a special case. These are games in which each user's payoff depends on a weighted sum of her neighbors' efforts, with weights given by those of the links of the network. In this chapter, we allow for both complements and substitutes, different strengths of interactions (weighted graphs), and unidirectional interactions (directed graphs). We are interested in the study of Nash equilibria, Pareto efficient effort profiles, and semi-cooperative equilibria (we define these as the effort profiles emerging when coalitions of users interact with one another). Our goal is to provide an understanding of how the aforementioned outcomes (i.e., the results of users' strategic interactions) are affected by the properties of the network. We will first identify necessary and sufficient conditions on the structure of the network for the uniqueness of Nash equilibria. We show that our finding unifies (and strengthens) existing results in the literature. We also identify conditions for the existence of Nash equilibria for the subclasses of games at the two extremes of our model, namely games of strategic complements and games of strategic substitutes. All identified conditions are based only on the network structure. We provide a graph-theoretical interpretation of users' efforts at Nash equilibria, Pareto efficient effort profiles, and semi-cooperative equilibria, by linking each user's decision to her centrality in the interaction network. Using this connection, we separate the effects of incoming and outgoing edges on users' efforts, and uncover an alternating effect over walks of different length in the network. This chapter includes work that has appeared in [96, 92].

The design of inter-temporal incentives (i.e., conditioning future cooperation on the history of past interactions) for disclosure in information sharing agreements is presented in Chapter 4. We propose a repeated game formulation of these agreements as repeated N-person prisoner's dilemma games, in order to understand firms' incentives for sharing their security information given the associated disclosure costs. Specifically, we show that a rating/assessment system can play a key role in enabling the design of appropriate incentives for supporting cooperation among firms. We further show that in the absence of a monitor, similar incentives can be designed if participating firms are provided with a communication platform, through which they can share their beliefs about others' adherence to the agreement. This chapter

includes work that has appeared in [94, 95].

A common assumption required in some parts of the preceding chapters is that the mechanism designer knows, or can estimate, users' efforts towards cyber security. For example, some incentive mechanisms require knowledge of users' levels of effort to determine appropriate taxes/rewards. A similar need arises when mitigating moral hazard and devising cyber insurance contracts with premium discrimination. Finally, in security information sharing agreements, there is a need for forming assessments of participants' adherence to the terms of the agreement. The aforementioned actions are commonly only known to the entity herself; an outside observer can in general only attain a noisy observation of these actions. Chapter 5 proposes a mechanism for crowdsourcing reputation (here, an estimate of users' investments or security related decisions) that allows the mechanism designer to improve her own prior assessment, without the need for monetary taxation/rewards, by crowdsourcing self-assessments and cross-observations from different users within the system. This chapter includes work that has appeared in [90, 88].

Chapter 6 concludes the dissertation with reflections and directions for future work.

1.5 Thesis contributions

The main contributions of this dissertation can be summarized as follows.

- Security as a non-excludable public good (Chapter 2)
  - It proposes the notion of *exit equilibrium* to describe strategic users' outside options from mechanisms for incentivizing the adoption of optimal security practices.
  - It shows *a general negative result* on simultaneously guaranteeing social optimality, voluntary participation, and weak budget balance, in all instances of security games.
  - By extending this negative result to risk-averse users, it highlights *the limitations of using cyber insurance* as an incentive for the adoption of

better security practices, which has been widely proposed in both theory and practice.

- Public good provision games on networks (Chapter 3)
  - It identifies the necessary and sufficient condition for *uniqueness* of Nash equilibria in public good provision games, in terms of the network structure. We show that previous results in the literature can be recovered as special cases of our result.
  - It further identifies the necessary and sufficient condition for the *existence* of Nash equilibria in two subclasses of our model, namely games with strategic substitutes and games with strategic complements, in terms of the network structure.
  - It presents a *graph theoretical characterization* of users' actions at different effort profiles, namely the Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria. Our characterization separates the effects of users' dependencies and influences on their efforts. It also uncovers an alternating effect over walks of different length in the network.
- Inter-temporal incentives in security information sharing agreements (Chapter 4)
  - It proposes the design of *inter-temporal incentives* for supporting cooperation on full disclosure in *security information sharing agreements*.
  - It illustrates *the role of a public rating/assessment system* in providing imperfect public monitoring, leading to coordination on cooperation in information sharing agreements.
  - It establishes the possibility of sustaining cooperative behavior in the absence of a public monitor, by introducing a platform for *communication* among firms through which they can exchange *their private beliefs* on others' adherence to the agreement.

- Crowdsourcing reputation in security games (Chapter 5)
  - It proposes *a simple reputation mechanism* that, by crowdsourcing assessments from a set of strategic users, is able to improve its own prior assessment of these users' reputation (security posture or related decisions), without the need for monetary incentives.

Chapter 2

Security as a Non-Excludable Public Good

2.1 Introduction

Despite advances in cyber-defense technologies, cyber attacks on organizations across all sectors remain rampant. From an economic perspective, it has been argued that this sub-optimal security status is due to lack of incentives for organizations to properly adopt existing software and best practices to improve their state of security [3, 54, 71]. In particular, as discussed in Chapter 1, security can be viewed as a non-excludable public good. This is because in a network of interdependent users, the expenditure in security measures by an entity affects not only herself, but also other users interacting with her. As a result, due to this public good nature, the optimal provision of security in a system of self-interested users is in general inefficient, and requires the design of appropriate incentive mechanisms.

To this end, the literature on security games has proposed incentive mechanisms for improving expenditures in cyber security; see Section 2.6 for related work. Similarly, our focus in the current chapter is on the use of monetary payments/rewards that achieve the following goals. First, the mechanism is designed to incentivize *socially optimal* security behavior, i.e., those minimizing the collective cost of security. In addition, it should ensure *voluntary participation* by all users. The importance of ensuring voluntary participation lies in the fact that, in general, the mechanism designer either lacks the authority, or is reluctant, to mandate user cooperation.

Finally, it should maintain a *weakly balanced budget*, so that the designer does not need to spend (potentially large amounts of) external resources to implement the mechanism.

This chapter studies such mechanisms with a focus on the issue of voluntary participation. It proposes the notion of exit equilibrium to describe users' ability to exercise their outside options from a proposed incentive mechanisms. We show a fundamental result that, due to the non-excludable nature of security, there exists no reliable mechanism which can incentivize socially optimal investments, while ensuring voluntary participation and maintaining a weakly balanced budget, for all instances of security games. This is in contrast to private goods and excludable public goods, for which tax-based incentive mechanisms such as the Externality and Pivotal (VCG) mechanisms can simultaneously guarantee social optimality, voluntary participation, and weak budget balance; see Section 2.6.1.

To further illustrate how the non-excludability of security leads to this result, we analyze the performance of the Pivotal and Externality mechanisms in a class of security games played on networks, referred to as *weighted effort* games. Through analysis and numerical simulations, we identify the effects of several features of the problem environment, including multiplicity of exit equilibria, and users' self-dependence levels, on the performance of these mechanisms. In addition, we identify two classes of users in these games: *main investors*, who receive a reward in return for improving their investment levels (from which themselves, as well as other users benefit), and *free-riders*, who pay a tax to benefit from a more secure environment. We highlight how voluntary participation constraints may fail to hold for users from either class. In other words, either free-riders or main investors may decide to opt out of a proposed mechanism.

We then extend the negative result by considering several variations of the main model. We illustrate how, given a mechanism, stable coalitions of participating users may emerge, leading to an improved, yet sub-optimal security status. We discuss the idea of bundling the role of the mechanism designer and a security vendor; the intent is to allow the vendor to leverage the additional revenue from the increased sale of security products to (partially) cover the deficit generated through the in-

centive mechanism. Finally, we extend the impossibility result to risk-averse users (as opposed to risk-neutral users in the basic model) who are offered cyber insurance contracts, and discuss the implications of this negative result on the viability of using cyber insurance for improving the state of cyber security.

2.1.1 Chapter contributions

The main contributions of this chapter can be summarized as follows:

- This chapter proposes the notion of exit equilibrium to describe strategic users' outside options from mechanisms for incentivizing the adoption of optimal security practices. This work hence formalizes the study of voluntary participation in security games, in which the assumption of compulsory compliance is commonly adopted.

- It shows the fundamental impossibility of simultaneously guaranteeing social optimality, voluntary participation, and weak budget balance in all instances of security games. By comparing this finding to existing possibility results (see Section 2.6), this work highlights the crucial effect of users' outside options on the design of any mechanism for improving users' security behavior. The insights are also applicable to other problems concerning the provision of non-excludable public goods over social and economic networks; see Section 2.6.

- We extend this impossibility result to risk-averse users who are offered cyber insurance contracts. This finding highlights the limitations of using cyber insurance as an incentive for the adoption of better security practices, which has been widely proposed in theory [76, 75, 100], as well as in practice, e.g., by the Department of Homeland Security [34].

- By finding restricted families of positive instances, we identify features of an environment that can affect the performance of incentive mechanisms for security games. We further analyze the role of a security vendor in extending the space of positive instances. These findings can guide the selection of a mechanism given additional information about the problem environment.

2.1.2 Chapter organization

The rest of this chapter is organized as follows. Section 2.2 presents the model for security games, followed by the impossibility result in Section 2.3. Section 2.4 illustrates this result and identifies restricted families of positive instances by analyzing the Pivotal and Externality mechanisms, and applying them to the weighted effort model. Several extensions of the impossibility result are presented in Section 2.5, followed by related work in Section 2.6. Section 2.7 concludes the chapter.

2.2 Security games

2.2.1 Model

Consider a network of N interconnected and interdependent users. These users can be the operators of computers on a network, different divisions within a larger organization, or various sectors of an economy. Each user has an initial wealth W_i, and is subject to suffering a loss of $0 < L_i \leq W_i$ if her security is compromised. To decrease the probability of a successful attack, each user can choose a level of *effort* or *investment* in security $x_i \in \mathbb{R}_{\geq 0}$. User i incurs a cost of $h_i(x_i)$ when exerting effort x_i, where $h_i(\cdot) : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is referred to as the *cost function*.

We assume that the effort x_i not only protects the user herself, but further benefits other users in the system. This is because a better protected user generates positive externalities to (some or all) other users by decreasing the probability of contagious infections or attacks using existing connections. Denote the vector of all users' efforts by $\mathbf{x} := \{x_1, \ldots, x_N\}$. The probability of a successful attack on user i, at a vector of efforts $\mathbf{x}$, is determined by the *risk function* $f_i(\mathbf{x}) : \mathbb{R}^N_{\geq 0} \to [0, 1]$. The dependence of user i's risk, $f_i(\cdot)$, on other users' efforts, $\mathbf{x}_{-i} := \{x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_N\}$, captures the interdependence of users' security.

The utility of user i is therefore given by

$$u_i(\mathbf{x}) = W_i - L_i f_i(\mathbf{x}) - h_i(x_i) \ . \tag{2.1}$$

We refer to the full information, one shot game among the N rational users, choosing actions $x_i \geq 0$, with utility functions given in (2.1), as the *security game*. We make the following assumptions on the risk and cost functions.

Assumption 2.1. *For all users i, $f_i(\cdot)$ is continuous, differentiable, decreasing (strictly decreasing in x_i), and strictly convex, in each argument x_j.*

Intuitively, the decreasing nature of this function in arguments $x_j, j \neq i$, models the positive externality of users' security decisions on one another. The convexity on the other hand implies that the effectiveness of security measures in preventing attacks (or the marginal benefit) is overall decreasing, as none of the available security measures can prevent all possible attacks.

Assumption 2.2. *For all users i, $h_i(\cdot)$ is continuous, differentiable, strictly increasing, and convex.*

Intuitively, the assumption of convexity entails that while implementing basic security measures is relatively cheap (e.g., limiting administrative privileges), protection may become increasingly costly as its effectiveness increases (e.g., implementing intrusion detection systems).

2.2.2 Social optimality and exit equilibria

Security games have been extensively studied, see [71, 81] for surveys. The most commonly studied aspect of these games is their *Nash equilibrium*, i.e, the vector of investments in security that emerge when each user chooses an optimal level of effort accounting for her costs and benefits, while also best-responding to the investments of other users. Formally, at a Nash equilibrium $\tilde{\mathbf{x}}$, for all users i,

$$\tilde{x}_i = \arg\max_{x \geq 0} \; u_i(x, \tilde{\mathbf{x}}_{-i}) \; .$$

Nevertheless, it is well known in the economics literature that, as users do not account for the externality of their decisions on one another, this effort profile is sub-optimal from a social welfare perspective. Formally, we define the welfare maximizing profiles as follows.

Definition 2.1. *A* socially optimal *profile* $\mathbf{x}^*$*, for users with utility functions* (2.1)*, is given by*

$$\mathbf{x}^* = \arg\max_{\mathbf{x}\succeq 0} \sum_{j=1}^{N} u_j(\mathbf{x}) \ . \tag{2.2}$$

We are interested in incentive mechanisms that use appropriately designed monetary taxation/rewards to implement the socially optimal effort profile. Formally, the mechanism assesses a tax t_i to each participating user i; this tax may be positive, negative, or zero, indicating payments, rewards, or no transaction, respectively. We assume that users' utilities are quasi-linear, i.e., linear in the tax term (see Section 2.5.2 for an extension of our results to risk-averse users). Therefore, the *total utility* of a user i when she is assigned a tax t_i is given by

$$\mathrm{v}_i(\mathbf{x}, t_i) := u_i(\mathbf{x}) - t_i \ , \tag{2.3}$$

where the tax t_i can in general be a function of the security investment profile $\mathbf{x}$.

In addition to inducing socially optimal behavior, the mechanism designer selecting the tax terms t_i wishes to satisfy two goals. On one hand, the designer attempts to ensure that the implementation of the mechanism does not require spending (a potentially large amount of) external resources.[1] Formally,

Definition 2.2. *An incentive mechanism assigning taxes* $\{t_i\}_{i=1,\ldots,N}$*, satisfies* weak budget balance (WBB) *if*

$$\sum_{i=1}^{N} t_i \geq 0 \ .$$

More importantly, for a successful implementation of an incentive mechanism, the designer has to ensure that users' *voluntary participation (VP)* constraints are satisfied, as the designer generally lacks the authority to enforce cooperation. Note the deliberate choice of the term voluntary participation as opposed to the commonly studied *individual rationality (IR)* constraint. Individual rationality is often used to

[1] At least at equilibrium, but ideally, both on and off equilibrium.

refer to the requirement that a user prefers participation in a proposed mechanism to an outcome in which she opts out and receives no allocation of the (private or excludable public) good at all. In contrast, we define voluntary participation constraints as those ensuring that a user prefers implementing the socially optimal outcome while being assigned a tax t_i, to the outcome attained had she unilaterally opted out, but in which she could still benefit from the (non-excludable) public good. Such distinction is important with non-excludable public good such as security, as a user can still benefit from the externalities generated by the participating users, and also potentially continue contributing to the production of the public good, even when opting out herself. This is in contrast to games with excludable public goods, where voluntary participation and individual rationality become equivalent.

Therefore, to formally state users' voluntary participation, we propose the notion of *exit equilibrium (EE)*: consider an *outlier*, who is contemplating unilaterally opting out of a proposed incentive mechanism. By the assumption of full information, the remaining participating users, who are choosing a welfare maximizing solution for their $(N-1)$-user system, will have the ability to predict the best-response of the outlier to their collective action, and thus choose their investments accordingly. As a result, the equilibrium investment profile where user i opts out is itself a Nash equilibrium (for the game between this outlier and the coalition of $N-1$ participating users). Formally, we define exit equilibria as follows.

Definition 2.3. *Assume a user i opts out of a given incentive mechanism, while the remaining $N-1$ users continue participating (i.e., implement the welfare maximizing solution to their $(N-1)$-user system). Then, the* exit equilibrium $\hat{\mathbf{x}}^{\mathbf{i}}$ *for outlier i is given by*

$$\begin{aligned} \hat{\mathbf{x}}^i_{-i} &= \arg\max_{\mathbf{x}_{-i}\succeq 0} \ \sum_{j\neq i} u_j(\mathbf{x}_{-i}, \hat{x}^i_i) \ , \\ \hat{x}^i_i &= \arg\max_{x_i\geq 0} \quad u_i(\hat{\mathbf{x}}^i_{-i}, x_i) \ . \end{aligned} \tag{2.4}$$

Using the definition of exit equilibrium, we can formally state users' voluntary participation constraints.

Definition 2.4. *A user i's* voluntary participation *constraint in an incentive mechanism, when assigned a tax t_i, is satisfied if*

$$v_i(\mathbf{x}^*, t_i) \geq u_i(\hat{\mathbf{x}}^i) \ .$$

We would like to highlight two important considerations in studying exit equilibria. First, note that the study of exit equilibria to understand users' unilateral deviations from socially optimal investment profiles is similar to the study of users' deviation from Nash equilibria: neither concept precludes the possibility that coalitions of deviating users can break the equilibrium. Nevertheless, it is necessary (although indeed not sufficient) for a mechanism to be resilient against unilateral exit strategies in order to incentivize cooperation. We therefore only focus on unilateral exit strategies. In fact, as shown in Section 2.3, there exists no incentive mechanism that can incentivize socially optimal efforts, while guaranteeing weak budget balance and voluntary participation, even against unilateral deviations, much less against higher order coalitions. Secondly, we also note that the proposed exit equilibrium is only an equilibrium under the assumption that the $N-1$ remaining users are cooperating in the mechanism; it is itself not necessarily stable as any of the remaining users may also prefer to opt out. In Section 2.5.1, we extend the definition of exit equilibrium to allow for multiple outliers, and present instances in which the resulting exit equilibrium yields a stable coalition of participating users, as well as instances in which no stable exit equilibrium (other than the degenerate case of Nash equilibrium) exists.

2.3 An impossibility result

In this section, we prove the following result.

Proposition 2.1. *There exists no tax-based incentive mechanism which can implement the socially optimal solution, while guaranteeing weak budget balance and voluntary participation simultaneously, in all instances of security games.*

We prove this impossibility through two families of counter-examples. The first

counter-example limits the network structure by considering a star topology. The second family limits users' preferences to the particular class of weakest link risk functions.

Counter-example I: The star topology

Assume some tax-based incentive mechanism $\mathcal{M}$ is proposed for security games. Consider N users connected through the star topology with user 1 as the center, such that the security decisions of the center affect all leaves, but each leaf's investment affects only herself and the center. Formally, set $W_i = W,\ L_i = L, \forall i$, and let the utility function of the center be given by

$$u_1(\mathbf{x}) = W - Lf(x_1 + \sum_{j=2}^{N} x_j) - cx_1 \ ,$$

and that of all leaves $j \in \{2, \ldots, N\}$ by

$$u_j(\mathbf{x}) = W - Lf(x_1 + x_j) - cx_j \ .$$

Here, $f(\cdot)$ is any function satisfying the assumptions in Section 2.2. The investment cost functions $h_i(\cdot)$ are linear, with the same unit investment cost c for all users.

We first solve (2.2) to find the socially optimal investment profile $\mathbf{x}^*$. It is easy to see that for this graph, only the center will be investing in security, while all leaves rely on the resulting externality. This socially optimal investment profile $\mathbf{x}^*$ is given by

$$\frac{\partial f}{\partial x}(x_1^*) = -\frac{c}{LN}, \quad x_j^* = 0, \forall j = 2, \ldots, N \ .$$

Now, assume the center user is considering stepping out of the mechanism. To find the exit equilibrium profile $\hat{\mathbf{x}}^1$ resulting from this unilateral deviation, first note that the leaves' security decisions will not affect one another, so that the socially optimal investment profile for the $N-1$ leaves is the same as their myopic decisions. User 1 will also be choosing her individually optimal level of investment. Therefore,

using (2.4), the exit equilibrium $\hat{\mathbf{x}}^1$ will satisfy

$$\frac{\partial f}{\partial x}(\hat{x}_1^1 + \sum_{j=2}^{N} \hat{x}_j^1) + \frac{c}{L} \geq 0 \ ,$$
$$\frac{\partial f}{\partial x}(\hat{x}_1^1 + \hat{x}_j^1) + \frac{c}{L} \geq 0, \quad \forall j = 2, \ldots, N \ .$$

We conclude that an exit equilibrium when user 1 unilaterally leaves the mechanism is such that

$$\frac{\partial f}{\partial x}(\hat{x}_1^1) = -\frac{c}{L}, \quad \hat{x}_j^1 = 0, \ \forall j = 2, \ldots, N \ .$$

Finally, if any leaf user $j \in \{2, \ldots, N\}$ leaves the mechanism, the exit equilibrium $\hat{\mathbf{x}}^j$ will satisfy

$$\frac{\partial f}{\partial x}(\hat{x}_1^j) = -\frac{c}{L(N-1)}, \quad \hat{x}_k^j = 0, \ \forall k = 2, \ldots, N \ .$$

We now use the socially optimal investment profile and the exit equilibria to evaluate voluntary participation and weak budget balance for mechanism $\mathcal{M}$. Assume $\mathcal{M}$ assigns a tax t_i^* to a participating user i. Then, voluntary participation will hold if and only if $v_i(\mathbf{x}^*, t_i^*) \geq u_i(\hat{\mathbf{x}}^i), \forall i$, which reduces to

$$t_1^* \leq L\left(f(\hat{x}_1^1) - f(x_1^*)\right) + c(\hat{x}_1^1 - x_1^*) \ ,$$
$$t_j^* \leq L\left(f(\hat{x}_1^j) - f(x_1^*)\right) , \ \forall j \in \{2, \ldots, N\} \ .$$

The sum of these taxes is thus bounded by

$$\sum_{i=1}^{N} t_i^* \leq L\left(f(\hat{x}_1^1) - f(x_1^*)\right) + c(\hat{x}_1^1 - x_1^*) + L(N-1)(f(\hat{x}_1^j) - f(x_1^*))$$

However, the above sum can be negative, e.g., when $f(z) = \exp(-z)$ or $f(z) = \frac{1}{z}$, indicating that weak budget balance will fail regardless of how the taxes are determined in the mechanism $\mathcal{M}$. For example, replacing $f(z) = \exp(-z)$, the

upper bound can be simplified as follows:

$$\begin{aligned}\sum_{i=1}^{N} t_i^* &\leq L\left(f(\hat{x}_1^1) - f(x_1^*)\right) + c(\hat{x}_1^1 - x_1^*) + L(N-1)(f(\hat{x}_1^j) - f(x_1^*)) \\ &= L(\frac{c}{L} - \frac{c}{LN}) + c(\ln\frac{L}{c} - \ln\frac{LN}{c}) + L(N-1)(\frac{c}{L(N-1)} - \frac{c}{LN}) \\ &= c(1 - \ln N)\ .\end{aligned}$$

The above upper bound is negative for any $N \geq 3$, indicating that the sum of taxes for any mechanism is necessarily negative as well.

Intuition: the failure of any mechanism $\mathcal{M}$ in guaranteeing social optimality, voluntary participation, and weak budget balance in this topology, is due to the fact that the center node (a main investor) asks for a reward that the leaves (free-riders) are not willing to subsidize. Note that if the users were facing a choice between the center investing the socially optimal level, and staying at the Nash equilibrium, this problem would have not arisen, and it would be possible to guarantee all three properties. Nevertheless, as outlier leaf nodes can still enjoy a lower level of security subsidized by other participating leaves, their willingness to pay is limited, consequently not financing the reward requested by the center, leading to the negative result of Proposition 2.1.

Counter-example II: Weakest link games

We next consider security games with a family of risk functions that approximate the *weakest link* risks $f_i(\mathbf{x}) = \exp(-\min_j x_j)$ [122, 71]. In particular, we use the approximation $\min_j x_j \approx -\frac{1}{\gamma}\log\sum_j \exp(-\gamma x_j)$, where the accuracy of the approximation is increasing in the constant $\gamma > 0$. User i's utility function is given by

$$u_i(\mathbf{x}) = W - L(\sum_{j=1}^{N}\exp(-\gamma x_j))^{1/\gamma} - cx_i\ ,$$

where investment cost functions $h_i(\cdot)$ are assumed to be linear, and users are homogeneous, with the same initial wealth W, loss L, and unit investment cost c.

In this game, the socially optimal investment profile $\mathbf{x}^*$ can be found by solving the first order conditions of (2.2), which are given by

$$N \exp(-\gamma x_i^*)(\sum_{j=1}^{N} \exp(-\gamma x_j^*))^{\frac{1}{\gamma}-1} = \frac{c}{L} \ , \forall i.$$

By symmetry, all users will be exerting the same socially optimal level of effort

$$x_i^* = \frac{1}{\gamma} \ln \frac{N}{(\frac{c}{L})^\gamma} \ , \forall i \ .$$

Next, assume a user i unilaterally opts out of the mechanism, while the remaining users continue participating. The exit equilibrium profile $\hat{\mathbf{x}}^i$ can be determined using the first order conditions on (2.4), leading to

$$\begin{aligned}(N-1) \exp(-\gamma \hat{x}_j^i)(\sum_{k \neq i} \exp(-\gamma \hat{x}_k^i) + \exp(-\gamma \hat{x}_i^i))^{\frac{1}{\gamma}-1} &= \frac{c}{L} \ , \\ \exp(-\gamma \hat{x}_i^i)(\sum_{k \neq i} \exp(-\gamma \hat{x}_k^i) + \exp(-\gamma \hat{x}_i^i))^{\frac{1}{\gamma}-1} &= \frac{c}{L} \ .\end{aligned}$$

Solving the above, we get

$$\begin{aligned}\hat{x}_i^i &= \frac{1}{\gamma} \ln \frac{2^{1-\gamma}}{(\frac{c}{L})^\gamma} \ , \\ \hat{x}_j^i &= \frac{1}{\gamma} \ln \frac{(N-1)2^{1-\gamma}}{(\frac{c}{L})^\gamma} \ , \forall j \neq i \ .\end{aligned}$$

Assume some tax-based incentive mechanism $\mathcal{M}$ is proposed in this game. We can use the socially optimal investment profile and the exit equilibria to analyze users' participation incentives in $\mathcal{M}$, as well as the budget balance conditions. Denote by t_i^* the tax assigned to user i by $\mathcal{M}$.

A user i's utilities when participating and staying out are given by

$$\begin{aligned} v_i(\mathbf{x}^*, t_i^*) &= W - L(N\exp(-\gamma x_i^*))^{\frac{1}{\gamma}} - cx_i^* - t_i^* \\ &= W - c(1 + x_i^*) - t_i^* \ . \\ u_i(\hat{\mathbf{x}}^i) &= W - L(\exp(-\gamma \hat{x}_i^i) + (N-1)\exp(-\gamma \hat{x}_j^i))^{\frac{1}{\gamma}} - c\hat{x}_i^i \\ &= W - c(2 + \hat{x}_i^i) \ . \end{aligned}$$

The voluntary participation condition for a user i will hold if and only if $v_i(\mathbf{x}^*, t_i^*) \geq u_i(\hat{\mathbf{x}}^i)$, which reduces to

$$c(1 + x_i^*) + t_i^* \leq c(2 + \hat{x}_i^i) \Leftrightarrow t_i^* \leq c(1 + \frac{1}{\gamma}\ln\frac{2^{1-\gamma}}{N}) \ . \tag{2.5}$$

On the other hand, for weak budget balance to hold, we need $\sum_i t_i^* \geq 0$. Nevertheless, by (2.5), we have

$$\sum_i t_i^* \leq cN(1 + \frac{1}{\gamma}\ln\frac{2^{1-\gamma}}{N}) \ .$$

It is easy to see that given γ and for any $N > e^\gamma 2^{1-\gamma}$, the above sum will always be negative, indicating a budget deficit for a general mechanism $\mathcal{M}$, regardless of how taxes are determined.

Intuition: note that the lack of any mechanism $\mathcal{M}$ occurs only when there is a sufficient number of players (given a finite γ). This is because with a sufficient number of participating users, the externality available to an outlier is high enough to dissuade her from participating. It is also interesting to point out that outside this region (i.e., $N \leq e^\gamma 2^{1-\gamma}$, the number of users is sufficiently small), we in fact have a positive instance, in which the Externality mechanism introduced in Section 2.4.1 can guarantee social optimality, budget balance, and voluntary participation.

A note on the nature of the impossibility result

We close this section by noting the implications of proving our impossibility result on a simultaneous guarantee of social optimality, voluntary participation, and weak budget balance, through counter-examples. We have shown that without prior knowledge of the graph structure or users' preferences, it is not possible for a designer to propose a *reliable* mechanism, i.e., one which can promise to achieve social optimality, voluntary participation, and weak budget balance, regardless of the realizations of utilities. This should be contrasted with environments with the same utility functions and information constraints, but excludable public goods, in which there exist reliable mechanisms to guarantee all three properties simultaneously; see Section 2.6.1. Nevertheless, as also suggested by the counter-example based on the weakest link games, it may still be possible to design reliable mechanisms for non-excludable public goods under a restricted problem space. With this in mind, we next analyze the class of weighted effort models, and aim to identify such positive instances, as well as the intuition behind the existence of each instance.

2.4 A tale of two mechanisms: Analysis of existing incentive schemes

In light of the negative result of Proposition 2.1, in this section we set out to better understand the performance of existing incentive mechanisms in security games, and identify features of the problem environment that affect the properties attainable through given mechanisms. We further find positive instances (in a restricted utility space) for which these existing mechanisms can guarantee all three properties of social optimality, voluntary participation, and weak budget balance. Specifically, we analyze the performance of the Pivotal and Externality mechanisms within the restricted class of security games played on networks, referred to as weighted effort security games.

2.4.1 The Pivotal and Externality mechanisms

Throughout this section, we will be studying the performance of two well-known tax-based incentive mechanisms, namely the Pivotal (VCG) and Externality mechanisms. We chose these mechanisms as they have been shown to simultaneously guarantee the achievement of social optimality, weak budget balance, and voluntary participation, in games of provision of *excludable* public goods. Our goal is hence to illustrate their inefficiencies in the provision of *non-excludable* public goods.

2.4.1.1 The Pivotal mechanism

Groves mechanisms [84, 103], also commonly known as Vickrey-Clarke-Groves (VCG) mechanisms, refer to a family of mechanisms in which, through the appropriate design of taxes for users with quasi-linear utilities, a mechanism designer can incentivize users to reveal their true preferences in dominant strategies, thus implementing the socially optimal solution. One particular instance of these mechanisms, the *Pivotal* (or Clarke) mechanism, has been shown to further satisfy the participation constraints and achieve weak budget balance in many private and public good games [103, 25, 56]; however, this is not necessarily the case in security games. The taxes in the Pivotal mechanism for security games are given by

$$t_i^P = \sum_{j \neq i} u_j(\hat{\mathbf{x}}^i_{-i}, \hat{x}^i_i) - \sum_{j \neq i} u_j(\mathbf{x}^*_{-i}, x^*_i) \ , \tag{2.6}$$

where $u_i(\mathbf{x})$ is user i's utility function, $\mathbf{x}^* = (\mathbf{x}^*_{-i}, x^*_i)$ is the socially optimal solution, and $\hat{\mathbf{x}}^i = (\hat{\mathbf{x}}^i_{-i}, \hat{x}^i_i)$ is the exit equilibrium under user i's unilateral deviation. In Appendix A, we show that the taxes in (2.6) incentivize users' voluntary participation and attain the socially optimal solution. However, these taxes may generate a budget deficit for the designer.

2.4.1.2 The Externality mechanism

We next introduce the Externality mechanism adapted from the work of Hurwicz in [59]. A main design goal of this mechanism is to guarantee a complete redistri-

bution of taxes, i.e., strong budget balance. This mechanism has been adapted in [115], where it is shown to achieve social optimality, guarantee voluntary participation, and maintain a balanced budget, in allocation of power in cellular networks (an excludable public good). However, this is again not the case in security games. The tax terms t_i^E at the equilibrium of the Externality mechanism in security games are given by

$$t_i^E(\mathbf{x}^*) = -\sum_{j=1}^{N} x_j^* L_i \frac{\partial f_i}{\partial x_j}(\mathbf{x}^*) - x_i^* \frac{\partial h_i}{\partial x_i}(x_i^*) \ . \tag{2.7}$$

The interpretation is that by implementing this mechanism, each user i will be financing part of user $j \neq i$'s reimbursement. According to (2.7), this amount is proportional to the positive externality of j's investment on user i's utility. In Appendix B, we show that the taxes in (2.7) attain the socially optimal solution and lead to a (strongly) balanced budget. However, they may fail to satisfy users' voluntary participation constraints in security games.

2.4.2 Weighted effort security games

The gap between the Nash equilibrium and the socially optimal investment profile of a security game, as well as users' participation incentives and possible budget imbalances, are dependent on the specifics of users' utility functions defined in (2.1). In particular, the risk function $f_i(\cdot)$ can model the types of connection and extent of interdependencies among users. Examples of existing security interdependence models include the *total effort*, *weakest link*, and *best shot* models considered in the seminal work of Varian [122], as well as the *weakest target* models studied in [54], the *effective investment* and *bad traffic* models in [64], and the *linear influence network* games in [85].

Here, we take the special case of *weighted effort* models[2], with exponential risks

[2]The weighted effort game is an instance of the public good provision games studied in Chapter 3. We use the term "weighted effort" to highlight the connection to the total effort model in the seminal work of Varian on security games [122].

and linear investment cost functions. This is a class of security games played on networks, in which the weights on the links among users determine the strength of their interdependencies. Formally, the total utility of user i is given by

$$\mathrm{v}_i(\mathbf{x}, t_i) = W_i - L_i \exp(-\sum_{j=1}^{N} a_{ij} x_j) - c_i x_i - t_i \ . \tag{2.8}$$

For simplicity, we assume $W_i = W$, $L_i = L = 1$, and $c_i = c$, for all i. The coefficients $a_{ij} \geq 0$ determine the dependence of user i's risk on user j's action. Consequently, user i's risk is dependent on a weighted sum of all users' efforts. We define the *dependence matrix* containing these coefficients as

$$A := \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{pmatrix} .$$

We isolate the effect of different features of the model on the performance of the two incentive mechanisms, by focusing on the following two sub-classes of this model:[3]

1. Varying users' self-dependence:

$$A = \begin{pmatrix} a & 1 & \cdots & 1 \\ 1 & a & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & a \end{pmatrix} , \tag{2.9}$$

for both $a > 1$ and $a < 1$.

2. Making all users increasingly dependent on a single user:

$$A = \begin{pmatrix} a & 1 & \cdots & 1 \\ a & 1 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ a & 1 & \cdots & 1 \end{pmatrix} , \tag{2.10}$$

for $a > 1$.

We present numerical results and intuitive interpretation for each of the above

[3]We refer the interested reader to [93] for an additional subclass, where we consider the effects of diversity by breaking users into two groups of self-dependent and reliant users.

Table 2.1: Effect of self-dependence in weighted effort security games

Parameter Conditions	Exit Equilibrium	VP in Externality	WBB in Pivotal
$a > 1$ with N and c s.t. $(1 + \frac{N-2}{a})^{N-1} > (\frac{a}{c})^{a-1}$	**CASE** α : $\hat{x}_i^i = 0, \hat{x}_j^i > 0$	No	No
$a > 1$ with N and c s.t. $(1 + \frac{N-2}{a})^{N-1} < (\frac{a}{c})^{a-1}$	**CASE** β : $\hat{x}_i^i > 0, \hat{x}_j^i > 0$	No	No
$a < 1$ with N and c s.t. $(1 + \frac{N-2}{a})^{a} > (\frac{a}{c})^{1-a}$	**CASE** γ : $\hat{x}_i^i = 0, \hat{x}_j^i > 0$	No	No
$a < 1$ with N and c s.t. $(1 + \frac{N-2}{a})^{a} < (\frac{a}{c})^{1-a}$	**CASE** γ : $\hat{x}_i^i = 0, \hat{x}_j^i > 0$ **CASE** ω : $\hat{x}_i^i > 0, \hat{x}_j^i = 0$ **CASE** ζ : $\hat{x}_i^i > 0, \hat{x}_j^i > 0$	Yes (iff ω or ζ)	Yes (iff ω or ζ)

scenarios; formal analysis is given in Appendices C and D.

2.4.2.1 Effects of self-dependence

Consider a network of N users, with the dependence matrix given by (2.9), and total utility functions[4]

$$\mathrm{v}_i(\mathbf{x}, t_i) = W - \exp(-ax_i - \sum_{j \neq i} x_j) - cx_i - t_i \ .$$

The following theorem characterizes the possible exit equilibria of this game under different parameter conditions, as well as whether the voluntary participation conditions are satisfied under the Externality mechanism, and whether the Pivotal mechanism can operate without a budget deficit. The results are summarized in Table 2.1.

Theorem 2.1. *For the weighted effort security game described by the dependence matrix* (2.9)*:*

(i) There exist five possible exit equilibria (cases $\alpha, \beta, \gamma, \omega,$ and ζ, summarized in Table 2.1) depending on the values of the number of players N, self-dependence a, and cost of investment c. In particular, note the multiplicity of exit equilibria

[4]We assume $c < a$, so as to ensure the existence of non-zero equilibria, i.e., at least one user exerts non-zero effort at any equilibrium of the game.

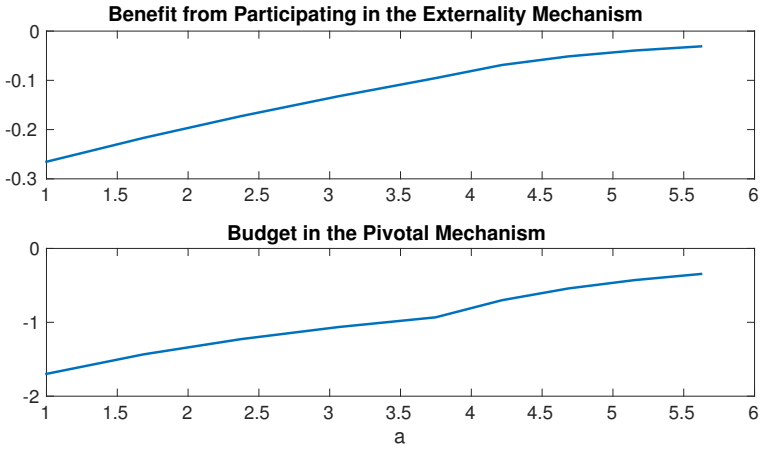

Figure 2.1: Increasing self-dependence in weighted effort games

under the parameter conditions of the last row in Table 2.1: either γ, ω, or ζ may be realized in such instances.

(ii) Either of the Externality or Pivotal mechanisms can guarantee social optimality, voluntary participation, and weak budget balance, if and only if the realized exit equilibrium is ω or ζ of Table 2.1.

The proof, including formal derivations of the exit equilibria, as well as the analysis of the Pivotal and Externality mechanisms under each exit equilibrium, are presented in Appendix C.

Using numerical simulations, we further examine the effect of changing a on the mechanisms' performance. In particular, we plot the sum of all taxes, $\sum_i t_i^P$, in the Pivotal mechanism. For the Externality mechanism, we plot $\mathrm{v}_i(\mathbf{x}^*, t_i^E) - u_i(\hat{\mathbf{x}}^i)$ per user i, i.e., the benefit of participation (in terms of increase in payoff) for that user. We set $N = 6$ and $c = 1$. We then increase a, starting from $a = 1$, hence moving gradually from the exit equilibrium in [Case α] to [Case β]. Intuitively, by increasing users' self-dependence, a unit of investment becomes more effective for the user. As a result, we move towards an exit equilibrium in which outliers exert non-zero effort. Figure 2.1 illustrates the results. From our analysis and simulations, we make the following observations:

Higher self-dependence improves performance of mechanisms: from Fig. 2.1 we observe that (as predicted by the analysis) the Pivotal mechanism will always carry a deficit, while the Externality mechanism will always fail to guarantee voluntary participation. Nevertheless, as self-dependence increases, the performance of both mechanisms improves. This is because higher self-dependence (equivalently, lower interdependence) leads to closer to optimal investments by individual users in their exit equilibrium. Such users require smaller incentives to move to the optimal state, hence the reduced budget deficit of the Pivotal mechanism, and smaller participation costs in the Externality mechanism.

Coordinating on the least beneficial exit equilibrium for the outlier: from Table 2.1, we observe that if selection among multiple exit equilibria is possible, the Pivotal and Externality mechanism can simultaneously guarantee social optimality, voluntary participation, and weak budget balance under the less beneficial exit equilibrium. A less beneficial equilibrium can be one that requires a free-rider to become an investor when leaving the mechanism, or one that requires an investor to continue exerting effort when out (although possibly at a lower level). This can be seen by comparing Cases ω and ζ (in which outliers become the main investors or have to continue exerting effort when out, respectively) with Case γ (in which outliers become free-riders).

An exchange of favors: it is also interesting to highlight another feature of the positive instances of Cases ω and ζ of Table 2.1: as users are mainly dependent on others' investments under these parameter conditions ($a < 1$), the incentive mechanisms can facilitate coordination among them, so that each will increase their investments in return for improved investments by others.

2.4.2.2 Effects of a dominant user

Consider a collection of N users, with dependence matrix given by (2.10), and total utility functions

$$\mathrm{v}_i(\mathbf{x}, t_i) = W - \exp(-ax_1 - \sum_{j=2}^{N} x_j) - cx_i - t_i \ ,$$

Table 2.2: Effects of a single dominant user in weighted effort security games

Parameter Conditions	Exit Equilibrium	VP in Externality	WBB in Pivotal
$a < N-1$	**CASE** α : $\hat{x}_1^1 = 0, \hat{x}_j^1 > 0, \forall j \neq 1$ $\hat{x}_1^i > 0, \hat{x}_j^i = 0, \forall i, j \neq 1$	No	No
$a > N-1$	**CASE** β : $\hat{x}_1^1 > 0, \hat{x}_j^1 = 0, \forall j \neq 1$ $\hat{x}_1^i > 0, \hat{x}_j^i = 0, \forall i, j \neq 1$	No	No

where $c < 1 < a$, and user 1 is the dominant user. In Appendix D, we show that in a socially optimal profile, as well as for exit equilibria of non-dominant users, only user 1 will be exerting effort. When the dominant user opts out of the mechanism, however, she may become either a main investor or free-rider, depending on the problem parameters.

The following theorem characterizes the possible exit equilibria and parameter conditions for which each is possible, as well as the performance of both mechanisms. The results are summarized in Table 2.2.

Theorem 2.2. *For the weighted effort security game described by the dependence matrix* (2.10)*:*

(i) There exist two possible exit equilibria (cases α and β, summarized in Table 2.2) depending on the values of the number of players N and dependence on the dominant user a.

(ii) Neither of the Externality or Pivotal mechanisms can guarantee social optimality, voluntary participation, and weak budget balance, regardless of the exit equilibrium.

The proof is presented in Appendix D. Using numerical simulations, we further illustrate the effect of increasing a, the dependence on the dominant user, on users' benefits from participating in the Externality mechanism (i.e., $\mathrm{v}_i(\mathbf{x}, t_i^E) - u_i(\hat{\mathbf{x}}^i)$), as well as the budget of the Pivotal mechanism (i.e., $\sum_i t_i^P$). We set $N = 10$, $c = 0.45$, and $a \in [1, 15]$. As a increases, the dominant user's exit equilibrium switches from free-riding to investing when opting out.

We make the following observation based on this analysis:

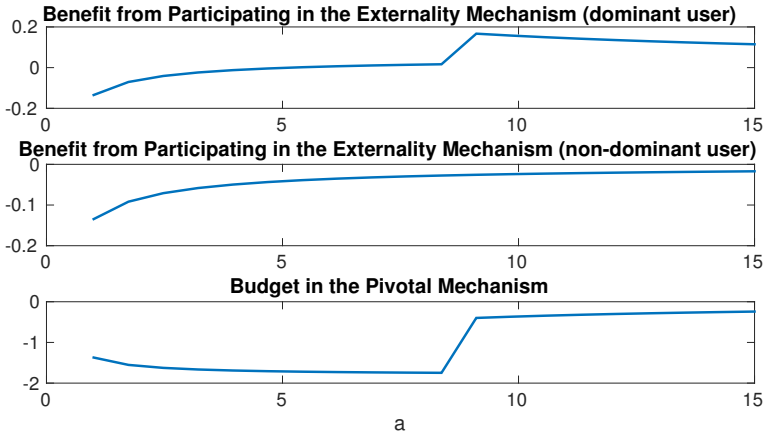

Figure 2.2: Increasing dependence on a single dominant user in weighted effort games

Either main investors or free-riders may opt out: Through our analysis, and as illustrated in Fig. 2.2, we observe that the voluntary participation conditions of non-dominant users in the Externality mechanism are never satisfied: these users can avoid paying taxes to the dominant user, while others pay her to increase her investment. More interesting however, is the fact that the voluntary participation conditions for the main investor may also fail to hold. This is because when user 1's exit equilibrium does not require her to exert effort, and the externality generated by her is small (i.e., small a), the collected taxes are not enough to persuade this dominant user to increase her effort level. Furthermore, we observe that although the Pivotal mechanism needs to give out a smaller reward to the dominant user as a increases (hence the jump in the third plot in Fig. 2.2), it still fails to avoid a deficit due to the small willingness of free-riders to pay the taxes required to cover this reward.

2.5 Discussion

2.5.1 Extending exit equilibria: Finding stable coalitions

As mentioned in Section 2.2, the definition of exit equilibrium considers unilateral deviations of users from mechanisms incentivizing socially optimal efforts, assuming all remaining users continue participating in the mechanism; the rationale is that it is a necessary (although not sufficient) condition for the users not to have incentives to unilaterally deviate if the socially optimal outcome is to be incentivized while maintaining voluntary participation and weak budget balance. In other words, if any one user's utility while participating in the proposed mechanism is lower than that she can attain at her exit equilibrium, then the mechanism fails to guarantee voluntary participation.

In this section, we look further into mechanisms which fail to guarantee voluntary participation. We are interested in identifying stable coalitions of participating users (a subset of all N users) that may emerge under a given incentive mechanism. To do so, we consider the possibility of multiple users opting out of the proposed mechanism (as opposed to only unilateral deviations considered in Section 2.2). We extend the definition of exit equilibrium, allowing E users to exit the mechanism while the remaining $N - E$ users continue participating.

Specifically, when a subset of users $\mathcal{E} \subset \mathcal{N}$ exit the proposed incentive mechanism, the resulting exit equilibrium, $\hat{\mathbf{x}}^{\mathcal{E}} := (\hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{E}}, \hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{N}-\mathcal{E}})$, is given by

$$\begin{aligned} \hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{N}-\mathcal{E}} &= \arg\max_{\mathbf{x}\geq 0} \ \sum_{j\notin\mathcal{E}} u_j(\mathbf{x}, \hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{E}}) \ , \\ \hat{x}^{\mathcal{E}}_k \ &= \arg\max_{x\geq 0} \ u_k(x, \hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{E}-\{k\}}, \hat{\mathbf{x}}^{\mathcal{E}}_{\mathcal{N}-\mathcal{E}}) \ , \forall k \in \mathcal{E}. \end{aligned} \tag{2.11}$$

A stable coalition

Through an illustrative example, we identify stable coalitions under this extended definition of exit equilibrium. In particular, similar to the counter-example in Section 2.3, we consider (approximations) of the weakest link risk function, i.e, users' utility

functions are given by

$$u_i(\mathbf{x}) = W - L(\sum_{j=1}^{N} \exp(-\gamma x_j))^{1/\gamma} - cx_i \ .$$

First, using the first order conditions on (2.2), and by symmetry, all users will be exerting the same socially optimal level of effort

$$x_i^* = \frac{1}{\gamma} \ln \frac{N}{(\frac{c}{L})^\gamma} \ , \forall i \ .$$

Consider the subset of E users $\mathcal{E} \subset \mathcal{N}$ who exit the proposed incentive mechanism. The exit equilibrium under the deviation of these users can be derived using the first order conditions on (2.11), given by

$$(N-E)\exp(-\gamma \hat{x}_k^{\mathcal{E}})(\sum_{j \notin \mathcal{E}} \exp(-\gamma \hat{x}_j^{\mathcal{E}}) + \sum_{j \in \mathcal{E}} \exp(-\gamma \hat{x}_j^{\mathcal{E}}))^{\frac{1}{\gamma}-1} = \frac{c}{L} \ , \ \forall k \notin \mathcal{E} \ ,$$
$$\exp(-\gamma \hat{x}_k^{\mathcal{E}})(\sum_{j \notin \mathcal{E}} \exp(-\gamma \hat{x}_j^{\mathcal{E}}) + \sum_{j \in \mathcal{E}} \exp(-\gamma \hat{x}_j^{\mathcal{E}}))^{\frac{1}{\gamma}-1} = \frac{c}{L} \ , \ \forall k \in \mathcal{E} \ .$$

Solving the above, the exit equilibrium $\hat{\mathbf{x}}^{\mathcal{E}}$ is given by

$$\hat{x}_k^{\mathcal{E}} = \frac{1}{\gamma} \ln \frac{(N-E)(E+1)^{1-\gamma}}{(\frac{c}{L})^\gamma}, \qquad \forall k \notin \mathcal{E} \ ,$$
$$\hat{x}_k^{\mathcal{E}} = \frac{1}{\gamma} \ln \frac{(E+1)^{1-\gamma}}{(\frac{c}{L})^\gamma}, \qquad \forall k \in \mathcal{E} \ .$$

Now, assume a user $i \notin \mathcal{E}$ is considering exiting the mechanism as well. Again, using the first order conditions on (2.11), and following similar steps as the above, this time for the subset of users $\mathcal{E} \cup \{i\}$ exiting the mechanism, the exit equilibrium

$\hat{\mathbf{x}}^{\mathcal{E}\cup\{i\}}$ will be given by

$$\hat{x}_k^{\mathcal{E}\cup\{i\}} = \frac{1}{\gamma}\ln\frac{(N-E-1)(E+2)^{1-\gamma}}{(\frac{c}{L})^{\gamma}}, \qquad \forall k \notin \mathcal{E}\cup\{i\}\ ,$$
$$\hat{x}_k^{\mathcal{E}\cup\{i\}} = \frac{1}{\gamma}\ln\frac{(E+2)^{1-\gamma}}{(\frac{c}{L})^{\gamma}}, \qquad \forall k \in \mathcal{E}\cup\{i\}\ .$$

For a stable coalition of $N-E$ users to form, we need to find the smallest set $\mathcal{E}$, such that $u_i(\hat{\mathbf{x}}^{\mathcal{E}\cup\{i\}}) \leq v_i(\hat{\mathbf{x}}^{\mathcal{E}}, t_i)$, where t_i is the tax assigned to a participating user i in some proposed incentive mechanism. Substituting for the exit equilibria derived above, user i's utilities when participating and staying out are given by

$$v_i(\hat{\mathbf{x}}^{\mathcal{E}}, t_i) = W - c(E+1+\hat{x}_i^{\mathcal{E}}) - t_i\ ,$$
$$u_i(\hat{\mathbf{x}}^{\mathcal{E}\cup\{i\}}) = W - c(E+2+\hat{x}^{\mathcal{E}\cup\{i\}})\ .$$

The voluntary participation condition therefore simplifies to

$$t_i \leq c\left(1+\frac{1}{\gamma}\ln\frac{1}{N-E}(\frac{E+2}{E+1})^{1-\gamma}\right)\ ,\forall i \notin \mathcal{E}\ .$$

Let E^* be the smallest number for which $N \leq E + e^{\gamma}(\frac{E+2}{E+1})^{1-\gamma}$ holds (note that we always have $E^* < N$). Given E^*, the Externality mechanism of Section 2.4.1 can lead to a stable coalition of size $M = N - E^*$ implementing the socially optimal solution in their M-user system, with the remainder E^* users not participating.[5]

It is also interesting to mention that the condition attained for having a stable coalition of all users (by setting $E = 0$) coincides with the positive instance of Section 2.3: if $N < e^{\gamma}2^{1-\gamma}$, the Externality mechanism can achieve the socially optimal solution, while guaranteeing voluntary participation and budget balance.

[5]The taxes assigned to the M participating users can be found using (2.7), and will in fact be zero at equilibrium (due to symmetry of the users). Also, the resulting effort profile will be an improved, yet sub-optimal solution for the N-user system.

A negative example

We close this section by noting that a stable coalition does not necessarily emerge in all problem environments. In particular, consider the following family of total effort games:

$$u_i(\mathbf{x}) = W - \exp(-\sum_{j=1}^{N} x_j) - c_i x_i \ .$$

Users are indexed such that $c_1 < c_2 < \ldots < c_N$. Also, assume $c_1 < \frac{c_2}{N-1}$. Consider a set of E users, $\mathcal{E} \subset \mathcal{N}$, exiting the mechanism. The resulting exit equilibrium, $\hat{\mathbf{x}}^{\mathcal{E}}$, depends primarily on user 1's participation choice.

$$\begin{aligned} &\text{If user } 1 \in \mathcal{E}: \quad \hat{x}_1^{\mathcal{E}} = \ln \frac{1}{c_1}, && \hat{x}_k^{\mathcal{E}} = 0, \ \forall k \neq 1 \ , \\ &\text{If user } 1 \notin \mathcal{E}: \quad \hat{x}_1^{\mathcal{E}} = \ln \frac{N-E}{c_1}, && \hat{x}_k^{\mathcal{E}} = 0, \ \forall k \neq 1 \ . \end{aligned}$$

First, note that once user 1 has already opted out, participation or opting out yields equivalent utilities for users $k \neq 1$. Therefore, any possible coalition has to include user 1.

Next, for any user to remain in a stable coalition $\mathcal{N} - \mathcal{E}$, we need to have $u_i(\hat{\mathbf{x}}^{\mathcal{E}\cup\{i\}}) \leq u_i(\hat{\mathbf{x}}^{\mathcal{E}}) - t_i$, where t_i is the tax assigned by the incentive mechanism to a participating user i. These conditions simplify to

$$\begin{aligned} &t_1 \leq c_1(1 - \tfrac{1}{N-E} - \ln(N-E)), \\ &t_k \leq \tfrac{c_1}{(N-E)(N-E-1)}, \ \forall k \notin \mathcal{E}, k \neq 1. \end{aligned}$$

For the mechanism to maintain weak budget balance, we need

$$\sum_{i \notin \mathcal{E}} t_i \leq c_1(1 - \ln(N-E)) \ .$$

However, the above is always negative for $N \geq 3$, indicating that there exists no

mechanism that can sustain such coalitions while maintaining weak budget balance. Also, note that the outcome for a coalition with $N = 2$ (user 1 and some user $k \neq 1$) will be equivalent to the Nash equilibrium. We therefore conclude that, regardless of the design of the mechanism, there exists no stable coalition in this family of total effort games.

2.5.2 Risk-averse users and cyber insurance contracts

In this section, we present an extension of the impossibility result of Section 2.3 to risk-averse users. Considering risk-averse users is of particular interest in studying the design of cyber insurance contracts. Cyber insurance has been widely proposed as a method for incentivizing the adoption of better security practices by users through strategies such as premium discrimination; see e.g., [57, 76]. Following the majority of the existing literature, we consider a monopolist cyber insurer (e.g., the government).[6] We assume the insurer is interested in improving the state of cyber security to its socially optimal levels (e.g., as required or directed by the government) through appropriately designed insurance contracts. The weak budget balance assumption ensures positive profits for this insurer, while voluntary participation models voluntary purchase of insurance from this provider.

CRRA functions for modeling risk aversion

Consider N interdependent users, with initial wealth W_i and loss L_i, each choosing an effort x_i. The cost of investment x_i is given by $h_i(x_i)$, with the probability of a successful attack given by $f_i(\mathbf{x})$. The utility function of user i is therefore

$$u_i(\mathbf{x}) = f_i(\mathbf{x})U_i(W_i - L_i) + (1 - f_i(\mathbf{x}))U_i(W_i) - h_i(x_i) \ .$$

In general, for risk-averse users, the function $U_i(\cdot)$ is a concave function. Here, we model risk aversion using CRRA (constant relative risk aversion) utility functions

[6] The assumption of a monopolist cyber insurer is in fact indispensable for this analysis. This is because, as mentioned in Section 1.3.2, the competition among multiple cyber insurers will inevitably lead to contracts that incentivize sub-optimal investments by the users.

[84], defined as follows:

$$U(c) = \begin{cases} \frac{1}{1-\theta} c^{1-\theta}, & \text{for } \theta > 0, \theta \neq 1 \ , \\ \ln c, & \text{for } \theta = 1 \ . \end{cases}$$

Note also that a CRRA utility with $\theta = 0$ represents risk-neutral users.

Assume users have the option of purchasing insurance contracts, specifying a *premium* ρ_i and an *indemnification payment* (coverage) level I_i. When insurance is purchased, the utility of user i will be given by

$$v_i(\mathbf{x}, \rho_i, I_i) = f_i(\mathbf{x}) U_i(W_i - \rho_i - L_i + I_i) + (1 - f_i(\mathbf{x})) U_i(W_i - \rho_i) - h_i(x_i) \ .$$

We now show the following negative result. Similar to Proposition 2.1, the proof is through a counter-example.

Proposition 2.2. *There exists no set of insurance contracts which can implement the socially optimal solution, while guaranteeing weak budget balance (no loss for the insurer) and voluntary participation (voluntary purchase of insurance contracts by users) simultaneously, in all instances of security games with risk-averse users with CRRA utilities.*

Proof. Similar to Section 2.3, we consider the (approximations) of weakest link risk functions $f_i(\mathbf{x}) = (\sum_{j=1}^{N} \exp(-\gamma x_j))^{1/\gamma}$, and set $W_i = W$, $L_i = L$, $h_i(x_i) = cx_i, \forall i$.

In this game, the socially optimal investment profile $\mathbf{x}^*$ is determined using (2.2), leading to

$$N \exp(-\gamma x_i^*)(\sum_{j=1}^{N} \exp(-\gamma x_j^*))^{\frac{1}{\gamma}-1} = \frac{c}{U(W)-U(W-L)} \ .$$

By symmetry, all users will be exerting the same socially optimal level of effort

$$x_i^* = \ln \frac{N^{1/\gamma}(U(W) - U(W-L))}{c} \ , \forall i \ .$$

The utility of users under this outcome, while also purchasing the optimal insurance

contract, is given by

$$u_i(\mathbf{x}^*) = \tfrac{-c}{U(W)-U(W-L)}(U(W-\rho)-U(W-L-\rho+I)) + U(W-\rho) - c\ln\frac{N^{1/\gamma}(U(W)-U(W-L))}{c} .$$

The exit equilibrium profile $\hat{\mathbf{x}}^i$ can be determined using the first order conditions on (2.4), leading to

$$(N-1)\exp(-\gamma\hat{x}_j^i)(\sum_{k\neq i}\exp(-\gamma\hat{x}_k^i)+\exp(-\gamma\hat{x}_i^i))^{\frac{1}{\gamma}-1} = \tfrac{c}{U(W)-U(W-L)} ,$$
$$\exp(-\gamma\hat{x}_i^i)(\sum_{k\neq i}\exp(-\gamma\hat{x}_k^i)+\exp(-\gamma\hat{x}_i^i))^{\frac{1}{\gamma}-1} = \tfrac{c}{U(W)-U(W-L)} .$$

Solving the above, we get

$$\hat{x}_i^i = \ln\frac{2^{1/\gamma-1}(U(W)-U(W-L))}{c} ,$$
$$\hat{x}_j^i = \ln\frac{(N-1)^{1/\gamma}2^{1/\gamma-1}(U(W)-U(W-L))}{c} , \forall j\neq i .$$

The utility of the outlier i under the exit equilibrium is given by

$$u_i(\hat{\mathbf{x}}^{\mathbf{i}}) = -2c + U(W) - c\ln\frac{2^{1/\gamma-1}(U(W)-U(W-L))}{c} .$$

We now proceed to the analysis of insurance contracts. First note that the insurer has the following total profit:

$$P^* := \sum_k \rho_k - \sum_k I_k f_k(\mathbf{x}^*) = N\rho - N\tfrac{c}{U(W)-U(W-L)}I \geq 0 . \qquad (2.12)$$

where, $\rho_k = \rho$ and $I_k = I$ for all users due to symmetry.

The voluntary participation condition for a user i to purchase insurance is given

by

$$\frac{-c}{U(W)-U(W-L)}(U(W-\rho)-U(W-L-\rho+I))+U(W-\rho)$$
$$-c\ln\frac{N^{1/\gamma}(U(W)-U(W-L))}{c}\geq -2c+U(W)-c\ln\frac{2^{1/\gamma-1}(U(W)-U(W-L))}{c}\ .$$

Define the following:

$$K_1:=\frac{c}{U(W)-U(W-L)},\quad K_2:=c(2+\frac{1}{\gamma}\ln\frac{2^{1-\gamma}}{N})\ .$$

Then, the voluntary participation conditions can be re-written as

$$U(W)-U(W-\rho)\leq K_2-K_1(U(W-\rho)-U(W-L-\rho+I))\ . \tag{2.13}$$

Can the insurer attain social optimality, voluntary participation, and weak budget balance in this game? Take equations (2.12) and (2.13) together. First, for all inequalities (2.13), relax the requirement by assuming $L=I$, that is, users are offered full coverage. Note that if this inequality fails for $L=I$, it will certainly fail for any $0\leq I<L$. Also, for the inequality in (2.12), assume $I=0$. Again, if (2.12) fails for $I=0$, it will certainly fail for all $0<I\leq L$. We show that the set of relaxed inequalities are inconsistent, and consequently, by the above argument, the original conditions in (2.12) and (2.13) can not be satisfied simultaneously either. We are therefore looking to find the premiums ρ, such that

$$U(W)-U(W-\rho)\leq K_2\ ,\quad \rho\geq 0\ .$$

Take any function in the CRRA family, $U(c)=\frac{1}{1-\theta}c^{1-\theta}$. The above conditions simplify to

$$\rho\leq W-(W^{1-\theta}-(1-\theta)c(2+\frac{1}{\gamma}\ln\frac{2^{1-\gamma}}{N}))^{\frac{1}{1-\theta}}\ ,\quad \rho\geq 0\ .$$

Fix the approximation parameter of the weakest link risk function, $\gamma>0$, and that

of the CRRA risk aversion function to a $\theta < 1$. We observe that, if $2 + \frac{1}{\gamma} \ln \frac{2^{1-\gamma}}{N} \leq 0$, then $\rho < 0$, and the second inequality (on the insurer's profit) cannot be satisfied. Therefore, if the number of users is such that $N > e^{2\gamma} 2^{1-\gamma}$, it is impossible to design insurance contracts that guarantee social optimality, voluntary participation, and weak budget balance. □

We conclude that unless additional information on users' preferences or the network structure is available, it is in general not possible to design insurance contracts that can result in a socially desirable state of security, are voluntarily purchased by the users, and can generate revenue for the cyber insurer.

2.5.3 The role of a security software vendor

Given the potential budget deficit of the Pivotal mechanism when achieving social optimality and voluntary participation in some instances of security games, in this section, we consider the availability of additional external resources/payments to the designer of the Pivotal mechanism. In particular, we consider a security product vendor entering the game as the mechanism designer. The idea of bundling security product vendors and mechanism designers (more specifically, cyber insurers) has been studied in [72, 99]. The authors in [72] propose the idea of a provider investing in increasing the security of widely used software products, leading to a decrease in monoculture risks. The focus of [99] on the other hand is on the security product pricing problem as a method for generating additional revenue for the cyber insurer. Similarly, we consider the effects of such bundling on the performance of the Pivotal mechanism.

Specifically, we allow the vendor to leverage the profit from the additional sales in cyber security products resulting from the requirements of improved security imposed on users, to cover the deficit and generate additional profit. Through an illustrative example, we show that this modification can lead to an expansion of the space of positive instances, but nevertheless, that this profit is not necessarily enough to cover the budget deficit in all instances of the game.

Formally, consider the total effort security game with exponential risks and linear

investment costs, with uniform W and $L = 1$. The utility functions of users in this game are given by

$$u_i(\mathbf{x}) = W - \exp(-\sum_j x_j) + c_i x_i \ .$$

Users are indexed such that $c_1 < c_2 < \ldots < c_N$. Also, assume $c_1 < \frac{c_2}{N-1}$. The socially optimal solution and exit equilibria are given by

$$\begin{aligned}
\text{SO: } x_1^* &= \ln \frac{N}{c_1}, x_j^* = 0, \forall j \neq 1 \ , \\
\text{EE, } j \neq 1\text{: } \hat{x}_1^j &= \ln \frac{N-1}{c_1}, \hat{x}_k^j = 0, \forall k \neq 1 \ , \\
\text{EE, } j = 1\text{: } \hat{x}_1^1 &= \ln \frac{1}{c_1}, \hat{x}_k^1 = 0, \forall k \neq 1 \ .
\end{aligned}$$

We consider the Pivotal mechanism, as the taxes in it guarantee voluntary participation, and we can thus focus on budget balance issues. These taxes are given by

$$\begin{aligned}
t_j^P &= c_1(-\frac{1}{N} + \ln \frac{N}{N-1}), \forall j \neq 1 \ , \\
t_1^P &= -c_1 \frac{(N-1)^2}{N} \quad .
\end{aligned}$$

The sum of all taxes will be given by

$$T^P := \sum_i t_i^P = c_1(N-1)(-1 + \ln \frac{N}{N-1}) \ .$$

The above is negative for all N, indicating a budget deficit for any number of users in the absence of external resources.

Alternatively, assume the Pivotal mechanism is implemented by a security product vendor. For simplicity, we assume that the marginal cost of production of security products is negligible for the vendor. Therefore, by the introduction of the Pivotal mechanism, the vendor makes the following additional profit from the increased se-

curity adoption (compared to the Nash equilibrium):

$$\Delta P := \sum_i c_i x_i^* - \sum_i c_i \tilde{x}_i = c_1 \ln \frac{N}{c_1} - c_1 \ln \frac{1}{c_1} = c_1 \ln N \ .$$

Considering the vendor's profit, the total budget following the introduction of the Pivotal mechanism is given by

$$T^P + \Delta P = c_1(N-1)(-1 + \ln \frac{N}{N-1}) + c_1 \ln N \ .$$

The above is positive if and only if $N = 2$. We conclude that the space of positive instances has expanded (albeit slightly) once the profit of additional product sales enters the market. In other words, with 2 users, a security vendor can introduce taxes that achieves the socially optimal levels of security, are voluntarily adopted by the users, and generate positive revenue for the designer/vendor. The budget deficit continues to hold for $N \geq 3$.

2.6 Related Work

2.6.1 Existing possibility and impossibility results

The presented impossibility results are different from those in the existing literature, in either the selected equilibrium solution concept, the set of properties the mechanism is required to satisfy, or the space of utility functions. For example, the Myerson and Satterthwaite result [84] (stronger version of Hurwicz's impossibility on dominant strategy implementation) establishes impossibility of Bayesian Nash implementation with optimality, individual rationality, and strong budget balance when users have quasi-linear utilities; our result differs in (1) solution concept (we are considering full Nash implementation), and (2) by only requiring a weaker condition of *weak* budget balance (thus making it a stronger impossibility result in this sense).

The most closely related impossibility result to our work is that of [112], which also studies impossibility results in the provision of non-excludable public goods. Our

adoption of the term voluntary participation as opposed to individual rationality is similar to this work. However, our work differs from [112] in two main aspects. First, in terms of users' preferences, [112] studies Cobb-Douglas utilities, whereas we consider quasi-linear utilities, as well as risk averse users with CRRA utilities. More importantly, [112] considers the production of a single (non-excludable) good with constant return to scale technology. As a result, although outliers benefit from the spill-over of the produced good, they no longer contribute to its provision. This is in contrast to the goods studied herein, e.g., the security of an outlier can still affect those of the participants. The notion of exit equilibrium is introduced to fully capture this distinction.

The current work should also be viewed in conjunction with existing possibility results, notably [25, 56, 103, 115], which consider the provision of excludable public goods (i.e., zero outside options) for users with the same utility functions and under the same informational constraints as the current work. As a result, they show that the Externality and Pivotal mechanisms simultaneously guarantee social optimality, voluntary participation, and weak budget balance. Therefore, the goal of this chapter is not solely to prove the impossibility of the design, but to highlight the important distinction users' outside options make in the choice of a mechanism.

2.6.2 Incentivizing improved cyber security

Similar to the work in this chapter, existing literature has proposed the introduction of monetary taxes/rewards or transfers for incentivizing better security behavior. One such theoretically attractive incentive mechanism that may result in optimal levels of investment is the *liability rule* [70, 122], where users are required to compensate others for the damages caused by their under-investment in security. However, these mechanisms are costly in practice, as it is difficult to accurately determine the cause of a damage. Alternatively, [122] proposes assigning a level of *due care*, in which following a security incident, a user is penalized only if her level of investment is lower than a pre-specified threshold. In [55], users can be incentivized to improve their investment in security if they are assigned bonuses/penalties based on their

security outcome (e.g., users get a reward if their security has not been breached), or get subsidized/fined based on their effort (e.g., users are given discounts if they buy security products). The current work differs in that it explicitly models budget balance and participation incentives in these mechanisms.

Our findings in Section 2.5.2 are most related to the study of cyber insurance contracts. Cyber insurance has been widely proposed as both a method for mitigating cyber risks, and as an incentive mechanism for internalizing the externalities of security investments; see e.g., [53, 57, 76, 100, 75, 12]. In particular, [57, 76, 100] have shown that by engaging in premium discrimination, a monopolistic profit-neutral cyber insurer can induce socially optimal security investments in an interdependent systems where security decisions are binary (i.e., invest or not). However, participation in these studies is assumed to be mandatory, e.g., users are enforced through policy mandates to purchase insurance. Our findings show that this assumption is indispensable; it is not in general possible to design non-compulsory insurance contracts that induce socially optimal behavior and generate profits for the cyber insurer.

The work in this chapter is also related to [64, 85]. The weighted effort risk model studied in Section 2.4.2 is a generalization of the total effort model in [122], and is similar to the effective investment model in [64] and the linear influence network game in [85]. The authors in [85] identify properties of the interdependence matrix affecting the existence and uniqueness of the Nash equilibrium in the linear influence model. Using a similar effective investment model, [64] determines a bound on the price of anarchy gap, i.e., the gap between the socially optimal and Nash equilibrium investments, depending on the adjacency matrix. Our work on the above model fills a gap within this literature as well, by (1) introducing the study of exit equilibria, and (2) analyzing the general mechanism design problem, in both this model, as well as more general environments.

Finally, the problem of incentivizing optimal security investments in an interconnected system is one example of problems concerning the provision of non-excludable public goods in social and economic networks. Other examples include creation of new parks or libraries at neighborhood level in cities [7], reducing pollution by neigh-

boring towns [38], or spread of innovation and research in industry [16]. Section 3.1.3 reviews additional related work in this area.

2.7 Conclusion

This chapter introduced the notion of exit equilibrium to study voluntary participation of users in mechanisms for provision of non-excludable public goods, such as security. This equilibrium concept accounts for both the spill-over of the public good produced by the participants on an outlier, as well as the continued influence of the outlier on the provision of the public good (here, the state of security). We have shown the fundamental result that, given these outside options, it is not possible to design a tax-based incentive mechanism to implement the socially optimal solution while guaranteeing voluntary participation and maintaining a weakly balanced budget, without additional information on the graph structure and users' preferences. We showed that despite the lack of a reliable mechanism for general problem instances, we can identify positive instances under restricted parameter conditions, for which it is possible to guarantee social optimality, weak budget balance, and voluntary participation using well-known incentive mechanisms. Alternatively, for instances in which the three properties are not attainable, we may be able to identify stable coalitions of participating users, by using an extended definition of exit equilibrium which accounts for possibly multiple outliers. We extended our result to risk-averse users purchasing cyber insurance contracts, highlighting the possible limitations of using cyber insurance as a tool for improving the state of cyber security to its socially desirable level.

An important implication of our result is that, when a designer lacks additional information about the specifics of the problem environment and users' preferences, she may choose to forgo the social optimality requirement, instead focusing on reliably attaining a sub-optimal solution while guaranteeing full voluntary participation and weak budget balance. Characterizing mechanisms that can lead to such sub-optimal solutions remains an interesting direction of future work.

Chapter 3

Public Good Provision Games on Networks

3.1 Introduction

3.1.1 Motivation: Beyond security games

In this chapter, we study the class of public good provision games on networks; i.e., the strategic interactions in a given network of agents who exert effort towards the provision of a public good. In these settings, the effort exerted by an agent affects not only herself, but also other agents interacting with her. The model in this chapter includes the weighted effort games studied in Section 2.4.2 as a special case. The setting applies to many social and economic applications. We present some examples.

First, consider the *spread of information and innovation* in networks. New technologies developed by one entity/agent in the network may later be adopted by other agents in the network. The interactions determining these innovation spillovers can in general depend on factors such as geographic location [5] and the interacting agents' access to resources [43]. Given this network, the possibility of spillovers can affect the decision of agents for investing in innovation or experimenting with new methods, leading to possible free-riding behavior. Specifically, a neighbor's effort can be either a *substitute* or a *complement* to an agent's own effort. Strategic substitutes (complements) are defined by the property that an increase of effort by an agent decreases (increases) her neighbors' marginal utilities, leading them to decrease (in-

crease) their effort levels in response. For instance, if farmers in a village have the option of experimenting with a new variety of seeds, then those whose neighbors are experimenting are less likely to do so themselves [43]. In this example, neighbors' efforts are a substitute to an agent's own effort. It may also be the case that an agent needs to increase her levels of experimentation in response to that of her neighbors, in order to remain competitive in her industry. In that case, the neighbors' efforts are a complement to the agent's own effort.

Another setting of interest is *investments in security* by interdependent entities. Security has been commonly viewed as a public good; examples include the model of airline security in [70], as well as the studies of cyber-security in [122, 64, 85, 54, 93]. Investment in security by a neighbor can act as either a substitute or a complement to an agent's own effort. For example, in weakest target games [54], neighbors' efforts are complementary since the agent with the lowest security will be selected as the target by an attacker. For total effort games [122, 54], on the other hand, neighbors' efforts act as substitutes, as an agent's overall security is assumed to be determined by the sum of her own investment and her neighbors' efforts.

In addition to the above applications, creation of new community parks or libraries in cities [7], investment in pollution reduction measures by neighboring towns [39], and even the states of happiness of individuals on a social network [44], can be studied using this framework.

3.1.2 Chapter overview

The public good provision game studied in this chapter belongs to the growing literature on games on networks; see [62, 17] for recent surveys. Specifically, we consider games in which, given the network structure, an agent's payoff depends on her own effort, as well as a *weighted sum* of her neighbor's efforts. Our model allows for both complements and substitutes, different strengths of interactions (weighted graphs), and unidirectional interactions (directed graphs). We are interested in the study of Nash equilibria, Pareto efficient effort profiles, and semi-cooperative equilibria (we define these as the effort profiles emerging when coalitions of agents interact with one

another). Our results provide an understanding of how the aforementioned outcomes (i.e., the results of agents' strategic interactions) are affected by the properties of the network.

Our first result identifies necessary and sufficient conditions on the structure of the network (in terms of the dependence matrix) that guarantee that a Nash equilibrium exists and is *unique*. We will show that previous results on the uniqueness of the Nash equilibrium [92, 85, 7, 18] can be recovered as corollaries of our first theorem.

In addition to studying uniqueness, we identify (weaker) necessary and sufficient conditions for the *existence* of Nash equilibria in two classes of games at the extremes of our model, namely games with strategic complements and games with strategic substitutes. The identified conditions (for both existence and uniqueness) are solely based on the structure of the network.

We then establish a connection between the agents' centrality in their dependence network, and the effort they exert at interior Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria. We separate the effects of dependencies (outgoing edges of the interaction network) and influences (incoming edges of the interaction network) on agents' effort decisions. We further discuss how the formation of coalitions is reflected in the centrality-effort characterization. We then uncover an *alternating effect* along walks of different length in the network. We show that in a network with strategic substitutes, this alternating effect implies that changes along each walk of odd (even) length will negatively (positively) affect the agent's final decision. We provide additional intuition and examples for general networks in Section 3.4.3.

3.1.3 Related work

Public good provision games, and network games in general, have recently received increasing attention. We refer the interested reader to [62, 17] for surveys on this general area. Here, we present the work most related to the current chapter.

Most of the existing work has studied the Nash equilibrium of network games. Previous work on identifying conditions for existence and uniqueness of Nash equi-

libria in public good provision games include [92, 85, 7, 18]. Both [85, 92] identify a similar sufficient condition for the uniqueness of the Nash equilibrium in public good provision games. The authors of [18] present a different sufficient condition for the uniqueness of the Nash equilibrium. Their result illustrates the role of the *lowest* eigenvalue of the network in determining the outcome of strategic interactions. Finally, [7] provides necessary and sufficient conditions for the uniqueness of the Nash equilibrium in a class of games with hidden complementarities. In addition to identifying the necessary condition for uniqueness of Nash equilibria in general networks, we show that the sufficiency results of [92, 85, 7, 18] can be recovered as corollaries of our main theorem. This comparison will further illustrate the key role of the lowest eigenvalue in (asymmetric) games with complementarities (in addition to the symmetric networks and particular classes of asymmetric networks studied in [18]).

Our work is also closely related to [39, 8], which provide graph-theoretical interpretations of agents' efforts in terms of their centralities in a suitably defined network. The work of Elliott and Golub in [39] focuses mainly on the implementation of Pareto efficient outcomes. The current work and [39] differ in the network used as the basis of analysis: rather than working directly on the dependence matrix, [39] focuses on a *benefits matrix* that is derived from the network graph; an entry B_{ij} of the matrix is the marginal rate at which i's effort can be substituted by the externality of j's action. The authors show that Lindahl outcomes can be interpreted as node centralities in this benefits matrix. Ballester et al. [8], on the other hand, study the Nash equilibrium of a linear quadratic interdependence model, and relate the equilibrium effort levels to the nodes' Bonacich centralities in a suitably defined matrix of local complementarities. Despite the difference in the base models, both games have the same linear best-reply functions. As a result, the characterization of Nash equilibria based on Bonacich centralities (used in [8]) and alpha-centralities (used in this chapter) are equivalent (see footnote 6). We will see that using (the more general measure of) alpha-centrality allows us to provide graph-theoretical interpretations of Pareto efficient efforts and semi-cooperative equilibria as well.

3.1.4 Chapter contributions

The main contributions of this chapter are summarized as follows:

- This chapter identifies the necessary and sufficient condition for *uniqueness* of Nash equilibria in public good provision games. We show that our result unifies, and strengthens, previous results in the literature.
- It further identifies the necessary and sufficient condition for the *existence* of Nash equilibria in two subclasses of our model, namely games with strategic substitutes and games with strategic complements.
- It presents a *graph theoretical characterization* of agents' actions at different effort profiles, namely the Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria (in terms of node centralities). Our characterization separates the effects of agents' dependencies and influences. It also uncovers an alternating effect over walks of different length.

3.1.5 Chapter organization

The remainder of this chapter is organized as follows. We present the model for public good provision games in Section 3.2, followed by conditions for the existence and uniqueness of Nash equilibria in Section 3.3. Section 3.4 discusses the graph theoretical characterization of different effort outcomes. In Section 3.5, we generalize the graph-theoretical characterization to games in which agents belong to different coalitions. Section 3.6 concludes the chapter.

3.2 Model and preliminaries

3.2.1 Public good provision games

We study the strategic interactions of N agents constituting the vertices of a directed network $\mathcal{G} = (\mathcal{N}, \mathcal{E})$; where $\mathcal{N}$ and $\mathcal{E}$ denote the set of agents and links, respectively. Each agent $i \in \mathcal{N}$ chooses to exert *effort* $x_i \in \mathbb{R}_{\geq 0}$ towards the provision

of a public good.[1] Agent i's payoff depends on her own effort, as well as the effort exerted by other agents in her local neighborhood $\mathcal{N}_i := \{j | \{i \to j\} \in \mathcal{E}\}$. An edge $\{i \to j\}$ indicates that agent i *depends* on agent j. The strength and type of this dependence are determined by the weight $g_{ij} \in \mathbb{R}$ of the edge $\{i \to j\}$. In particular, $g_{ij} > 0$ (< 0) indicates that j's effort is a substitute (complement) to i's effort. Let $\mathbf{G} = (g_{ij})$ denote the *dependence* matrix of the graph.

Let $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$ denote the profile of efforts exerted by all agents. The utility of agent i at this effort profile is given by

$$u_i(\mathbf{x}; \mathbf{G}) = b_i(x_i + \sum_{j \in \mathcal{N}_i} g_{ij} x_j) - c_i x_i \ . \tag{3.1}$$

Here, $c_i > 0$ is the marginal cost of effort for agent i, and $b_i(\cdot)$ is a twice-differentiable, strictly increasing, and strictly concave function, determining the benefit to agent i from the aggregate effort she experiences.

This model has been used to study the local provision of public goods in [92, 18, 16]. In the context of security (when viewed as a public good), it is a generalization of the total effort model used in the seminal work of Varian [122], and is similar to the effective investment model of [64] and the linear influence network game of [85].

3.2.2 Characterizing effort outcomes

We now consider the problem of finding the efforts at two outcomes of public good provision games: the Nash equilibria and Pareto efficient effort profiles. A Nash equilibrium is an effort profile at which no agent has an incentive to unilaterally deviate from her strategy given other agents' efforts. This is an effort profile that emerges at the status quo as a result of strategic agents' interactions. A Pareto efficient outcome is an effort profile at which it is not possible to increase any agent's

[1] In this thesis, we follow the definition of Mas-Collel, Whinston, and Green [84], and define public goods as those that are non-rivalrous, i.e., goods for which consumption by an agent does not reduce its availability to others. As a result, in the public good provision games of this chapter, we allow for both complements and substitutes, as well as both excludable and non-excludable public goods. We only explicitly make the distinction based on excludability in Section 3.5, when studying effort profiles that emerge under coalitions.

utility without making at least one other agent worse off as a result. It is therefore an indication of the profile's efficiency relative to other possible outcomes. These profiles can be attained through negotiation among agents, or following the introduction of appropriate incentives such as monetary taxes/rewards.

Nash equilibria

We start with the Nash equilibria of public good provision games.[2] A Nash equilibrium is a fixed point of the best-reply map. Formally, let $f_i(\mathbf{x}_{-i}; \mathbf{G})$ be the best reply of agent i; this is the effort that maximizes i's payoff given other agents' profile of efforts $\mathbf{x}_{-i}$ and the dependence matrix $\mathbf{G}$. For agents with utility (3.1), this best reply is given by

$$f_i(\mathbf{x}_{-i}; \mathbf{G}) = \max\{0, \bar{q}_i - \sum_{j \in \mathcal{N}_i} g_{ij} x_j\} \ , \tag{3.2}$$

where $\bar{q}_i$ is the effort level at which $b_i'(\bar{q}_i) = c_i$. In other words, $\bar{q}_i$ is the aggregate effort at which i's marginal utility equals her marginal cost.[3]

For each effort level x_i, define a corresponding complementary variable w_i. Then, finding a fixed point of the mapping (3.2) is equivalent to finding a solution to the

[2]We consider pure Nash equilibria of the game. Given the strict concavity of the payoffs in (3.1), playing the average of a set of effort levels leads to a higher payoff than a mixed strategy over that set. As a result, there is no mixed strategy Nash equilibrium for our games.

[3]It is worth mentioning that the best-response mapping of games with linear quadratic payoffs is also of the form (3.2). Formally, in a game with linear quadratic payoffs, the utility of agent i is given by [8]

$$u_i(\mathbf{x}; \mathbf{G}) = \bar{q}_i x_i - \frac{1}{2} x_i^2 - \sum_{j \neq i} g_{ij} x_i x_j \ ,$$

where $\bar{q}_i$ is a given constant. The delinquency games of [9] and Cournot competitions with heterogeneous goods and network collaboration (in which g_{ij} determines the degree of substitutability of i's good with j's output) are special cases of games with linear-quadratic payoffs; see [18, 7] for examples. All our results regarding Nash equilibria apply to these (as well as other games with linear best-replies of the form (3.2)) as well.

following problem:

$$\begin{aligned} \mathbf{w} - (\mathbf{I} + \mathbf{G})\mathbf{x} &= -\bar{\mathbf{q}} \ , \\ \mathbf{w} \succeq \mathbf{0} \ , \quad \mathbf{x} &\succeq \mathbf{0} \ , \\ \mathbf{w}^T \mathbf{x} &= 0 \ . \end{aligned} \tag{3.3}$$

Here, $\bar{\mathbf{q}} := \{\bar{q}_1, \ldots, \bar{q}_N\}$, and $\mathbf{I}$ is the $N \times N$ identity matrix. The optimization problem in (3.3) is an instance of *linear complementarity problems* (LCPs).

The Linear Complementarity Problem (LCP) refers to a family of problems which arise in solving linear programming and quadratic programming problems, as well as in finding Nash equilibria of bimatrix (two-player non-zero sum) games [30]. For example, the necessary first order optimality (KKT) conditions of a quadratic programming problem constitute an LCP. In addition to these direct connections, LCPs have found applications in the study of market equilibrium, computing Brouwer and Kakutani fixed points, and developing efficient algorithms for solving nonlinear programming problems [87].

Formally, an LCP $(\mathbf{M}, \mathbf{q})$ is the problem of finding vectors $\mathbf{x} \in \mathbb{R}^n$ and $\mathbf{w} \in \mathbb{R}^n$ satisfying

$$\begin{aligned} \mathbf{w} - \mathbf{M}\mathbf{x} &= \mathbf{q} \ , \\ \mathbf{w} \succeq \mathbf{0} \ , \quad \mathbf{x} &\succeq \mathbf{0} \ , \\ \mathbf{w}^T \mathbf{x} &= 0 \ . \end{aligned} \tag{3.4}$$

An LCP is therefore fully determined by an $n \times n$ square matrix $\mathbf{M}$ and a constant right-hand vector $\mathbf{q} \in \mathbb{R}^n$.

Comparing (3.4) with (3.3), we observe that finding the Nash equilibria for the public good provision game is equivalent to solving the LCPs $((\mathbf{I} + \mathbf{G}), -\bar{\mathbf{q}})$. In Section 3.3, we will identify conditions on the dependence matrix $\mathbf{G}$ such that solutions to (3.3) exist and are unique, *for all* right-hand vectors $\mathbf{q}$. In other words, we are interested in the structural properties of the interaction network that guarantee the existence and uniqueness of Nash equilibria, for any payoffs of the form (3.1),

irrespective of the realization of benefit functions or marginal costs of effort.

Remark (on the sign of $\mathbf{q}$*):* We note that the right-hand vector entries q_i in (3.4) can be either positive, negative, or zero. In particular, for $\mathbf{q} \succeq \mathbf{0}$ in (3.4), the LCP always has the solution $\mathbf{w} = \mathbf{q}$ and $\mathbf{x} = \mathbf{0}$. In the case of Nash equilibria with LCP $(\mathbf{I} + \mathbf{G}, -\bar{\mathbf{q}})$, $\bar{\mathbf{q}} \prec \mathbf{0}$ implies that the zero effort profile $\mathbf{x} = \mathbf{0}$ is always a possible Nash equilibrium. This observation can be intuitively explained as follows. Recall that q_i indicates the effort level at which agent i's marginal utility equals her marginal cost. A negative q_i therefore indicates that exerting effort is not cost-efficient for agent i. Hence, a zero effort equilibrium is indeed to be expected.

Pareto efficient effort profiles

We also consider Pareto efficient effort profiles of the public good provision game. Formally, we consider the solutions to the following problem:

$$\max_{\mathbf{x} \succeq \mathbf{0}} \sum_i \lambda_i u_i(\mathbf{x}) \ ,$$

where $\boldsymbol{\lambda} := \{\lambda_1, \cdots, \lambda_N\}$ is a vector of non-negative weights. By [84, Proposition 16.E.2], for the strictly concave utility functions $u_i(\cdot)$ given by (3.1), the set of solutions to this linear welfare maximization problem, as $\boldsymbol{\lambda}$ ranges over the set of all strictly positive weight vectors, leads to the Pareto optimal effort profiles. It is worth noting that solving for the Pareto efficient profile with unit vector of weights $\boldsymbol{\lambda} = \mathbf{1}$ in (3.5) will lead to the socially optimal profile of efforts $\mathbf{x}^* = \arg\max_{\mathbf{x} \succeq \mathbf{0}} \sum_i u_i(\mathbf{x})$.

We now proceed to characterizing these profiles. Consider the Pareto efficient effort profile $\mathbf{x}^{\boldsymbol{\lambda}}$ corresponding to the strictly positive weight vector $\boldsymbol{\lambda}$. That is,

$$\mathbf{x}^{\boldsymbol{\lambda}} = \arg\max_{\mathbf{x} \succeq \mathbf{0}} \sum_k \lambda_k u_k(\mathbf{x}) \ . \tag{3.5}$$

The first order condition on (3.5) with respect to x_i implies that at the Pareto efficient

solution, the following should hold:

$$b_i'(x_i^{\boldsymbol{\lambda}} + \sum_{j \in \mathcal{N}_i} g_{ij} x_j^{\boldsymbol{\lambda}}) + \sum_{k, \text{ s.t. } i \in \mathcal{N}_k} \frac{\lambda_k}{\lambda_i} g_{ki} b_k'(x_k^{\boldsymbol{\lambda}} + \sum_{j \in \mathcal{N}_k} g_{kj} x_j^{\boldsymbol{\lambda}}) = c_i - z_i, \ \forall i \ . \tag{3.6}$$

Here, z_i is a complementary variable corresponding to the effort level $x_i^{\boldsymbol{\lambda}}$.

Consider an *interior* Pareto efficient outcome in which all agents exert non-zero effort, i.e., $\mathbf{z} = \mathbf{0}$. We will study graph-theoretical characterizations of these outcomes, as well as interior Nash equilibria, in Section 3.4. Define $\mathbf{q}^{\boldsymbol{\lambda}}$ as the effort levels satisfying:

$$b_i'(q_i^{\boldsymbol{\lambda}}) + \sum_{k, \text{ s.t. } i \in \mathcal{N}_k} \frac{\lambda_k}{\lambda_i} g_{ki} b_k'(q_k^{\boldsymbol{\lambda}}) = c_i \ , \forall i. \tag{3.7}$$

Intuitively, $\mathbf{q}^{\boldsymbol{\lambda}}$ is the vector of efforts at which the marginal *social* benefits equal the marginal (social) costs of effort. When $\mathbf{I} + \mathbf{G}$ is invertible, we can find the following alternative expression for $q_i^{\boldsymbol{\lambda}}$ by solving the system of equations in (3.7):

$$b_i'(q_i^{\boldsymbol{\lambda}}) = \left((\mathbf{I} + \boldsymbol{\Lambda}^{-1} \mathbf{G}^T \boldsymbol{\Lambda})^{-1} \mathbf{c}\right)_i \ .$$

Here, $q_i^{\boldsymbol{\lambda}}$ can be interpreted as the aggregate effort level at which agent i's marginal benefit equals her *modified* marginal cost. The modification depends on the graph structure, as well as the weights $\boldsymbol{\lambda}$. We will elaborate further in Section 3.4.3.

Finding such interior Pareto efficient effort profile $\mathbf{x}^{\boldsymbol{\lambda}}$ is equivalent to finding a solution with $\mathbf{w} = \mathbf{0}$ to the following problem:

$$\begin{aligned} \mathbf{w} - (\mathbf{I} + \mathbf{G})\mathbf{x} &= -\mathbf{q}^{\boldsymbol{\lambda}} \ , \\ \mathbf{w} \succeq \mathbf{0} \ , \quad \mathbf{x} &\succeq \mathbf{0} \ , \\ \mathbf{w}^T \mathbf{x} &= 0 \ . \end{aligned} \tag{3.8}$$

In other words, finding interior Pareto efficient outcomes is equivalent to finding solutions to the LCP $((\mathbf{I} + \mathbf{G}), -\mathbf{q}^{\boldsymbol{\lambda}})$ with $\mathbf{w} = \mathbf{0}$. We study conditions under which such solutions exist in Section 3.4.1.

3.3 Existence and uniqueness of Nash equilibria

In this section, we study conditions under which Nash equilibria for public good provision games exist, and in particular, conditions under which these profiles are unique. We contrast our result with those in the existing literature, and show how existing conditions on the uniqueness of Nash equilibria can be recovered as corollaries of our main theorem.

3.3.1 Existence and uniqueness

Using the LCP formulations of the problems for finding the Nash equilibria in (3.3), we identify conditions for the existence and uniqueness of the corresponding effort profile. We begin with the following definition.

Definition 3.1. *A square matrix* $\mathbf{M}$ *is a* P-matrix *if the determinants of all its principal minors (i.e., the square submatrix obtained from* $\mathbf{M}$ *by removing a set of rows and their corresponding columns) are strictly positive.*

The following theorem provides the necessary and sufficient condition under which the Nash equilibrium exists and is unique.

Theorem 3.1 (Uniqueness)**.** *The public good provision game has a unique Nash equilibrium if and only if* $\mathbf{I}+\mathbf{G}$ *is a P-matrix.*

The proof follows from results on the uniqueness of solutions of LCPs; see e.g., [86, Theorem 4.2]. We illustrate Theorem 3.1 through an example.

Example 3.1. Consider a network of two nodes. We study the Nash equilibria of a public good provision game with payoffs:

$$u_i(\mathbf{x};\mathbf{G}) = 1 - \exp(-x_i - g_{ij}x_j) - \frac{1}{e}x_i, \text{for } i \in \{1,2\}, j \neq i \ .$$

Note that $\mathbf{I}+\mathbf{G}$ is a P-matrix if and only if $g_{12}g_{21} < 1$.

(i) First, let $g_{12} = g_{21} = \frac{1}{2}$. Then, by Theorem 3.1, this game should have a unique Nash equilibrium. Indeed, this unique equilibrium is given by $x_1 = x_2 = \frac{2}{3}$.

(ii) Next, consider $g_{12} = g_{21} = 2$. Then $\mathbf{I} + \mathbf{G}$ is not a P-matrix, and the game need not have a unique Nash equilibrium. For the given payoffs, there are three possible Nash equilibria: $(x_1, x_2) = (0, 1)$, $(x_1, x_2) = (1, 0)$, and $(x_1, x_2) = (\frac{1}{3}, \frac{1}{3})$.

(iii) Finally, let $g_{12} = g_{21} = -2$. Again, $\mathbf{I} + \mathbf{G}$ is not a P-matrix, and hence by Theorem 3.1, the corresponding game need not have a unique equilibrium. In fact, under the assumed payoff functions, the game will have no Nash equilibrium.

We now turn to the more general question of *existence* of Nash equilibria. We are interested in weaker conditions than those of Theorem 3.1 that guarantee at least one Nash equilibrium exists. Unlike uniqueness, there is no simple characterization of matrices $\mathbf{M}$ for which an LCP $(\mathbf{M}, \mathbf{q})$ has a solution. Nevertheless, we can identify existence results on two particular subclasses of games, namely games of strategic substitutes and games of strategic complements. Recall that for a game of strategic substitutes (complements), $g_{ij} \geq 0$ $(g_{ij} \leq 0), \forall i, j \neq i$.

Theorem 3.2 (Existence in games with strategic substitutes)**.** *A public good provision game with strategic substitutes always has at least one Nash equilibrium.*

Proof. By [86, Theorem 5.2], for a given non-negative matrix $\mathbf{M}$, the corresponding LCP $(\mathbf{M}, \mathbf{q})$ has a solution for all $\mathbf{q}$ if and only if $m_{ii} > 0$. For a game with substitutes, $\mathbf{I} + \mathbf{G}$ is a non-negative matrix, and the diagonal entries are all 1. Therefore, for LCP (3.3), a solution (Nash equilibrium) always exists. □

We next consider the existence of Nash equilibria in games where agents' efforts are complements to their neighbors'. Let $\rho(\mathbf{G}) := \max\{|\lambda| \text{ s.t. } \mathbf{G}\mathbf{v} = \lambda\mathbf{v}\}$ denote the spectral radius of $\mathbf{G}$. Also, define the following classes of matrices.

Definition 3.2 (Z-matrix, L-matrix, S-matrix)**.**

- *A square matrix* $\mathbf{M}$ *is a* Z-matrix *if* $m_{ij} \leq 0, \forall i, j \neq i$.
- *A square matrix* $\mathbf{M}$ *is an* L-matrix *if it is Z-matrix and* $m_{ii} > 0, \forall i$.
- *A matrix* $\mathbf{M}$ *is an* S-matrix *if there exists* $\mathbf{x} \succ \mathbf{0}$ *such that* $\mathbf{M}\mathbf{x} \succ \mathbf{0}$.

Theorem 3.3 (Existence in games with strategic complements)**.** *For a public good provision game with strategic complements, if a Nash equilibrium exists for all* $\bar{\mathbf{q}}$*, i.e., for all payoff realizations, then it is unique. Specifically, the game has a Nash equilibrium if and only if* $\rho(\mathbf{G}) < 1$*.*

Proof. First, note that for this game, $\mathbf{I} + \mathbf{G}$ is an L-matrix. For an LCP $(\mathbf{M}, \mathbf{q})$, if $\mathbf{M}$ is an L-matrix, the LCP has at least one solution for all $\mathbf{q}$ if and only if $\mathbf{M}$ is an S-matrix; see [87, p. 282]. Therefore, the LCP (3.3) has a solution if and only if $\mathbf{I} + \mathbf{G}$ is an S-matrix. A Z-matrix is an S-matrix if and only if it is a P-matrix [101]. Therefore, the condition for existence and uniqueness in games with complements are the same. In other words, if a Nash equilibrium is guaranteed to exist, it is also unique.

Also, for a Z-matrix $\mathbf{G}$, $\mathbf{I}+\mathbf{G}$ is an S-matrix if and only $\rho(\mathbf{G}) < 1$ [11]. Therefore, a solution exists and is unique if and only if $\rho(\mathbf{G}) < 1$.[4] □

It is worth noting the difference between Theorems 3.2 and 3.3 and a previous result on the existence of Nash equilibria in concave n-person games. Rosen [111] shows that for an n-person game, if agents' payoffs are concave in their own effort, and agents' strategies are limited to a convex, closed, and bounded set, then the corresponding n-person game always has a Nash equilibrium [111, Theorem 1]. The latter assumption does not hold in the current model, as we allow an unbounded effort space $x_i \in \mathbb{R}_{\geq 0}$.

However, similar to [111], Theorem 3.2 concludes that for games of strategic substitutes, a Nash equilibrium always exists. In this case, each agent's strategy space can be effectively bounded by q_i, where $b_i'(q_i) = c_i$, i.e., agent i may exert effort lower than q_i (due to positive externalities from her neighbors), but will never exert an effort higher than q_i, as her marginal cost to do so will be higher than her marginal benefit. Thus in this case the existence result given by Theorem 3.2 is equivalent to that given in [111], though arrived at using a different methodology.

[4]The statement of Theorem 3.3 is similar to Theorem 1 in [28], which also uses an LCP formulation in the study of Nash equilibria on unweighted and undirected networks where agents have linear quadratic payoffs. This can be explained by observing that both games have best replies of the form (3.2), and hence have similar conclusions; cf. footnote 3.

For games of strategic complements on the other hand, a similar upper bound on agents' strategies does not exist. Specifically, when an agent i's neighbors increase their efforts, she will experience a negative externality, and will therefore increase her own level of effort to compensate for the lost benefit. As a result, agents' efforts can grow unbounded, and an equilibrium may not exist; the sufficient and necessary condition given in Theorem 3.3 thus goes beyond that considered in [111]. If the strategy spaces were bounded in this scenario, then agents would exert the upper bound effort, leading to the existence result of [111].

3.3.2 Comparison with existing results

We now show how existing results in [18, 7, 92, 85] on the uniqueness of the Nash equilibrium of public good provision games can be recovered as corollaries of Theorem 3.1. These comparisons also illustrate that some well-known matrices, namely, symmetric positive definite, strongly diagonally dominant, and (a subclass of) Z-matrices, belong to the family of P-matrices.

We begin with the uniqueness result of [18] on networks of symmetric relations. We note that [18] only states the sufficient condition; we also show the necessary condition in the following corollary using Theorem 3.1.

Corollary 3.1 (Uniqueness on symmetric networks [18])**.** *Consider a network with a symmetric dependence matrix* $\mathbf{G}$*. Then, if and only if* $|\lambda_{min}(\mathbf{G})| < 1$*, the Nash equilibrium is unique.*

Proof. By [86, Theorem 1.9] a square symmetric matrix is a P-matrix if and only if it is positive definite. Therefore, by Theorem 3.1, the Nash equilibrium is unique if and only if $\mathbf{I} + \mathbf{G}$ is positive definite, which occurs if and only if $|\lambda_{min}(\mathbf{G})| < 1$. □

The results of [18] are the first to show the importance of the lowest eigenvalue in determining outcomes of strategic interactions on networks, leading to several interesting insights on equilibria stability and network structure; we refer the interested reader to [18] for details.

It is also worth noting that Theorem 3.1 generalizes [18] on both symmetric and asymmetric matrices:

(i) Symmetric matrices: [18] uses the theory of potential games to show that a positive definite $\mathbf{I}+\mathbf{G}$ is a sufficient condition for uniqueness of the Nash equilibrium. Our result shows that this condition is necessary as well.

(ii) Asymmetric matrices: For directed, asymmetric graphs, the results of [18] apply if $|\lambda_{min}(\frac{\mathbf{G}+\mathbf{G}^T}{2})| < 1$; i.e., if $\mathbf{I} + \frac{\mathbf{G}+\mathbf{G}^T}{2}$ is positive definite. This is equivalent to $\mathbf{I} + \mathbf{G}$ being positive definite [86, Result 1.9]. In contrast, we only require that $\mathbf{I} + \mathbf{G}$ be a P-matrix, providing a more general (weaker) sufficient condition (as well as a necessary condition). This is because there exist (asymmetric) P-matrices that are not positive definite [86, Theorem 1.10]. Hence, positive definite matrices are in general a subset of P-matrices.

We next show that the uniquness result of [7] can also be recovered as a corollary of Theorem 3.1.

Corollary 3.2 (Uniqueness on networks with hidden complementarities [7])**.** *Let* $\mathbf{T}$ *be a Z-matrix such that* $\mathbf{T}(\mathbf{I} + \mathbf{G})$ *is both a Z-matrix and an S-matrix. Then, the Nash equilibrium is unique if and only if* $\mathbf{I} + \mathbf{G}$ *is an S-matrix. In particular, if* $\mathbf{G}$ *is a Z-matrix (i.e., a game with complementarities), the equilibrium is unique if and only if* $\rho(\mathbf{G}) < 1$.

Proof. By Theorem 3.1, we know that the Nash equilibrium is unique if and only if $\mathbf{I} + \mathbf{G}$ is a P-matrix. On the other hand, a matrix $\mathbf{I} + \mathbf{G}$ satisfying the conditions of the corollary is a *hidden Z-matrix* [101]. By [101, Theorem 1], a hidden Z-matrix is a P-matrix if and only if it is an S-matrix. Therefore, the Nash equilibrium is unique if and only if $\mathbf{I} + \mathbf{G}$ is an S-matrix. Finally, when $\mathbf{G}$ is a Z-matrix, $\mathbf{I} + \mathbf{G}$ is an S-matrix if and only $\rho(\mathbf{G}) < 1$ [11]. □

We also prove an alternative expression for Corollary 3.2.

Corollary 3.3. *If* $\mathbf{G}$ *is a Z-matrix, a unique Nash equilibrium exists if and only if* $|\lambda_{min}(\mathbf{G})| < 1$.

Proof. For a Z-matrix $\mathbf{G}$, $-\mathbf{G}$ is a non-negative matrix. Then, by the Perron-Frobenius theorem, $-\mathbf{G}$ has a positive eigenvalue equal to its spectral radius, $\lambda_{max}(-\mathbf{G}) = \rho(-\mathbf{G})$. Noting that $\rho(-\mathbf{G}) = \rho(\mathbf{G})$ and $\lambda_{max}(-\mathbf{G}) = -\lambda_{min}(\mathbf{G})$, we conclude that for Z-matrices, $\rho(\mathbf{G}) < 1$ if and only if $|\lambda_{min}(\mathbf{G})| < 1$. □

Comparing the above with Corollary 3.1, we conclude that the lowest eigenvalue of the dependence matrix has the key role in determining sufficient (and necessary) conditions for the uniqueness of Nash equilibria in (asymmetric) networks with complementarities (in addition to the symmetric networks and some subclasses of directed networks shown in [18]).

Finally, we show that the result of [92, 85] can also be recovered as a corollary of Theorem 3.1.

Corollary 3.4 (Uniqueness on strictly diagonally dominant networks [92, 85])**.** *If* $\mathbf{I} + \mathbf{G}$ *is strictly diagonally dominant, i.e.,* $\sum_i |g_{ij}| < 1, \forall i$*, there is a unique Nash equilibrium.*

Proof. We prove the theorem by showing that if $\mathbf{I}+\mathbf{G}$ is strictly diagonally dominant, then it is a P-matrix. This is because by the Gershgorin circle theorem, for a strictly diagonally dominant matrix with positive diagonal elements, all real eigenvalues are positive. Following a similar argument, all real eigenvalues of all sub-matrices of $\mathbf{I}+\mathbf{G}$ are also positive. Since the determinant of a matrix is the product of its eigenvalues, and as for real matrices, the complex eigenvalues appear in pairs with their conjugate eigenvalues, it follows that $\mathbf{I}+\mathbf{G}$, as well as all its square sub-matrices, have positive determinants. Therefore, $\mathbf{I} + \mathbf{G}$ is a P-matrix. The uniqueness then follows from Theorem 3.1. □

3.4 Efforts as node centralities

In this section, we focus on *interior* effort profiles of the public good provision game; these are outcomes in which all agents exert strictly positive efforts. We establish a connection between agents' actions at different effort outcomes and their

centralities in the dependence network. Using this connection, we can identify the effects of dependencies (outgoing edges in $\mathbf{G}$) and influences (incoming edges in $\mathbf{G}$), as well as walks of different length, on the efforts exerted by agents.

3.4.1 Existence and uniqueness of interior effort profiles

We first identify conditions under which a game with payoffs (3.1) has interior Nash equilibria and Pareto efficient effort profiles. We begin with a definition.

Definition 3.3 (Positive cone)**.** *The positive cone (or positive linear span) of a set of vectors* $\mathbf{v} = \{v_1, v_2, \ldots, v_n\}$ *is given by* $pos(\mathbf{v}) := \{\sum_i \alpha_i v_i |\ \alpha_i \geq 0, \forall i\}$.

For a Nash equilibrium (or a Pareto efficient effort profile) to be interior, the corresponding LCP (3.3) (or (3.8)) should have a solution with $\mathbf{x} \succeq \mathbf{0}, \mathbf{w} = \mathbf{0}$.[5]

Theorem 3.4 (Existence and uniqueness of interior effort profiles)**.** *The LCP* (3.3) *(or* (3.8)*) leads to an interior Nash equilibrium (or Pareto efficient effort profile) if and only if* $\bar{\mathbf{q}}$ *(or* $\mathbf{q}^{\boldsymbol{\lambda}}$*) is in the positive cone generated by the columns of* $\mathbf{I} + \mathbf{G}$*. Furthermore, the interior effort profile (when one exists) is unique if and only if* $\mathbf{I} + \mathbf{G}$ *is a P-matrix.*

Proof. Solving the LCP (3.3) for interior solutions is equivalent to finding a solution to:

$$(\mathbf{I} + \mathbf{G})\mathbf{x} = \bar{\mathbf{q}}, \quad \mathbf{x} \succeq \mathbf{0} \ .$$

The theorem then follows from Definition 3.3. The same argument applies to finding interior Pareto efficient profiles using (3.8). It is also worth mentioning that given $\mathbf{G}$, non-interior solutions will necessarily exist for some $\mathbf{q} \in \mathbb{R}^n$, as we need at least $n + 1$ vectors to positively span $\mathbb{R}^n$ [31, Theorem 3.8]. In other words, as expected, there is no network structure for which solutions are guaranteed to be interior.

Finally, when $\mathbf{I} + \mathbf{G}$ is a P-matrix, the LCPs in (3.3) and (3.8) will have unique solutions [86, Theorem 4.2]. Therefore, under this condition, when a solution with $\mathbf{x} \succeq \mathbf{0}, \mathbf{w} = \mathbf{0}$ exists, it is also the unique solution. □

[5] With a slight abuse of terminology, we consider solutions with $x_i = 0, w_i = 0$ to be interior as well.

We now proceed to establishing a connection between interior effort profiles (when they exist) and agents' centralities in their interaction network, starting with an overview of centrality measures.

3.4.2 Alpha-centrality: An overview

Centrality measures have been used extensively in the graph theory and network analysis literatures as indicators of importance of nodes in their interaction network. Some of these measures (e.g., degree centrality) take into account the number of connections of a node in determining her centrality. In contrast, another class of measures (e.g. eigenvector centrality) account for the importance of the connections as well, such that a node's centrality is (recursively) related to those of her neighbors. *Alpha-centrality*, considered herein, belongs to the latter family. This measure was introduced by Bonacich and Lloyd in [14], mainly as an extension of eigenvector centrality that is applicable to networks of asymmetric relations.

Formally, denote the centrality of node i by x_i. Let $\mathbf{G}$ be the adjacency matrix of a network, where g_{ij} determines the dependence of node i on node j. Then, the eigenvector centrality of nodes will be proportional to $\mathbf{Gx}$. Alpha-centrality generalizes this measure by allowing the nodes to additionally experience an exogenous source of centrality $\mathbf{e}$, such that,

$$\mathbf{x} = \alpha\mathbf{G}\mathbf{x} + \mathbf{e} \ .$$

Here, α is a constant that determines a tradeoff between the endogenous (eigenvector) and exogenous centrality factors. The nodes' alpha-centralities are therefore given by

$$c_{\text{alpha}}(\mathbf{G}, \alpha, \mathbf{e}) = (\mathbf{I} - \alpha\mathbf{G})^{-1}\mathbf{e} \ . \tag{3.9}$$

On the interpretation of α: As mentioned above, α determines the tradeoff between the endogenous and exogenous sources of centrality. We will now illustrate that powers of α also appear as weights of walks of different length in determining

nodes' centralities.

We do so by noting the connection between alpha-centrality and the measure proposed by Katz [68]. Katz centrality defines a weighted sum of powers of the adjacency matrix $\mathbf{G}$ as an indicator of nodes' importance; intuitively, longer paths are weighed differently (and often less favorably) in determining nodes' centralities. Formally, Katz' measure is given by

$$c_{\text{katz}}(\mathbf{G}, \alpha) = (\sum_{i=1}^{\infty} \alpha^i \mathbf{G}^i)\mathbf{1} \ ,$$

where α is an attenuation factor. In particular, if $\alpha < \frac{1}{|\lambda_{max}(\mathbf{G})|}$, the infinite sum converges, so that,

$$(\sum_{i=1}^{\infty} \alpha^i \mathbf{G}^i)\mathbf{e} = (-I + (I - \alpha\mathbf{G})^{-1})\mathbf{e} \ . \tag{3.10}$$

Comparing (3.9) and (3.10), we conclude that the parameter α of alpha-centrality can be similarly interpreted as a weight assigned to the walks of different length in determining the effect of endogenous centralities on the overall centrality of a node. [6]

[6] Alpha centrality is also similar to the measure introduced earlier by Bonacich in his seminal work [13]. Formally, Bonacich's centrality is defined as $c_{\text{bonacich}}(R, \beta, \alpha) = \beta(I - \alpha R)^{-1} R\mathbf{1}$. Here, R is a symmetric matrix of relationships, with main diagonal elements equal to zero. The parameter β only affects the length of the final measures, and has no network interpretation. The parameter α on the other hand can be positive or negative, and determines the extent and direction of influences. On symmetric matrices, Katz' measure is essentially equivalent to Bonacich centrality; in fact, $c_{\text{katz}}(R, \alpha) = \sum_{i=1}^{\infty} \alpha^i R^i \mathbf{1} = \alpha c_{\text{bonacich}}(R, \alpha, 1)$. To summarize, taking the three measures on a symmetric matrix A, and setting $\mathbf{e} = \mathbf{1}$ for the alpha-centralities, we have:

$$c_{\text{alpha}}(A, \alpha, \mathbf{1}) = 1 + \alpha c_{\text{bonacich}}(A, \alpha, 1) = 1 + c_{\text{katz}}(A, \alpha) \ .$$

Therefore, in essence, alpha-centrality generalizes Bonacich and Katz centralities, allowing for vectors of exogenous status $\mathbf{e}$. Using the above equivalence, we can show that our characterization of Nash equilibrium based on alpha-centralities in Theorem 3.5, and the Nash-Bonacich linkage established in [8] are equivalent (see also footnote 3).

3.4.3 A centrality-effort connection

We now establish the connection between agents' efforts at interior profiles, and their alpha-centralities in the interaction network.

Theorem 3.5 (Centrality-effort connection)**.** *(i) Consider an interior Nash equilibrium* $\mathbf{x}^*$*. Then,*

$$\mathbf{x}^* = c_{alpha}\left(\mathbf{G}, -1, \bar{\mathbf{q}}\right) \ ,$$

where $\bar{\mathbf{q}}$ *is such that* $b_i'(\bar{q}_i) = c_i$*.*

(ii) Consider an interior Pareto efficient effort profile $\mathbf{x}^{\boldsymbol{\lambda}}$*. Then,*

$$\mathbf{x}^{\boldsymbol{\lambda}} = c_{alpha}\left(\mathbf{G}, -1, \mathbf{q}^{\boldsymbol{\lambda}}\right) \ ,$$

where $\mathbf{q}^{\boldsymbol{\lambda}}$ *is such that* $b_i'(q_i^{\boldsymbol{\lambda}}) = c_{alpha,i}(\boldsymbol{\Lambda}^{-1}\mathbf{G}^T\boldsymbol{\Lambda}, -1, \mathbf{c})$*.*

Proof. (i) An interior Nash equilibrium is a solution to LCP (3.3) with $\mathbf{w} = \mathbf{0}$, i.e.,

$$(\mathbf{I} + \mathbf{G})\mathbf{x} = \bar{\mathbf{q}}, \quad \mathbf{x} \succeq \mathbf{0} \ .$$

Therefore, when such solution exists, $\mathbf{x}^* = (\mathbf{I} + \mathbf{G})^{-1}\bar{\mathbf{q}}$. Comparing this expression with (3.9) establishes the connection.

(ii) An interior Pareto efficient profile with weights $\boldsymbol{\lambda}$ is a solution to LCP (3.8) with $\mathbf{w} = \mathbf{0}$, i.e.,

$$(\mathbf{I} + \mathbf{G})\mathbf{x} = \mathbf{q}^{\boldsymbol{\lambda}}, \quad \mathbf{x} \succeq \mathbf{0} \ .$$

Therefore, when such solution exists, $\mathbf{x}^{\boldsymbol{\lambda}} = (\mathbf{I} + \mathbf{G})^{-1}\mathbf{q}^{\boldsymbol{\lambda}}$. Also, by definition, we know that $\mathbf{q}^{\boldsymbol{\lambda}}$ satisfies $b_i'(q_i^{\boldsymbol{\lambda}}) = \left((\mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{G}^T\boldsymbol{\Lambda})^{-1}\mathbf{c}\right)_i$. Comparing these expressions with (3.9) establishes the connection. □

The connection established in Theorem 3.5 leads to several interesting insights. Recall that an entry $g_{ij} \neq 0$ in $\mathbf{G}$ indicates that agent i's payoff depends on agent j's action; we therefore refer to $\mathbf{G}$ as the *dependence* matrix. On the other hand, an

entry $g_{ji} \neq 0$ in the $\mathbf{G}^T$ indicates that agent j's effort influences agent i's payoff. We will therefore refer to $\mathbf{G}^T$ as the *influence* matrix.

Perceived costs at different effort profiles: comparing parts (i) and (ii) of Theorem 3.5, we observe that the only difference when determining nodes' efforts is in the corresponding vectors of exogenous centralities. These vectors are determined by efforts at which agents' marginal benefits equal their (perceived) marginal costs. At the Nash equilibrium, each agent acts independently and perceives only her own cost of effort, leading to $b_i'(\bar{q}_i) = c_i$. On the other hand, for Pareto efficient solutions to emerge, the cost perceptions are modified according to agents' positions in the network, as well as the importance placed on each agent's welfare, as determined by λ_i. Consequently, both $\mathbf{G}$ and $\boldsymbol{\lambda}$ play a role in determining agents' perceived marginal costs, leading to $b_i'(q_i^{\boldsymbol{\lambda}}) = c_{\text{alpha},i}(\boldsymbol{\Lambda}^{-1}\mathbf{G}^T\boldsymbol{\Lambda}, -1, \mathbf{c})$.

Effects of dependencies: consider agents' dependencies (outgoing edges in the network). We observe that by the definition of alpha-centrality (3.9), the matrix of dependencies $\mathbf{G}$ shapes the endogenous component of the centrality measure, determining a node's centrality as a function of her neighbors' centrality. Similarly, $\mathbf{G}$ in Theorem 3.5 indicates that the dependence of an agent on her neighbors (and the efforts they have exerted) will shape her final effort. Note also that this is the case for both Nash equilibria (part (i)) and Pareto efficient efforts (part (ii)): an agent benefits of any neighbor's effort regardless of the solution concept, or the mechanism or negotiations through which the effort profile is implemented.

Effects of influences: we further observe the effects of agents' influences (incoming edges in the network) on the outcomes of their strategic interactions. The matrix of influences $\mathbf{G}^T$ appears when determining the perceived costs of agents in Pareto efficient solutions. Intuitively, an agent with higher influence on others (as determined by her alpha-centrality in the network of influences) will have a lower perceived marginal cost, hence a higher exogenous centrality (due to concavity of $b_i(\cdot)$), which in turn increases a node's alpha-centrality (i.e, her level of effort/contribution). Note also that the matrix of influences $\mathbf{G}^T$ does not appear in the characterization of the Nash equilibria in Theorem 3.5. This is because at a Nash equilibrium, an

agent only accounts for her own marginal costs when selecting an effort level.

Alternating effect – the role of α: most interestingly, we note that the alpha parameter of all the alpha-centralities in Theorem 3.5 is $\alpha = -1$. Recall that, as shown in Section 3.4.2, α^k is a weight associated with a walk of length k in determining an agent's centrality.[7] Let $i_0, i_1, \ldots, i_k$ be the agents along this walk. Then, for walks of odd length, $\alpha = -1$ induces a sign reversal on the weight $g_{i_0 i_1} g_{i_1 i_2} \ldots g_{i_{k-1} i_k}$ of the walk. For walks of even length on the other hand, $\alpha = -1$ leaves the sign on the weight associated with the walk unchanged.

To better highlight the intuition behind this observation, consider a network of substitutes, i.e., $g_{ij} \geq 0, \forall i, j$. Consider a walk of length one by choosing a neighbor j of i. If agent j increases her effort, agent i benefits from the positive externality of j's increased effort, and can in turn reduce her effort. Thus, changes along this walk of odd length negatively affect agent i's effort decision; this is consistent with $(-1)^1 g_{ij} < 0$. Now, consider a neighbor k of j. Therefore, there is a walk of length 2 from i to k. By the same argument as above, if k increases her effort, j will decrease her effort in response. To compensate for the lost externality, agent i will now have to increase her own effort. Thus, a change along this walk of even length positively affects agent i's effort decision, which is again consistent with $(-1)^2 g_{ij} g_{jk} > 0$. The same argument extends to walks of longer lengths.

Alternatively, consider a network of complementarities, i.e., $g_{ij} \leq 0, \forall i, j \neq i$. Again, consider a walk of length one from i to j. If agent j increases her effort, agent i's benefit is reduced, and so she will increase her level of effort in response. Thus, a change along this walk of odd length positively affects agent i's effort decision, which is consistent with $(-1)^1 g_{ij} > 0$. Now, consider a walk of length 2 from i to k (j's neighbor). By the same argument as above, if k increases her effort, j will increase her effort in response, and so agent i will have to increase her effort as well. Thus, the change along this walk of even length also positively affects agent i's effort decision; this is again consistent with $(-1)^2 g_{ij} g_{jk} > 0$. The same argument extends

[7] Given $\alpha = -1$, the condition $\alpha < \frac{1}{|\lambda_{max}(\mathbf{G})|}$ holds for all adjacency matrices $\mathbf{G}$. Therefore, the alpha-centralities can be interpreted as the limit of a weighted sum of powers of the adjacency matrix, and the interpretation of α as a weight on walks of different length in applicable.

to walks of longer lengths.

3.4.4 Numerical examples

We illustrate the centrality-effort connection through some examples.

Example 3.2 (Alternating effect of α)**.** Consider the three node network of Fig. 3.1, and a public good provision game of strategic substitutes (i.e., $g_{12}, g_{13}, g_{21} > 0$) played on this network. Set $g_{12} = g_{21} = 0.2$. Let agents' payoffs be given by

$$u_i(\mathbf{x}; \mathbf{G}) = 1 - \exp(-x_i - \sum_{j \neq i} g_{ij} x_j) - \frac{1}{e} x_i \ .$$

Consider the edge between agents 1 and 3. Assume we increase the weight g_{13}, and want to know how this change affects the efforts of the agents at the Nash equilibrium. The results are given in the bottom two networks of Fig. 3.1, and can be explained as follows.

• Agent 1: the edge $1 \to 3$ is on all the outgoing walks of *odd* length from node 1. Increasing g_{13} increases the weights of these walks. However, given $\alpha = -1$, each walk weight is multiplied by $(-1)^{2k+1} = -1$ (this is the alternating effect induced by α). Therefore, the increase in g_{13} should negatively affect agent 1's effort decision, leading her to decrease her effort levels in response.

• Agent 2: the edge $1 \to 3$ is on (some of) the outgoing walks of *even* length from node 1. Increasing g_{13} increases the weights of these walks. Given $\alpha = -1$, each walk weight is multiplied by $(-1)^{2k} = 1$. Therefore, the increase in g_{13} should positively affect agent 2's effort decision, leading her to increase her effort levels in response.

• Agent 3: we are changing the weight of an *incoming* edge to agent 3. By Theorem 3.5, only outgoing edges and walks affect the agent's effort decisions at the Nash equilibrium. Therefore, we expect agent 3's effort to remain unchanged.

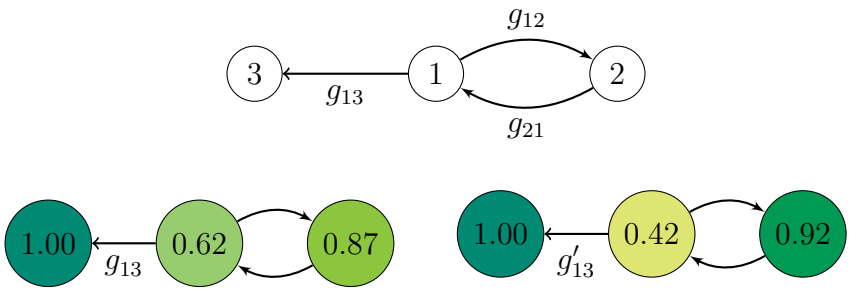

Figure 3.1: Alternating effect of α is illustrated by increasing $g_{13} = 0.2$ to $g'_{13} = 0.4$ (bottom left to bottom right) in this network (Example 3.2). Numbers inside nodes at the bottom networks indicate efforts at the Nash equilibrium.

Example 3.3 (Effects of incoming edges and perceived costs)**.** Consider the 4 agent network of Fig. 3.2. Agents' payoffs are given by

$$u_1(\mathbf{x}; \mathbf{G}) = 1 - \exp(-x_1 - g_o \sum_{j \neq i} x_j) - \tfrac{1}{e} x_1 \ ,$$
$$u_k(\mathbf{x}; \mathbf{G}) = 1 - \exp(-x_k - g_i x_1) - \tfrac{1}{e} x_k \ , \ k \neq 1 \ .$$

We consider the socially optimal effort profile in this network, i.e., $\mathbf{x}^* := \arg\max_{\mathbf{x} \geq 0} \sum_i u_i(\mathbf{x})$. This corresponds to a Pareto efficient solution of (3.5) with weights $\boldsymbol{\lambda} = \mathbf{1}$. Thus, according to Theorem 3.5, the vector of perceived costs of agents at this outcome is given by $(\mathbf{I} + \mathbf{G}^T)^{-1}\mathbf{c}$.

Fix $g_o = 0.2$. To illustrate the effect of incoming edges on agents' perceived costs, and consequently their efforts, we increase g_i from 0.2 to 0.3. The vector of perceived costs of agents will change from $[0.17, 0.33, 0.33, 0.33]$ to $[0.04, 0.36, 0.36, 0.36]$. Therefore, the perceived cost of agent 1 (the center) decreases considerably when her influence on others increases, leading her to exert higher effort as a result. Furthermore, as the center invests more, the leaves now have an incentive to decrease their investment (alternating effect of α). These effects combined lead the center (leaves) to exert higher (lower) effort when the weight of incoming edges, g_i, increases.

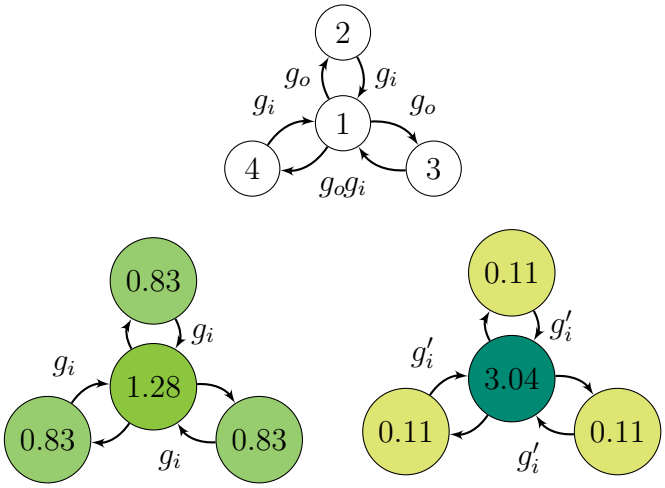

Figure 3.2: Effect of incoming edges on perceived costs is illustrated by changing $g_i = 0.2$ to $g'_i = 0.3$ (bottom left to bottom right) in this network (Example 3.3). Numbers inside nodes indicate efforts exerted at the socially optimal outcome.

3.5 Extension to coalitions

In this section, we extend the results of Section 3.4 to effort profiles that emerge when agents belong to different coalitions. For this analysis, we distinguish between *excludable* and *non-excludable* public goods. With excludable goods, each coalition may choose to exclude other coalitions from experiencing the externalities of its produced good. If this is the case, each coalition can be studied in isolation, and therefore the results of the previous sections will be directly applicable. For non-excludable goods on the other hand, such separation is not possible; each coalition needs to further account for the externalities from and on other coalitions. Throughout this section, we are interested in the provision of such non-excludable goods. We do not explicitly model coalition formation or stability; we assume each coalition has emerged through either collaboration or appropriate incentive mechanisms. We present a centrality-effort connection, and the corresponding intuition, for the effort profiles emerging as the result of strategic interactions of such coalitions.

3.5.1 Semi-cooperative equilibrium

Let agents form K coalitions, denoted by the collection of disjoint sets $\mathcal{C} := \{\mathcal{C}_1, \ldots, \mathcal{C}_K\}$, such that $\mathcal{C}_1 \cup \ldots \cup \mathcal{C}_K = \mathcal{N}$. We refer to $\mathcal{C}$ as the coalition partition. Individual agents are allowed to form their own one-member coalition. The effort profile emerging from the interactions of these coalitions is affected by both *intragroup* and *intergroup* decisions.

Intragroup decisions refer to those adopted within each coalition. Specifically, we assume that the members within a coalition $\mathcal{C}_i$ agree (either cooperatively or through the implementation of an incentive mechanism) on a vector of welfare weights $\boldsymbol{\lambda}^i := \{\lambda_k^i, \text{ for } k \in \mathcal{C}_i\}$, and implement the corresponding Pareto efficient solution in (3.5), i.e.,

$$\bar{\mathbf{x}}_{\mathcal{C}_i}^{\boldsymbol{\lambda}^i} = \arg\max_{\mathbf{x}_{\mathcal{C}_i} \geq \mathbf{0}} \quad \sum_{k \in \mathcal{C}_i} \lambda_k^i u_k(\mathbf{x}_{\mathcal{C}_i}, \mathbf{x}_{\mathcal{N}\setminus\mathcal{C}_i}) \ , \tag{3.11}$$

where $\mathbf{x}_{\mathcal{N}\setminus\mathcal{C}_i}$ denotes the efforts of agents outside the coalition. The profile $\bar{\mathbf{x}}_{\mathcal{C}_i}^{\boldsymbol{\lambda}^i}$ is therefore a Pareto efficient effort profile with weights $\boldsymbol{\lambda}^i$ for the agents in $\mathcal{C}_i$.

At the intergroup level, each coalition is viewed as a *super-agent*, playing a non-cooperative game with other coalitions/super-agents, and best-responding to their decisions. The resulting equilibrium effort profile $\bar{\mathbf{x}}_{\mathcal{C}}^{\boldsymbol{\lambda}} := \left(\bar{\mathbf{x}}_{\mathcal{C}_1}^{\boldsymbol{\lambda}^1}, \ldots, \bar{\mathbf{x}}_{\mathcal{C}_K}^{\boldsymbol{\lambda}^K}\right)$ is the Nash equilibrium among these super-agents, i.e., a solution to the system of equations determined by (3.11). We refer to $\bar{\mathbf{x}}_{\mathcal{C}}^{\boldsymbol{\lambda}}$ as a *semi-cooperative* equilibrium for coalition partition $\mathcal{C}$ with weights $\boldsymbol{\lambda}$.[8]

Similar to the characterization of interior Pareto efficient outcomes in Section 3.2.2, the problem of characterizing interior semi-cooperative equilibria can be formulated as an LCP. Assume agents are indexed in an order consistent with the index of their coalition memberships. Also, to simplify notation, denote the semi-cooperative equilibrium by $\bar{\mathbf{x}}$; dependence on the coalition partition $\mathcal{C}$ and the weights

[8] A semi-cooperative equilibrium is an "equilibrium" in the sense that, assuming binding coalition memberships, the effort profile resulting from intergroup interactions is the fixed-point of a best-response mapping. It has the limitation that it does not preclude the possibility of agents moving to other coalitions if the memberships are not binding or appropriately incentivized.

$\boldsymbol{\lambda}$ is implied. Then, the first order condition on (3.11) with respect to $x_i,\ i \in \mathcal{C}_i$, implies that at the interior Pareto efficient solution, the following should hold:

$$\lambda_i b_i'(\bar{x}_i + \sum_{j \in \mathcal{N}_i} g_{ij}\bar{x}_j) + \sum_{k \in \mathcal{C}_i,\ \text{s.t. } i \in \mathcal{N}_k} \lambda_k g_{ki} b_k'(\bar{x}_k + \sum_{j \in \mathcal{N}_k} g_{kj}\bar{x}_j) = \lambda_i c_i,\ \forall i\ .$$

Define $q_i^{\mathcal{C},\boldsymbol{\lambda}}$ as the effort levels at which

$$b_i'(q_i^{\mathcal{C},\boldsymbol{\lambda}}) = \left((\mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{G}_{\mathcal{C}}^T\boldsymbol{\Lambda})^{-1}\mathbf{c}\right)_i\ .$$

Note that the only difference of these efforts with the $q_i^{\boldsymbol{\lambda}}$ defined for Pareto efficient outcomes is in the *coalition-modified* dependence matrix $\mathbf{G}_{\mathcal{C}}$, which is defined as follows: for each row k corresponding to an agent in coalition $\mathcal{C}_i$, set the entries $g_{kl}, \forall l \notin \mathcal{C}_i$ to zero. This matrix is therefore equivalent to the dependence matrix of a network obtained by removing all edges between coalitions.

Then, finding an interior semi-cooperative equilibrium $\bar{\mathbf{x}}$ is equivalent to finding a solution with $\mathbf{w} = \mathbf{0}$ to the following LCP:

$$\begin{aligned} \mathbf{w} - (\mathbf{I} + \mathbf{G})\mathbf{x} &= -\mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}\ , \\ \mathbf{w} \succeq \mathbf{0}\ ,\quad \mathbf{x} &\succeq \mathbf{0}\ , \\ \mathbf{w}^T\mathbf{x} &= 0\ . \end{aligned} \tag{3.12}$$

Using a similar procedure as Section 3.4.1, such profile exists under the following condition.

Theorem 3.6 (Existence and uniqueness of interior semi-cooperative equilibria)**.** *The public good provision game has an interior semi-cooperative equilibrium $\bar{\mathbf{x}}_{\mathcal{C}}^{\boldsymbol{\lambda}}$ if and only if the corresponding $\mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}$ is in the positive cone generated by the columns of $\mathbf{I} + \mathbf{G}$. Furthermore, the interior effort profile (when one exists) is unique if and only if $\mathbf{I} + \mathbf{G}$ is a P-matrix.*

3.5.2 A centrality-effort connection

We now present a centrality-effort characterization of interior semi-cooperative equilibria.

Theorem 3.7 (Centrality-effort connection for semi-cooperative equilibria)**.** *Consider an interior semi-cooperative equilibrium* $\bar{\mathbf{x}}^{\boldsymbol{\lambda}}_{\mathcal{C}}$*. Then,*

$$\bar{\mathbf{x}}^{\boldsymbol{\lambda}}_{\mathcal{C}} = c_{alpha}\left(\mathbf{G}, -1, \mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}\right) \ ,$$

where $\mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}$ *is such that* $b_i'(q_i^{\mathcal{C},\boldsymbol{\lambda}}) = c_{alpha,i}(\boldsymbol{\Lambda}^{-1}\mathbf{G}_{\mathcal{C}}^T\boldsymbol{\Lambda}, -1, \mathbf{c})$.

Proof. An interior semi-cooperative equilibrium for coalition partition $\mathcal{C}$ and weights $\boldsymbol{\lambda}$ is a solution to LCP (3.12) with $\mathbf{w} = \mathbf{0}$, i.e.,

$$(\mathbf{I} + \mathbf{G})\mathbf{x} = \mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}, \quad \mathbf{x} \succeq \mathbf{0} \ .$$

Therefore, when such solution exists, $\bar{\mathbf{x}}^{\boldsymbol{\lambda}}_{\mathcal{C}} = (\mathbf{I} + \mathbf{G})^{-1}\mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}$. Also, by definition, we know that $\mathbf{q}^{\mathcal{C},\boldsymbol{\lambda}}$ satisfies $b_i'(q_i^{\mathcal{C},\boldsymbol{\lambda}}) = \left((\mathbf{I} + \boldsymbol{\Lambda}^{-1}\mathbf{G}_{\mathcal{C}}^T\boldsymbol{\Lambda})^{-1}\mathbf{c}\right)_i$. Comparing these expressions with (3.9) establishes the connection. □

The implications of the centrality-effort connection on effects of incoming and outgoing edges and the alternating effect induced by $\alpha = -1$ are applicable to the characterization of Theorem 3.7 as well. The main difference resulting from the formation of coalitions is the following.

The effect of coalitions – benefiting from dependencies and ignoring influences: with non-excludable goods, an agent can benefit from the externalities of the effort exerted by her neighbor, whether or not that neighbor is a member of the agent's coalition. Consequently, the alpha-centralities (i.e, efforts) of agents are calculated on the full network of dependencies $\mathbf{G}$.

However, recall that the perceived costs of each agent are affected by the influence of the agent on those with whom she is cooperating to implement a Pareto efficient effort profile. The agents in a coalition account for their influences on others in their

group, but disregard their influence on all other agents. Therefore, agents' perceived costs $c_{\text{alpha},i}(\mathbf{\Lambda}^{-1}\mathbf{G}_{\mathcal{C}}^{T}\mathbf{\Lambda}, -1, \mathbf{c})$, while again evaluated on the network of influences (i.e., the transpose of the dependence matrix), are now evaluated on a coalition-modified matrix of influences $\mathbf{G}_{\mathcal{C}}^{T}$. In other words, when determining their perceived costs, agents act as if the dependence network is one in which all edges between coalitions are removed.

3.6 Conclusion

We studied the provision of public goods on a network of strategic agents. We identified a necessary and sufficient condition on the dependence matrix of the network that guarantees the uniqueness of the Nash equilibrium in these games. Our condition unifies (and strengthens) existing results in the literature. We also identified necessary and sufficient conditions for existence of Nash equilibria in subclasses of games that lie at the two extremes of our model, namely games of strategic complements and games of strategic substitutes. An interesting direction of future work is to identify similar conditions for a general model of games on networks, and in particular, games with non-linear best replies.

We further presented a graph theoretical characterization of different interior effort outcomes, namely, the Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria, in terms of agents' alpha-centralities in their dependence network. Using this characterization, we were able to identify the effects of incoming edges, outgoing edges, and coalitions, as well as an alternating effect over walks of different length in the network. An interesting direction of future work is to use this connection for conducting comparative statics (e.g., the effects of adding/removing links), as well as for the design of targeted tax/subsidy policies that can incentivize the improved provision of the public good.

Chapter 4

Inter-temporal Incentives in Security Information Sharing Agreements

4.1 Introduction

In this chapter, we take a game-theoretic approach to understand firms' behavior and (dis)incentives in a particular class of security games, namely, security information sharing agreements. Improving the ability of analyzing cyber-incidents, and ensuring that the results are shared among organizations and authorities in a timely manner, has received increased attention in the recent years by governments and policy makers. This is because the availability of information on previous cyber-incidents can lead to better protection of the national infrastructure against potential cyber-attacks, allow organizations to invest in the most effective preventive and protective measures, and protect consumer rights. However, there is a conflict between individual and social goals in these agreements: despite the benefits of making such information available, the associated disclosure costs (e.g., drop in market value and loss of reputation) act as a disincentive for firms' full disclosure.

To capture this conflict, we model security information sharing agreements as an N-person prisoner's dilemma (NPD) game. In an NPD, there will be no information sharing at the state of equilibrium, as also predicted by similar game-theoretic models that consider one-shot information sharing games (see Section 4.1.1). 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 information sharing agreements) as remedies for encouraging information disclosure.

We take a different approach and account for the repeated nature of these agreements to propose the design of inter-temporal incentives that lead sufficiently patient firms to cooperate on information sharing. It is 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 [80]; conditions under which such cooperation is possible are known as *folk theorem*. The possibility of achieving efficient outcomes however depends on whether the monitoring of other participants' actions is perfect or imperfect, and private or public. In particular, for information sharing games, each firm or an outside monitor can (at best) only imperfectly assess the honesty and comprehensiveness of the shared information. Accordingly, we model these agreements as repeated games with imperfect monitoring, and consider two possible monitoring structures for these games.

First, we analyze the role of a rating/assessment system in providing an imperfect *public signal* about the quality of firms' reports in the agreement. We show that for the proposed NPDs equipped with a simple monitoring structure, the folk theorem of [45] holds in the repeated game, therefore making it possible to design appropriate inter-temporal incentives to support cooperation. We illustrate the construction of these incentives through an example, and discuss the effects of the monitoring accuracy on this construction.

We then consider the design of such incentives in the absence of a pubic monitoring system. Specifically, we assume that the firms have access to a communication platform, through which they are allowed to report their *private beliefs* on whether other firms are adhering to the agreement. We show that given a simple imperfect private monitoring structure by each firm, the folk theorem of [67] will be applicable to our proposed NPDs, again enabling the design of appropriate incentives for information sharing.

4.1.1 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.

The work by [98] and [73] consider the effects of security breach reporting *between firms and an authority*. The authors of [98] show that if the availability of shared information[1] 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, [73] 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.

Game-theoretic models of information sharing *among firms* have been proposed in [50] and [47]. Gordon et al. [50] show that, if security information from a partner firm is a *substitute* to a firm's own security expenditure, 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 an Information Sharing and Analysis Center (ISAC) for the losses of other members. Gal-Or and Ghose [47] on the other hand allow information sharing to be a *complement* to the firm's own security expenditure, as it may increase consumer confidence in a firm that is believed to take steps towards securing her system. 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.

In this chapter, similar to [50, 98, 73], we assume disclosure costs to be higher than potential demand-side benefits, therefore similarly predicting a lack of voluntary information sharing at equilibrium. Our proposed approach of considering the effects

[1]Firms' incentives for information disclosure or the mechanisms for ensuring breach disclosure have not been modeled in [98].

of repeated interactions as an incentive solution is however different from those proposed in the aforementioned literature, as they consider one-shot information sharing games.

4.1.2 Chapter contributions

The contributions of this chapter are the following:

- This chapter proposes the design of inter-temporal incentives for supporting cooperative behavior in security information sharing agreements. To this end, we model firms' interactions as an N-person prisoner's dilemma game and equip it with a simple monitoring structure.
- It illustrates the role of a public rating/assessment system in providing imperfect public monitoring, leading to coordination on cooperation in information sharing agreements.
- It establishes the possibility of sustaining cooperative behavior in the absence of a public monitor, by introducing a platform for communication among firms through which they can exchange their beliefs on others' adherence to the agreement.

4.1.3 Chapter organization

Section 4.2 presents the model and proposed monitoring structure for information sharing games. Section 4.3 discusses the role of a public monitoring system in the design of inter-temporal incentives. In Section 4.4, we illustrate the design of such incentives based on private observations and communication among firms. Section 4.5 concludes the chapter.

4.2 Information sharing games

4.2.1 The stage game

Consider N (symmetric) firms participating in an information sharing agreement (e.g. firms within an ISAC). Each firm can choose a level of expenditure in security measures to protect her infrastructure against cyber incidents. Examples include implementing an intrusion detection system, introducing employee education initiatives, and installing and maintaining up-to-date security software. We assume these measures are implemented independently of the outcome of the sharing agreement, and focus solely on firms' information sharing decisions.[2]

The information sharing agreement requires each firm i to share her security information with other participating firms. This disclosure can include information on both successful and failed attacks, as well as effective breach prevention methods and the firm's adopted security practices. A firm i should therefore decide whether to fully and honestly disclose such information. We denote the decision of firm i by $r_i \in \{0, 1\}$, with $r_i = 0$ denoting (partially) concealing and $r_i = 1$ denoting (fully) disclosing.[3] Denote the number of firms adopting a full disclosure decision by x; i.e., $x := |\{i|\ r_i = 1\}|$.

A decision of $r_i = 1$ is beneficial (to firms $j \neq i$, and also firm i herself) for the following reasons. On one hand, the disclosed information can allow other firms $j \neq i$ to leverage the acquired information to protect themselves against ongoing attacks and to adopt better security practices. Aside from this security-related implications, information disclosure $r_i = 1$ may further provide a competitive advantage to firms

[2]This assumption is adopted for two reasons. First, this allows us to focus only on firms' incentives for information sharing. More importantly, we assume the information shared by firm i to be a *substitute* to firm j's investment. That is, firm j can decrease her security expenditure when she receives information from firm i. This possible reduction in the positive externality from j's investments may therefore result in further disincentives for firm i for sharing her security information. We therefore remove these effects by decoupling the decisions on information sharing and security efforts, and assume fixed security expenditures. Analyzing the interplay of investment and sharing decisions remains a direction of future work.

[3]The results and intuition obtained in the following sections continue to hold when firms can choose one of finitely many disclosure levels, given an appropriate extension of utilities and monitoring structure.

$j \neq i$. This is because a firm j can increase her share of the market by strategically leveraging the attained information to attract a competitor i's customers. Finally, sharing of security information may be beneficial to firm i herself as well (especially when many other firms are disclosing as well), as it may garner trust from potential partners and customers. We denote all such applicable *information gains* to a firm, as a function of firm i's decision and the number of *other* firms making a full disclosure decision, by $G(r, z) : \{0, 1\} \times \{0, 1, \ldots, N-1\} \to \mathbb{R}_{\geq 0}$, with $G(0, 0) = 0$. We assume that given r, $G(r, \cdot)$ is increasing in z, the number of other firms sharing their information.

Despite the benefits that adopting $r_i = 1$ can have for all participating firms, sharing of security information may not be in firm i's interest; this disincentive for full disclosure is due to the associated costs. These costs include the man-hours spent in documenting and reporting security information, as well as potential losses in reputation, business opportunities with potential collaborators, stock market prices, and the like, following the disclosure of a breach or existing security flaws. In addition, it may be in i's interest to conceal methods for preventing ongoing threats, predicting that an attack on the competitor j will result in j's customers switching to i's products/services, increasing firm i's profits. Consequently, such potential market loss or competitors' gain in sales can further deter firms from adhering to information sharing agreements. We denote all these associated *disclosure costs* by $L(r, z) : \{0, 1\} \times \{0, \ldots, N-1\} \to \mathbb{R}^+$, with $L(0, 0) = 0$, where the cost can potentially depend on how many other firms, z, are disclosing their security information.

We can now define the utility of each user based on her disclosure decision. Given the number of firms that are sharing, x, and substituting $z = x - \mathbb{1}\{r_i = 1\}$ (where $\mathbb{1}(\cdot)$ denotes the indicator function), we define the following utilities for the cooperators ($r_i = 1$) and deviators ($r_i = 0$):

$$\begin{aligned} \text{Cooperator:} \qquad & C(x) := G(1, x-1) - L(1, x-1) \ , \\ \text{Deviator:} \qquad & D(x) := G(0, x) - L(0, x) \ . \end{aligned}$$

We impose the following two assumptions on the utility functions:

Assumption 4.1. *Non-cooperation dominates cooperation,*

$$\textit{(A1)} \qquad D(x-1) > C(x),\ \forall 1 \leq x \leq N\ .$$

Assumption 4.1 entails that the disclosure costs outweigh the gains from sharing for the firm, making $r_i = 0$ a dominant strategy. In other words, the marginal benefit from increased trust or approval due to disclosure is limited compared to the potential market and reputation loss due to disclosed security weaknesses. Therefore, the only Nash equilibrium of a one-shot information sharing game is for no firm to disclose her information. This observation is consistent with similar studies of one-shot information sharing games in [53, 73], 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.

Assumption 4.2. *Non-cooperation is inefficient,*

$$\textit{(A2)} \qquad C(N) > D(0) = 0\ .$$

Assumption 4.2 entails that the resulting non-disclosure equilibrium is suboptimal, particularly compared to the outcome in which all firms disclose. That is, full disclosure dominates the unique Nash equilibrium of the one-shot game. We may further be interested in imposing a more restrictive condition (although this is not necessary for our technical discussion).

$$\text{(A2')} \qquad xC(x) + (N-x)D(x) > (x-1)C(x-1) + (N-x+1)D(x-1), \quad \forall 1 \leq x \leq N\ .$$

Under (A2') (which indeed implies (A2)), non-disclosure by any firm decreases social welfare, making the full disclosure equilibrium $x = N$ the socially desired outcome.

Example 4.1. Consider the gain functions $G(1,z) = G(0,z) = zG$ and loss functions $L(0,z) = 0$ and $L(1,z) = L$. Here, each firm obtains a constant gain G from any other firm who is disclosing information, and incurs a constant loss L if she discloses

herself, both regardless of the number of other firms making a disclosure decision. It is easy to verify that these functions satisfy (A1). Furthermore, if $G > \frac{L}{N-1}$, conditions (A2) and (A2') hold as well. Note also that the 2-player prisoner's dilemma can be recovered as a special case when $N = 2$.

Example 4.2. Alternatively, consider the gain functions $G(1, z) = G(0, z) = f(z)G$, where $f(\cdot) : \{0, \ldots, N-1\} \to \mathbb{R}^+$ is an increasing and concave function, and loss functions $L(0, z) = 0$ and $L(1, z) = L$. The concavity of $f(\cdot)$ implies that as the number of cooperators increases, the marginal increase in information gain is decreasing due to potential overlap in the disclosed information. The utilities of cooperators and deviators will be given by:

$$C(x) = f(x-1)G - L, \text{ and, } D(x) = f(x)G \ .$$

Condition (A1) in Assumption 4.1 is satisfied. For Assumption 4.2, condition (A2) will hold if and only if $G > \frac{L}{f(N-1)-f(0)}$. However, unlike the previous example, for (A2') to hold we need additional restrictions beyond that required for (A2). Specifically, the full disclosure equilibrium will be the optimal solution only if the constants G and L are such that:

$$G[(N-x)\,(f(x) - f(x-1)) + (x-1)\,(f(x-1) - f(x-2))] > L, \forall x \ .$$

The described N-player game with Assumptions 4.1 and 4.2 is known as the N-person Prisoner's Dilemma (NPD) game; see e.g., [15, 48]. These games are used to model social situations in which there is a conflict between individual and social goals; examples include individual decisions whether to belong to unions, political parties, or lobbies, and problems of pollution or overpopulation [15]. The imposed assumptions then model the intuition that in such situations, any individual has a disincentive for cooperation, (A1), despite the fact that an outcome in which all cooperate would have been preferred by each participant, (A2).

4.2.2 Repeated interactions and the monitoring structure

Throughout the following sections, we are interested in the design of inter-temporal incentives that can incentivize firms to move away from the one-shot equilibrium of the information sharing game, and adopt full disclosure decisions when they interact repeatedly. Such inter-temporal incentives should be based on the history of firms' past interactions. We therefore formalize firms' monitoring capabilities, and the ensuing beliefs, of whether other firms are adhering to the information sharing agreement.

First, note that such monitoring is inevitably *imperfect*; after all, the goal of an information sharing agreement is to encourage firms to reveal their non-verifiable and private breach and security information. Furthermore, the monitoring can be either carried out independently by the firms, or be based on the reports of a central monitoring system. We consider both possibilities.

Imperfect private monitoring First, assume each firm conducts her own monitoring and forms a belief on other firms' disclosure decisions. Specifically, by monitoring firm j's externally observed security posture, firm i forms a *belief* b_{ij} about j's report. We let $b_{ij} = 1$ indicate a belief by firm i that firm j has been honest and is fully disclosing all information, and $b_{ij} = 0$ otherwise. In other words, $b_{ij} = 0$ indicates that firm i's monitoring provides her with evidence that firm j has experienced an undisclosed breach, has an unreported security flaw, or has fabricated an incident. Formally, we assume the following distribution on firm i's belief given firm j's report:

$$\pi(b_{ij}|r_j) = \begin{cases} \epsilon, & \text{for } b_{ij} = 0, r_j = 1 \\ 1-\epsilon, & \text{for } b_{ij} = 1, r_j = 1 \\ \alpha, & \text{for } b_{ij} = 0, r_j = 0 \\ 1-\alpha, & \text{for } b_{ij} = 1, r_j = 0 \end{cases} \tag{4.1}$$

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.

We assume the evidence available to firm i, and hence the resulting belief b_{ij}, is private to firm i, and independent of all other beliefs. Specifically, $b_{ij}, \forall i \neq j$ are i.i.d. samples of a Bernoulli random variable (with parameter α or ϵ, depending on r_j).

Imperfect 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 $\{b_{01}, \ldots, b_{0N}\}$ formed by the monitor is:

$$\hat{\pi}(\{b_{01}, \ldots, b_{0N}\}|\{r_1, \ldots, r_N\}) := \Pi_{j=1}^{N} \pi(b_{0j}|r_j) \ , \tag{4.2}$$

where the distributions $\pi(b_{0j}|r_j)$ follow (4.1), 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.[4]

4.3 Imperfect public monitoring: The role of centralized monitoring

The possibility of public monitoring (either perfect or imperfect) can enable the design of inter-temporal incentives for cooperation in repeated interactions. With perfect public monitoring, deviations from the intended equilibrium path are perfectly observable by all participants, and can be accordingly punished. As a result,

[4]It is worth mentioning that the binary beliefs are assumed for ease of exposition; the results of the subsequent sections continue to hold if the monitoring technology has finitely many outputs.

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. This has led to folk theorems under perfect monitoring; see e.g., [46]. With imperfect public monitoring on the other hand, deviations cannot 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 presenting some preliminaries on infinitely repeated games with imperfect public monitoring, and in particular, the folk theorem of [45] for these games. In Section 4.3.2, we show that this folk theorem applies to NPD information sharing games with monitoring given by (4.2).

4.3.1 The folk theorem with imperfect public monitoring

In this section, we present the folk theorem due to [45]. 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:

$$u_i(\mathbf{r}) = \sum_{b \in B} u_i^*(r_i, b)\pi(b|\mathbf{r}). \tag{4.3}$$

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 $\boldsymbol{\rho}^i$, with

$$\begin{aligned} \boldsymbol{\rho}^i_{-i} \in & \arg\min_{\boldsymbol{\rho}_{-i}} \left(\max_{r_i} \; u_i(r_i, \boldsymbol{\rho}_{-i}) \right) , \\ \rho^i_i \in & \arg\max_{r_i} \quad u_i(r_i, \boldsymbol{\rho}^i_{-i}) , \end{aligned}$$

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 containing 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 i 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 (see Section 4.4). 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 $\mathcal{F}^*$ 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. This means that the distribution over signals induced by some profile $\boldsymbol{\rho}$ are different from that induced by any $(\rho_i', \boldsymbol{\rho}_{-i})$ for $\rho_i' \neq \rho_i$. This condition is formally stated as follows.

Definition 4.1. *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 4.2. *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 folk theorem under imperfect public monitoring.

Theorem 4.1 (The imperfect public monitoring folk theorem, [45])**.** *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 $\mathbf{v} \in int\mathcal{F}^*$, there exists a discount factor $\underline{\delta} < 1$, such that for all $\delta \in (\underline{\delta}, 1)$, $\mathbf{v} \in \mathcal{E}(\delta)$.*

4.3.2 Cooperation in information sharing with public monitoring

We now verify that the above folk theorem applies to information sharing games with imperfect public monitoring structure given by (4.2). 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 informativeness of the public monitoring signal, hold for (4.2). First, note that the public signal b has 2^N possible outcomes; we view each signal as a binary string and assume the columns of the following matrices are ordered according to the decimal equivalent of these binary strings.

We first verify that the minmax profile of the repeated information sharing game has individual full rank for any firm. The minmax action profile for some firm i, $\hat{\mathbf{r}}^i$, is all firms concealing their information, i.e., $\hat{\mathbf{r}}^i = \mathbf{0}$. Consider deviations by firm 1 (by the symmetric nature of the game, the same argument holds for other firms). Then $A_1(\hat{\mathbf{r}}^i)$ is given by:

$$\begin{array}{c} \mathbf{b} = \\ r_1 = 0 \\ r_1 = 1 \end{array} \begin{array}{c} \begin{array}{cccccc} (0,0,\ldots,0) & (1,0,\ldots,0) & \ldots & (0,1,\ldots,1) & & (1,1,\ldots,1) \end{array} \\ \begin{pmatrix} \alpha^N & (1-\alpha)\alpha^{N-1} & \ldots & \alpha(1-\alpha)^{N-1} & & (1-\alpha)^N \\ \epsilon\alpha^{N-1} & (1-\epsilon)\alpha^{N-1} & \ldots & \epsilon(1-\alpha)^{N-1} & & (1-\epsilon)(1-\alpha)^{N-1} \end{pmatrix} \end{array}$$

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

We also need to verify that all pure strategy action profiles have pairwise full rank. We do so for $\mathbf{r}_k := (1, 1, \ldots, 1, 0, 0, \ldots, 0)$, where the first k firms disclose, and the remainder $N-k$ conceal; the result for other profiles can be shown similarly. For the profile $\mathbf{r}_k$, first consider the firms $i = 1$ and $j = N$. The matrix $A_{ij}(\mathbf{r}_k)$ is given by:

$$
\begin{array}{lcccc}
\mathbf{b} = & (0,0,\ldots,0) & (1,0,\ldots,0) & \ldots \\
r_1 = 0 & \alpha \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\alpha) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \ldots \\
r_1 = 1 & \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \ldots \\
r_N = 0 & \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \alpha & \ldots \\
r_N = 1 & \epsilon \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \epsilon & (1-\epsilon) \cdot \epsilon^{k-1} \cdot \alpha^{N-k-1} \cdot \epsilon & \ldots
\end{array}
$$

$$
\begin{array}{cccc}
\ldots & (0,1,\ldots,1) & (1,1,\ldots,1) & \\
\ldots & \alpha \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\alpha) \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & \\
\ldots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\epsilon)^{k} \cdot (1-\alpha)^{N-k} & \\
\ldots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k} & (1-\epsilon)^{k} \cdot (1-\alpha)^{N-k} & \\
\ldots & \epsilon \cdot (1-\epsilon)^{k-1} \cdot (1-\alpha)^{N-k-1} \cdot (1-\epsilon) & (1-\epsilon)^{k} \cdot (1-\alpha)^{N-k-1} \cdot (1-\epsilon) &
\end{array}
$$

Note that the rows corresponding to $r_1 = 1$ and $r_N = 0$ are the same: indeed when both firms follow the prescribed strategy, the distribution of the signals is consistent. It is straightforward to verify that the above has row rank 3; i.e., removing the common row, the three remaining rows are linearly independent. As a result, $\mathbf{r}_k$ has pairwise full rank for firms $i = 1$ and $j = N$. A similar procedure follows for other pairs of firms i, j, the remaining pure action profiles, verifying that all have pairwise full rank.

We have therefore established the following proposition:

Proposition 4.1. *The conditions of the folk theorem of Section 4.3.1 hold with*

	C	D
C	$G-L$, $G-L$	$-L$, G
D	G, $-L$	0, 0

Table 4.1: Firms' payoffs in a two-person prisoner's dilemma game

the public signals distributed according to (4.2). *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.*

4.3.3 Constructing public strategies: An example

In this section, we present the process through which equilibrium public strategies leading to a desired payoff profile are constructed. To simplify the illustration, we consider a two player prisoner's dilemma game with payoff matrix given by Table 4.1.

We first present an overview of the idea behind constructing the equilibrium strategies. The utility of firms at each step of the game can be decomposed into their current payoff, plus the continuation payoff; the latter is the expected payoff for the remainder of the game depending on the observed public monitoring output. Therefore, to achieve an average payoff profile $\mathbf{v}$ as equilibrium in the repeated game, the action profile and the continuation payoffs should be selected so as to maximize firms' expected payoff.

Formally, we say $\mathbf{v}$ is decomposed by $\mathbf{r}$ on a set W using a mapping $\gamma : B \to W$ if we have:

$$\begin{aligned} v_i = \;& (1-\delta)u_i(\mathbf{r}) + \delta E[\gamma_i(b)|\mathbf{r}] \\ \geq \;& (1-\delta)u_i(r_i', \mathbf{r}_{-i}) + \delta E[\gamma_i(b)|r_i', \mathbf{r}_{-i}],\ \forall r_i' \in R_i\ , \forall i. \end{aligned} \tag{4.4}$$

Here, the mapping γ determines firms' continuation payoffs (selected from a set W) following each signal $b \in B$. The goal is thus to set $W = \mathcal{E}(\delta)$ (the set of PPE payoffs), and find appropriate actions $\mathbf{r}$ and mappings γ decomposing (i.e., satisfying

(4.4) for) payoff profiles $\mathbf{v} \in \mathcal{E}(\delta)$. We can then conclude that any payoff profile $\mathbf{v}$ for which the above decomposition is possible, will be attainable as a PPE average payoff, as we can recursively decompose the selected continuation payoffs on $\mathcal{E}(\delta)$ as well. This procedure thus characterizes the set of payoffs that can be attained using public strategies.

However, the set of decomposable payoffs on arbitrary sets W is in general hard to characterize; let's instead consider the simpler decomposition on half-spaces $H(\lambda, \lambda\cdot\mathbf{v}) := \{\mathbf{v}' \in \mathbb{R}^N : \lambda\cdot\mathbf{v}' \leq \lambda\cdot\mathbf{v}\}$. With $W = H(\lambda, \lambda\cdot\mathbf{v})$, (4.4) can be re-written as:

$$\begin{aligned} v_i \;&=\; u_i(\mathbf{r}) + E[\bar{\gamma}_i(b)|\mathbf{r}] \;\;\geq\;\; u_i(r_i', \mathbf{r}_{-i}) + E[\bar{\gamma}_i(b)|r_i', \mathbf{r}_{-i}],\ \forall r_i' \in R_i\ , \forall i\ , \\ \text{and,}\quad & \lambda\cdot\bar{\gamma}(b) \leq 0,\ \forall b \in B\ , \end{aligned} \tag{4.5}$$

where $\bar{\gamma} : B \to \mathbb{R}^N$, and $\bar{\gamma}_i(b) = \frac{\delta}{1-\delta}(\gamma_i(b) - v_i)$. We refer to $\bar{\gamma}_i$ as the normalized continuation payoffs.

It can be shown (see [80]) that characterizing the set of attainable PPE payoffs $\mathcal{E}(\delta)$ is equivalent to finding the maximum average payoffs that can be decomposed on half-spaces using different actions $\mathbf{r}$ and in various directions λ. We therefore first find the maximum average payoffs $\mathbf{v}$ enforceable on half-spaces (i.e, satisfying (4.5), and with $\lambda\cdot\bar{\gamma}(b) = 0$ whenever possible), for each action profile $\mathbf{r}$ and direction λ. We will then select the best action $\mathbf{r}$ for each direction, and finally take the intersection over all possible directions λ to characterize $\mathcal{E}(\delta)$.[5]

We now find the average payoffs decomposable on half-spaces for the prisoner's dilemma game in Table 4.1. Let us first consider profile $\mathbf{r} = (1, 1)$,[6] and an arbitrary

[5]Define $k^*(\lambda; \mathbf{r}) := \lambda\cdot\bar{\mathbf{v}}$, where $\bar{\mathbf{v}}$ is the maximum payoff profile satisfying (4.5). It can be shown that $k^*(\lambda; \mathbf{r}) \leq \lambda\cdot u(\mathbf{r})$, and so the maximum is attained when $\mathbf{r}$ is *orthogonally enforced* (whenever possible); this is $\lambda\cdot\bar{\gamma}(b) = 0$ in (4.5). Let $k^*(\lambda) = \sup_{\mathbf{r}} k^*(\lambda; \mathbf{r})$. Intuitively, $k^*(\lambda)$ is a bound on the average payoff for firms for which the incentive constraints are satisfied. Let $H^*(\lambda) := H(\lambda, k^*(\lambda))$ be the corresponding *maximal half-space*. Then, that the set of PPE payoffs is contained in the intersection of these maximal half-spaces, i.e., $\mathcal{E}(\delta) \subseteq \cap_\lambda H^*(\lambda) := \mathcal{M}$, and that the reverse is also true for sufficiently large δ; i.e, $\lim_{\delta\to 1} \mathcal{E}(\delta) = \mathcal{M}$. We refer the interested reader to [80] for more details.

[6]Note that decomposing using $(0, 0)$ is not considered as it leads to the maximal half-space $\mathbb{R}^2$. It thus provides no information on the set of attainable payoffs as we already know that $\mathcal{E}(\delta) \subseteq \mathbb{R}^2$.

direction $\lambda = (\lambda_1, \lambda_2)$. Setting $\lambda \cdot \bar{\gamma}(b) = 0$, (4.5) reduces to:

$$\begin{aligned}
G - L &= G - L + \left(\epsilon^2\bar{\gamma}_1(0,0) + \epsilon(1-\epsilon)\bar{\gamma}_1(0,1) + (1-\epsilon)\epsilon\bar{\gamma}_1(1,0) + (1-\epsilon)^2\bar{\gamma}_1(1,1)\right) \\
&\geq \quad G \ + (\epsilon\alpha\bar{\gamma}_1(0,0) + \alpha(1-\epsilon)\bar{\gamma}_1(0,1) + (1-\alpha)\epsilon\bar{\gamma}_1(1,0) + (1-\epsilon)(1-\alpha)\bar{\gamma}_1(1,1)) \\
&\text{and ,} \\
G - L &= G - L + \left(\epsilon^2\bar{\gamma}_2(0,0) + \epsilon(1-\epsilon)\bar{\gamma}_2(0,1) + (1-\epsilon)\epsilon\bar{\gamma}_2(1,0) + (1-\epsilon)^2\bar{\gamma}_2(1,1)\right) \\
&\geq \quad G \ + (\epsilon\alpha\bar{\gamma}_2(0,0) + \epsilon(1-\alpha)\bar{\gamma}_2(0,1) + (1-\epsilon)\alpha\bar{\gamma}_2(1,0) + (1-\epsilon)(1-\alpha)\bar{\gamma}_2(1,1)) \\
&\text{and ,} \\
&\lambda_1\bar{\gamma}_1(b) + \lambda_2\bar{\gamma}_2(b) = 0, \quad \forall b \in B \ .
\end{aligned}$$

Substituting for $\bar{\gamma}_2(b)$ using the last equation, and writing the inequalities as equalities, finding the normalized continuation payoffs is equivalent to solving the following system of equations:

$$\begin{pmatrix} \epsilon^2 & \epsilon(1-\epsilon) & (1-\epsilon)\epsilon & (1-\epsilon)^2 \\ \alpha\epsilon & \alpha(1-\epsilon) & (1-\alpha)\epsilon & (1-\alpha)(1-\epsilon) \\ \epsilon^2 & \epsilon(1-\epsilon) & (1-\epsilon)\epsilon & (1-\epsilon)^2 \\ \alpha\epsilon & (1-\alpha)\epsilon & \alpha(1-\epsilon) & (1-\alpha)(1-\epsilon) \end{pmatrix} \begin{pmatrix} \bar{\gamma}_1(0,0) \\ \bar{\gamma}_1(0,1) \\ \bar{\gamma}_1(1,0) \\ \bar{\gamma}_1(1,1) \end{pmatrix} = \begin{pmatrix} 0 \\ -L \\ 0 \\ L\frac{\lambda_2}{\lambda_1} \end{pmatrix} .$$

The first and third rows represent the same equations (corresponding to the equilibrium outcome). Removing the third row and performing row-reduction on the remaining matrix, the continuation payoffs should satisfy the following set of equations:

$$\begin{aligned}
\epsilon\bar{\gamma}_1(0,0) + (1-\epsilon)\bar{\gamma}_1(0,1) &= \frac{-L}{\alpha\kappa}\frac{1-\epsilon}{\epsilon} \\
\epsilon\bar{\gamma}_1(1,0) + (1-\epsilon)\bar{\gamma}_1(1,1) &= \frac{L}{\alpha\kappa} \\
-\bar{\gamma}_1(0,1) + \bar{\gamma}_1(1,0) &= \frac{L}{\epsilon\alpha\kappa}(\frac{\lambda_2}{\lambda_1} + 1) \ ,
\end{aligned}$$

where $\kappa := \frac{1-\epsilon}{\epsilon} - \frac{1-\alpha}{\alpha} > 0$. The above is an underdetermined system, and thus has

	$\bar{\gamma}_1(b)$	$\bar{\gamma}_2(b)$
b=(0,0)	$\frac{L}{\epsilon\alpha\kappa}\frac{1-\epsilon}{\epsilon}(\frac{\lambda_2}{\lambda_1}-1)$	$\frac{L}{\epsilon\alpha\kappa}\frac{1-\epsilon}{\epsilon}(\frac{\lambda_1}{\lambda_2}-1)$
b=(0,1)	$-\frac{\lambda_2}{\lambda_1}\frac{L}{\epsilon\alpha\kappa}$	$\frac{L}{\epsilon\alpha\kappa}$
b=(1,0)	$\frac{L}{\epsilon\alpha\kappa}$	$-\frac{\lambda_1}{\lambda_2}\frac{L}{\epsilon\alpha\kappa}$
b=(1,1)	0	0

Table 4.2: An example of normalized continuation payoff choices in repeated information sharing games

infinitely many solutions depending on the designer's choice of continuation payoffs. We construct and interpret one such possibility.

Set $\bar{\gamma}_1(1,1) = 0$, implying $\bar{\gamma}_2(1,1) = 0$ as well. This means if the signal indicates that both firms are cooperating with high probability, there is no need for punishments, so that both firms expect their continuation payoff to remain unchanged (i.e., equal to their current payoff). Given this choice, we can solve for the remaining normalized continuation payoffs, illustrated in Table 4.2.

Intuition. These normalized continuation payoffs can be intuitively interpreted as follows. Fix a direction with $\lambda_1, \lambda_2 > 0$ (similar interpretations follow for other directions). Then, given a signal $b = (1,0)$, which is more likely under a deviation by firm 2, firm 1 expects a higher continuation payoff ($\bar{\gamma}_1(1,0) > 0$), while the suspect deviator expects a lower one ($\bar{\gamma}_2(1,0) < 0$).[7] A similar intuition applies to the continuations under the signal $(0,1)$. On the other hand, with $b = (0,0)$, either firm 1 or 2 will be punished, depending on the direction λ. Specifically, for a direction $\lambda_1 = \lambda_2$, neither firm expects a change in her continuation payoff. Note that with $\lambda_1 = \lambda_2$, the change in continuation payoffs between the outcomes $(0,1)$ and $(1,0)$, as well as among firms in either outcome, are also of equal size. Note also that both firms are never punished simultaneously under any outcome, so as to maintain a high average payoff.

Finally, it is worth noting the effect of the monitoring accuracy, α and ϵ, on the normalized continuation payoffs. Consider direction $\lambda_1 = \lambda_2 = 1$, and fix $L = 1$. First, note that $\epsilon\alpha\kappa = \alpha(1-\epsilon) - \epsilon(1-\alpha)$ is increasing in α and decreasing in

[7] It is worth emphasizing that due to the equilibrium construction, firms are both playing $r_i = 1$; nevertheless, punishments on the equilibrium path happen due to imperfect monitoring.

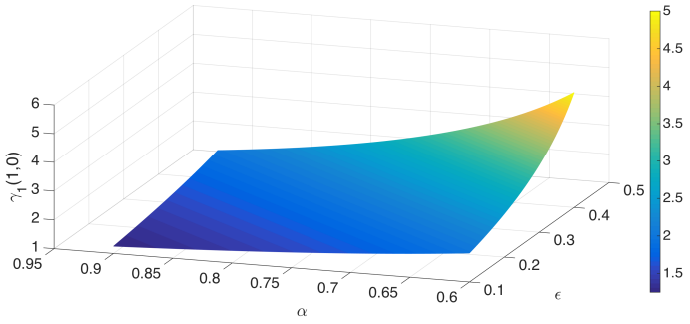

Figure 4.1: Effect of monitoring accuracy on normalized continuation payoffs, $\bar{\gamma}_1(1,0)$

ϵ. This is illustrated in Figure 4.1, which shows the dependence of $\bar{\gamma}_1(1,0)$ on the monitoring parameters. As a result, as the monitoring technology becomes more accurate, i.e., α increases and/or ϵ decreases, the size of the normalized continuation payoffs for firms, when $(1,0)$ or $(0,1)$ is observed, becomes smaller. This is because, as monitoring becomes accurate, signals indicating deviations (despite equilibrium being played) are more likely to be due to monitoring errors rather than actual deviations, and therefore the required continuation punishments/rewards become less severe to maintain high average payoffs for firms.

We conclude that in general, using the described procedure, we can decompose payoff profiles in the half-spaces $H(\lambda, k^*(\lambda,(1,1)))$, where $k^*(\lambda,(1,1)) = \lambda \cdot u(1,1) = (G-L)(\lambda_2+\lambda_1)$, using the action profile $\mathbf{r} = (1,1)$ and continuation payoffs determined as above. Using a similar procedure, the corresponding half-spaces for the remaining action profiles will have $k^*(\lambda,(0,1)) = G\lambda_1 - L\lambda_2$ and $k^*(\lambda,(1,0)) = G\lambda_2 - L\lambda_1$.

We next choose, for a given direction λ, the action for which the corresponding half-spaces cover a larger set of average payoffs; these are $k^*(\lambda) = \max_{\mathbf{r}}\{G\lambda_2 -$

$L\lambda_1, G\lambda_1 - L\lambda_2, (G-L)(\lambda_1+\lambda_2)\}$. We therefore have:

$$k^*(\lambda) = \begin{cases} G\lambda_2 - L\lambda_1 & \lambda_2 \geq \frac{G}{L}\lambda_1 \\ (G-L)(\lambda_1+\lambda_2) & \frac{L}{G}\lambda_1 \leq \lambda_2 \leq \frac{G}{L}\lambda_1 \\ G\lambda_1 - L\lambda_2 & \lambda_1 \geq \frac{G}{L}\lambda_2 \end{cases}$$

Finally, it is straightforward to show that the intersection of half-spaces $H(\lambda, k^*(\lambda))$, as λ ranges over $\mathbb{R}^2$, is equivalent to the set of feasible and strictly individually rational payoffs of the two-person prisoner's dilemma game of Table 4.1. That is, it is possible to find an action profile $\mathbf{r}$ and the corresponding continuation payoff mapping γ (constructed as described above), so as to incentivize any feasible strictly individually rational payoff profile.

4.4 Imperfect private monitoring: The role of communication

In this section, we consider the use of private monitoring in providing intertemporal incentives for information sharing. Unlike repeated games with imperfect public monitoring, relatively less is known about games with private monitoring [66].

In particular, we are interested in a folk theorem for this repeated game; a folk theorem in this scenario is a full characterization of payoffs that can be achieved when firms only have private observations, if firms are sufficiently patient. As discussed in Section 4.3, with imperfect public monitoring, [45] presents a folk theorem under relatively general conditions. The possibility of this result hinges heavily on that firms 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. However, a similar folk theorem with private monitoring remained an open problem until recently,[8] mainly due to the

[8] A recent advance in the field is by Sugaya [117, 118], who presents a folk theorem for repeated games with imperfect private monitoring, without requiring cheap talk communication or public randomization. The conditions on the private monitoring structure required by Sugaya's folk

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 these games. Examples include:

- Games in which firms are allowed to communicate (cheap talk) after each period. This approach has been proposed in [27, 67], and in essence, uses the signals collected through communication as a public signal, allowing participants to coordinate on cooperation.

- Games in which firms have public actions (e.g., announcement of sanctions) in addition to private decisions (here, disclosure decisions), as proposed by [102] for the study of international trade agreements. Intuitively, public actions serve a similar purpose as communication, allowing participants 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, [79] proves a folk theorem for almost-perfect and almost-public monitoring.

In this section, we present the folk theorem with private monitoring and communication due to [67] in Section 4.4.1, and in Section 4.4.2, verify that it applies to NPD information sharing games with monitoring given by (4.1).

4.4.1 The folk theorem with imperfect private monitoring and communication

In this section, we state the folk theorem with imperfect private monitoring, allowing for communication between players, due to [67].

The stage game is similar to the setup of [45] in Section 4.3.1. Consider N rational players. At the stage game, each player i chooses an action $r_i \in R_i$. Let

theorem are however more restrictive than those of Kandori and Matsushima's folk theorem with (cheap talk) communication [67]. Therefore, we analyze the application of the folk theorem with communication of [67]; this will further allow us to draw a closer parallel with the public monitoring structure of the previous section.

$\mathbf{r} \in R := \prod_{i=1}^{N} R_i$ denote a profile of actions. At the end of an stage, each player privately observes an outcome $b_i \in B_i$, where B_i is a finite set of possible signals. The probability of observing the profile of private signals $\mathbf{b} \in B := \prod_{i=1}^{N} B_i$ following $\mathbf{r}$ is given by the joint distribution $\pi(\mathbf{b}|\mathbf{r})$. Assume π has full support, i.e., $\pi(\mathbf{b}|\mathbf{r}) > 0,\ \forall \mathbf{b}, \forall \mathbf{r}$. Let $u_i^*(r_i, b_i)$ be the utility of player i when she plays r_i and observes the signal b_i. Note that i's utility depends on others' actions only through b_i, and thus the stage payoffs are not informative about others' actions. The expected stage game payoff for user i when $\mathbf{r}$ is played is therefore given by

$$u_i(\mathbf{r}) = \sum_{\mathbf{b} \in B} u_i^*(r_i, b_i)\pi(\mathbf{b}|\mathbf{r}). \tag{4.6}$$

The definition of the minmax action profiles $\boldsymbol{\rho}^i$ and the set of feasible and strictly individually rational payoffs $\mathcal{F}^* \subset \mathbb{R}^N$ are the same as those in Section 4.3.1. However, in addition to the private nature of signals b_i, the current model differs from the setup in Section 4.3.1 in that we allow the players to communicate in this game. Formally, after choosing the action r_i and observing the signal b_i, each player i will publicly announce a message $m_i \in M_i$, selected from the finite set of possible messages M_i. Let $M = \Pi_{i=1}^{N} M_i$.

Consequently, the strategy $s_i = (r_i, m_i)$ of each player consists of both an action r_i and a message m_i. In particular, when the game is played repeatedly often, the strategy specifies a choice for each time step t, i.e, $r_i = (r_i(t))_{t=0}^{\infty}$ and $m_i = (m_i(t))_{t=0}^{\infty}$, where,

$$\begin{aligned} r_i(t) &: R_i^{t-1} \times B_i^{t-1} \times M^{t-1} \to \Delta(R_i)\ , \\ m_i(t) &: R_i^{t} \times B_i^{t} \times M^{t-1} \to \Delta(M_i)\ . \end{aligned}$$

Let $r_i^t = (r_i(0), \ldots, r_i(t))$. Define b_i^t and m^t similarly. Then, the private history of player i at the end of time t is given by $h_i^t := (r_i^t, b_i^t)$, and the public history is $h^t := m^t$. Therefore, players' strategies depend on both private and public histories. Given the strategy profiles $\mathbf{s} = (s_1, \ldots, s_N)$, and assuming that players discount future payoffs by a discount factor δ, a player's average payoff throughout the repeated game is given

by $(1-\delta)\sum_{t=0}^{\infty}\delta^t u_i(\mathbf{s}(t))$. Each player i is choosing her strategy s_i to maximize the expected value of this expression.

We are interested in characterizing the payoffs attainable by the strategy profiles $\mathbf{s}$ that are a *sequential equilibrium* of the game. Formally, $\mathbf{s}$ is a sequential equilibrium of the game if for every player and her history (h_i^t, h^t), $s_i|_{(h_i^t,h^t)}$ is a best reply to $E[s_{-i}|_{h_{-i}^t,h^t}|h_i^t]$. That is, a player is best-responding according to her belief over private histories of other players, in particular those which are consistent with her own private history. Let $V(\delta)$ denote the set of sequential equilibrium average payoffs when the discount factor is δ. We are interested in identifying conditions under which $V(\delta)\subseteq\mathcal{F}^*$.

Recall that in the game of imperfect public monitoring, conditioning of strategies on the publicly observed signal allowed players to coordinate, and to recover a recursive structure in the game. The possibility of communication allows for recovering a similar recursive structure in games with private monitoring. The equilibrium strategies leading to nearly efficient payoffs will be constructed as follows. At the end of each period t, each player i is asked to report her privately observed signal as her message, i.e, $m_i(t)=b_i(t)$. To make sure that players truthfully report their signals, the equilibrium strategies use this private information to determine *other* players' deviations and future payoffs, and maintain i's payoff independent of her report. As a result, truthful reporting of privately observed signals will be a (weak) best-response.[9,10] It remains to ensure, following a rationale similar to the folk theorem of Section 4.3.1, that the available signals are informative enough, in the sense that they allow players to distinguish between different deviations of individuals, and to differentiate among deviations by different players.

The required conditions on the informativeness of the players' signals are as

[9]It is also possible to make truth reporting a *strict* best-response if players' privately observed signals are mutually correlated; see [67, Section 4.2].

[10]Note that unlike Section 4.3, each player will be playing a private strategy at equilibrium, as she is using her private information as her message m_i. However, the choice of action r_i will be based only on the public information (i.e., the disclosed messages available to all players).

follows. First, define the following vectors:

$$\begin{aligned} p_{-i}(\mathbf{r}) &:= (\pi_{-i}(\mathbf{b}_{-i}|\mathbf{r}))_{\mathbf{b}_{-i}\in B_{-i}} \ , \\ p_{-ij}(\mathbf{r}) &:= (\pi_{-ij}(\mathbf{b}_{-ij}|\mathbf{r}))_{\mathbf{b}_{-ij}\in B_{-ij}} \\ Q_{ij}(\mathbf{r}) &:= \{p_{-ij}(\mathbf{r}_{-i}, r_i')|r_i' \in R_i \backslash \{r_i\}\} \ , \end{aligned}$$

where $B_{-i} := \Pi_{k\neq i} B_k$, $B_{-ij} := \Pi_{k\neq i,j} B_k$, and π_{-i} and π_{-ij} are marginal distributions of the joint distribution $\pi(\mathbf{b}|\mathbf{r})$ of privately observed signals. The three sufficient conditions on signals can be expressed accordingly.

Condition 4.1. *At the minmax strategy profile of a player i, $\hat{\boldsymbol{\rho}}^i$, for any player $j \neq i$ and any mixed strategy $\rho_j' \in \Delta(R_j)$, either*

$$\begin{aligned} &(i) \quad p_{-j}(\hat{\boldsymbol{\rho}}^i) \neq p_{-j}(\hat{\boldsymbol{\rho}}^i_{-j}, \rho_j') \quad \textit{or,} \\ &(ii) \quad p_{-j}(\hat{\boldsymbol{\rho}}^i) = p_{-j}(\hat{\boldsymbol{\rho}}^i_{-j}, \rho_j') \ \textit{and} \ u_j(\hat{\boldsymbol{\rho}}^i) \geq u_j(\hat{\boldsymbol{\rho}}^i_{-j}, \rho_j'). \end{aligned}$$

Condition 4.1 states that at the minmax profile of any player, a deviation by another player is either statistically distinguishable, and if not, it reduces the payoff of the deviator, and is hence not profitable. This condition ensures that we can provide incentives to players to punish (minmax) one another. Note that this requirement is similar to (but weaker than) the individual full rank condition in the folk theorem of Section 4.3.1.

Condition 4.2. *For each pair of players $i \neq j$, and each pure action equilibrium $\mathbf{r}$ leading to an extreme point of the payoff set $\mathcal{F}^\dagger$, we have:*

$$p_{-ij}(\mathbf{r}) \notin co(Q_{ij}(\mathbf{r}) \cap Q_{ji}(\mathbf{r})) \ ,$$

where $co(X)$ denotes the convex hull of the set X.

Recall that $Q_{ij}(\mathbf{r})$ denotes the vector of distribution of beliefs of players other than i and j, when player i is deviating. Condition 4.2 therefore requires that a deviation by either i or j (but not both) is statistically detected by the remaining players.

Condition 4.3. *For each pair of players $i \neq j$, and each pure action equilibrium* $\mathbf{r}$ *leading to an extreme point of the payoff set* $\mathcal{F}^{\dagger}$*, we have:*

$$co(Q_{ij}(\mathbf{r}) \cup p_{-ij}(\mathbf{r})) \cap co(Q_{ji}(\mathbf{r}) \cup p_{-ij}(\mathbf{r})) = \{p_{-ij}(\mathbf{r})\} \ .$$

Finally, Condition 4.3 requires that players other than i, j can statistically distinguish deviations by i from deviations by j, as the resulting distribution on $\mathbf{b}_{-ij}$ will be different under either player's deviation. In other words, the only consistent distribution arises when neither player is deviating. It is worth mentioning that Conditions 4.2 and 4.3 hold when the pairwise full rank condition of the folk theorem of Section 4.3.1 holds.

Therefore, given adequate private monitoring signals and communication, we have the following folk theorem under imperfect private monitoring.

Theorem 4.2. (The Imperfect private monitoring with communication folk theorem, [67]) *Assume that there are more than two players ($N > 2$), and the set of feasible and strictly individual rational payoffs $\mathcal{F}^* \subset \mathbb{R}^N$ has non-empty interior (and therefore dimension N). Then, if the monitoring of players satisfy Conditions 4.1, 4.2, and 4.3, any interior payoff profile $\mathbf{v} \in int\mathcal{F}^*$ can be achieved as a sequential equilibrium average payoff profile of the repeated game with communication, when δ is close enough to 1.*

4.4.2 Cooperation in information sharing with private monitoring and communication

We now verify that the above folk theorem applies to information sharing games with imperfect private monitoring structure given by (4.1). That is, when the firms are sufficiently patient, they can sustain cooperation on full security information sharing in a repeated setting, by truthfully revealing their private signals, and making their disclosure decisions based only on the imperfect, publicly announced collective observation about their past actions. To this end, we need to verify that the three conditions of the folk theorem on the informativeness of the monitoring signal hold for the joint distribution of the private signals in (4.1). However, note that once

these signals are truthfully reported, it is as if we have access to $N-1$ independent realizations of the public monitoring distribution in (4.2). Assume that to test the conditions of the folk theorem, we randomly choose one of the available cross-observations about a possible deviator (from all players other than the suspect for Condition 4.1, or other than the two suspects for Conditions 4.2 and 4.3) and test the statistical distinguishability of the signal. With this method, the joint distribution of the private signal that is being tested will be equivalent to the public monitoring distribution of (4.2).

On the other hand, as mentioned in Section 4.4.1, the conditions of the folk theorem for imperfect private monitoring are weaker than the full rank conditions required by the folk theorem with imperfect public monitoring: Condition 4.1 is a weaker condition than individual full rank, and Conditions 4.2 and 4.3 are weaker versions of pairwise full rank. This is because the individual and pairwise full rank conditions require linear independence of their corresponding signal distributions, while Conditions 4.1-4.3 are stated in terms of convex combinations. As shown in Section 4.3.2, the distribution in (4.2) satisfies the full rank conditions, and consequently, the joint private distribution resulting from (4.1) satisfies Conditions 4.1-4.3. We have therefore established the following proposition.

Proposition 4.2. *The conditions of the folk theorem of Section 4.4.1 hold with private monitoring* (4.1) *and communication. As a result, when the firms are sufficiently patient, i.e., they value the future outcomes of their information sharing agreement, and are allowed to communicate their private signals, it is possible for them to nearly efficiently cooperate on full information disclosure through repeated interactions.*

4.5 Conclusion

We modeled information sharing agreements among firms as an N-person prisoner's dilemma game, and equipped it with a simple binary monitoring structure. 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 showed

that a rating/monitoring system can play a crucial role in providing a common public signal which, despite being imperfect, can be used to design inter-temporal incentives that lead firms to cooperate on information sharing. We also showed that in the absence of a monitor, if the firms are provided with a platform to communicate their privately observed beliefs on each others' adherence to the agreement, it is again possible to design similar inter-temporal incentives.

An important requirement for the folk theorem, and consequently the design of inter-temporal incentives, is to ensure that firms are sufficiently patient (i.e., they place significant value on their future interactions), as characterized by having discount factors higher than $\underline{\delta}$. Despite the fact that the proposed binary monitoring structures in (4.1) and (4.2) are informative enough for the folk theorem to hold, their accuracy, (α, ϵ), will impact the requirement on firms' patience, $\underline{\delta}$. Characterizing the dependence of $\underline{\delta}$ on (α, ϵ) is a direction of future work. Particularly, we have only considered one method for using firms' communication of their privately observed signals to establish the folk theorem of Section 4.4.1. Determining the optimal method for combining firms' inputs to ensure the lowest $\underline{\delta}$ remains an interesting question.

Another possible extension is to consider the design of inter-temporal incentives when both types of public and private monitoring are available. It is indeed still possible to have firms coordinate based on the public monitoring system's report alone (i.e., use public strategies); nevertheless, it may also be possible to employ *private strategies*, in which firms use both their own observations, as well as the public signal. Private strategies may lead to higher payoffs than those attainable through public strategies alone [80, Chapter 10], thus making their study of interest to either lower the required discount factor, or to implement such strategies when the monitoring signals are not informative enough for a public monitoring folk theorem to hold.

Finally, studying the possibility of sustaining inter-temporal incentives under generalizations of the current model are of interest. While we have assumed that all firms benefit equally from others' disclosed information, a generalized model can introduce a network structure among firms to capture the importance of firms' information to

one another. We have also assumed that the monitoring, as well as its accuracy, are fixed and available to firms at no additional cost. Analyzing the effects of costly monitoring, as well as the dependence of the accuracy on the underlying security incident, are other interesting questions for future work.

Chapter 5

Crowdsourcing Reputation in Security Games

5.1 Introduction

A common assumption required in some classes of incentive mechanisms, including the Externality mechanism of Section 2.4.1, is that users' levels of effort are observable by the regulator, and can thus be used to determine appropriate taxes/rewards. Similarly, to mitigate the problem of moral hazard when designing cyber insurance contracts, or in order to devise insurance contracts with premium discrimination, an insurer needs to observe the investments of the insured, or at least have access to an accurate estimate of their effort levels. Finally, in security information sharing agreements, there is a need for forming assessments of participants' adherence to the terms of the agreement.

In general, the exact amount of security expenditure or security decisions by an autonomous entity remains her private information. Nevertheless, we can assume that a noisy observation of each entity's level of effort is available to the regulator, as well as to other entities. Such observation can be based on analyzing the exchanged traffic, past data on security incidents [32, 124, 104], externally observable mismanagement symptoms [127], or reputation blacklists [22, 60, 105, 37]. These observations can be mapped to appropriate security scores [77], which can then be aggregated to form an opinion about the security status, or *reputation*, of each entity.

While the (noisy) reputation assessment gleaned from an external vantage point

can be used for offering insurance contracts, determining taxes, or assessing adherence to the sharing agreement, we ask whether it is possible to design a reputation mechanism that can improve upon this assessment, by incorporating input from all entities in the environment. Ultimately, the designer aims to find the best estimate of each entity's security status. The entities themselves are on one hand attempting to have the best estimate of others' efforts to choose an action, such as the optimal level of self-protection, in response. On the other hand, each entity attempts to inflate the perception of other entities, as well as the regulator, about her own effort, so as to spend less resources while appearing to have a better security posture, consequently enjoying more favorable contracts, higher rewards, lower taxes, or continued benefits from the information sharing agreement.

In the remainder of this chapter, we focus on the general problem of reputation mechanism design, and when appropriate, interpret the model and results in the context of network (security) reputation.

5.1.1 Related work

The theory of mechanism design has been increasingly used to address problems of resource allocation in informationally decentralized systems with strategic users. Pricing schemes, e.g. [78], and auctions, e.g. [58], are two popular approaches in the design of allocation schemes in communication systems. The use of pricing allows the system to align individual users' objectives with global performance goals to implement optimal outcomes. Despite the possibility of using monetary taxation in our setting, alternative forms of leverage may be preferred in incentivizing user cooperation, though they are relatively hard to identify; two notable exceptions are [107, 126]. In [107], the authors study the problem of using the downlink rate allocated to a user as an alternative commodity to induce socially optimal uplink rate allocation in a multi-access broadcast channel with selfish users. The work of [126] proposes an *intervention* mechanism that uses the *commodity of interest* as the means for preventing users' deviation from their designated strategies. Specifically, a monitoring device is used to estimate the transmit power profile of selfish

users in a wireless network; it then chooses to transmit at a positive power level if users deviate to higher transmission powers, thus negatively affecting users' utilities. The punish-reward mechanism proposed in Section 5.3.1 of this chapter also relies on a credible threat of punishment to deter non-cooperative users from deviation. However, the above intervention mechanism only exercises punishment, while our punish-reward mechanism can also reward users' cooperative actions using the commodity of interest.

The work presented in this chapter is also closely related to elicitation and prediction mechanisms used for aggregating the predictions of agents about an event, see e.g. [125, 106, 49]. Scoring rules [125] incentivize an agent to truthfully reveal her prediction by offering rewards based on the accuracy of the agent's estimation as compared to the actual realization of the event. Although these rules can be used to quantify the performance of forecasters, they rely on the observation of an objective ground truth. A class of *peer prediction* methods can be used to eliminate the need for such verification by requesting an agent's own assessment, as well as her prediction of other agent's assessment. For example, the elicitation methods in [106] and [49] result in truthful revelation even for subjective assessments. However, in all aforementioned work, the users are essentially rewarded in accordance with their participation, but do not attach any value to the realization of the event, or the outcome that the elicitor may be building using the aggregated data. In this chapter, users attach value to the outcome built by the estimator. This allows our proposed mechanism to use non-monetary rewards (a vector of accurate reputation indices) to incentivize participation, whereas in elicitation methods monetary rewards are used to incentivize cooperation. Furthermore, although we are studying a problem of elicitation about an objective ground truth, this event is not observable by the elicitor.

5.1.2 Chapter contributions

The contribution of this chapter is to propose a simple punish-reward reputation mechanism that, by soliciting both self-assessments and cross-assessments from a

set of strategic users, is able to improve its own prior assessment of these users' reputation (security posture or related decisions), without the need for monetary incentives.

5.1.3 Chapter organization

Section 5.2 presents the model and preliminaries. The design and analysis of the punish-reward mechanism is presented in Section 5.3. Section 5.4 concludes the chapter.

5.2 The reputation system model

Consider a collection of $K \geq 2$ entities[1], denoted by $N_1, N_2, \ldots, N_K$. In the context of network reputation, a user N_i may refer to a network in a system of interconnected networks. Each user N_i's overall quality is described by a quantity r_{ii}, which we refer to as the *real* or *true* quality of N_i, or simply the *truth*. We assume without loss of generality that $r_{ii} \in [0, 1], \forall i$. With a slight abuse of notation we will use K to denote both the number of users and the set of user indices $\{1, 2, \cdots, K\}$ whenever there is no ambiguity. We assume that each user N_i is aware of her own condition, and therefore knows r_{ii} precisely, but this is her private information. We do note however that while it is technically feasible for any entity to obtain r_{ii} by monitoring her own actions/interactions (e.g. a network is aware of its expenditures in security measures), this is by no means always the case due to reasons such as resource constraints.

There is a central *reputation system* that is responsible for soliciting input from participants, and coming up with the system estimates. For instance, this could be a regulator or insurer in the network reputation example. Specifically, the system proposes a mechanism, according to which it collects input from participants and uses it to build a global quality assessment, in the form of a *reputation index*, for each of the K users in the system. Its goal is to have the reputation index reflect

[1]We will use the terms *users*, *entities* and *participants* interchangeably.

the true quality r_{ii} as accurately as possible.

In general, each user N_j independently monitors her interactions with another user N_i to form an estimate R_{ji} based on her observations. For example, a network N_j can monitor the inbound traffic from network N_i to form an opinion. However, N_j's observation is in general an *incomplete* view of N_i, and may contain errors depending on the monitoring and estimation technique used. We will thus assume that R_{ji} is described by a normal distribution $\mathcal{N}(\mu_{ji}, \sigma_{ji}^2)$, which itself may be unbiased ($\mu_{ji} = r_{ii}$) or biased ($\mu_{ji} \neq r_{ii}$) [2]. We will further assume that this distribution (but not its realization) is common knowledge.

The reputation system itself may also be able to monitor the actions of each user N_i so as to form its own estimate of N_i's condition. This will be denoted by R_{0i}, again a random variable for the same reason given above. In the network reputation example, the system's observations can be gathered by monitoring the outgoing traffic of a network, and by considering externally observable security indicators, including mismanagement symptoms and malicious activity data. As before we will assume that R_{0i} is normally distributed with $\mathcal{N}(\mu_{0i}, \sigma_{0i}^2)$, and that this distribution is known to user N_i.

Reputation mechanism's objective The reputation system operates as follows. It can collect a vector $(x_{ij})_{j \in K}$ of *reports* from each user N_i. It consists of *cross-reports* x_{ij}, $i, j = 1, \cdots, K$, $j \neq i$, which represent N_i's assessment of N_j's quality, and *self-reports* $x_{ii}, i = 1, 2, \cdots, K$, which are the users' *self-advertised* quality measure. The mechanism may be such that only a subset of these reports are collected. Furthermore, there is no a priori guarantee that the participants will report any of these quantities truthfully.

The reputation system's goal is to derive the reputation index for each user N_i so as to accurately reflect the true quality r_{ii}. This objective is different from what is commonly studied, e.g., revenue maximization. Formally, the reputation mechanism

[2] The assumption of a normal distribution is made for simplicity and concreteness. Our proposed framework is also applicable given other distributions.

is designed to solve the following problem:

$$\min \sum_i |\hat{r}_i - r_{ii}| \ . \tag{5.1}$$

Here, we have used the absolute error as a performance measure; other error functions may be adopted as well.

A *reputation mechanism* specifies a method used by the reputation system to compute the reputation indices, i.e., what input to solicit and how the inputs are used to generate output estimates. As users are entities acting in self-interest and the truth is their own private information, the key to a successful mechanism (one that attains the solution to (5.1)) is to induce the users to provide useful, or ideally truthful, input. Such a reputation mechanism will also be referred to as a *collective revelation* mechanism, a term borrowed from [49]. It is assumed that the mechanism is common knowledge among all K participating users.

Individual users' objectives In modeling users' objectives, we identify two elements of a user's payoff.

- *Truth*: Each user N_i may wish to obtain accurate estimates on a *set of users of interest* $\mathcal{I}_i \subset K$. Formally, this part of the objective function is given by

  $$I_i = -\sum_{j \in \mathcal{I}_i} f_i(|\hat{r}_j - r_{jj}|) \ .$$

  Here, $f_i(\cdot) \geq 0$ are increasing and convex functions. This element captures a user's interest in having accurate assessment of other users' quality so that she can properly regulate her actions. For instance, each network in an interdependent system requires an accurate assessment of others' security effort in order to decide on her own optimal level of investment.

- *Image*: Each user N_i may further wish to obtain as high as possible an estimate

on *herself*. Formally,

$$II_i = g_i(\hat{r}_i) \ .$$

Here, $g_i(\cdot) \geq 0$ are concave and increasing. This element reflects a user's interest in having a high reputation herself as it translates into other benefits as mentioned earlier. For instance, a network can enjoy higher rewards or lower taxes as a result of appearing more secure.

A general preference model of a legitimate, non-malicious user may consist of both elements, possibly weighted; that is, user N_i may be captured by

$$u_i = -\lambda \sum_{j \in \mathcal{I}_i} f_i(|\hat{r}_j - r_{jj}|) + (1-\lambda) g_i(\hat{r}_i) \ , \tag{5.2}$$

for some constant $0 \leq \lambda \leq 1$. Throughout this chapter, for simplicity and tractability, we assume the two elements are weighed equally by all users, and that $f_i(z) = g_i(z) = z, \forall i$; extension to other choices of $\lambda, g_i(\cdot)$, and $f_i(\cdot)$, is straightforward. Our work in [90] contains the study of reputation mechanisms for several other environments, including environments with only truth type (utilities with only a truth element), only image type (utilities with only an image element), or a mix of truth and image type users.

By defining these two utility elements, we assume a user's preference is in general increasing in the accuracy of others' quality estimate, and increasing in her own quality estimate. We assume these two characteristics to be common knowledge.

It should be noted that $\hat{r}_j$ is a function of the proposed game form $(\mathcal{M}, h)$, where $\mathcal{M}$ denotes users' message space, and $h(\cdot)$ denotes the function mapping users' messages to an outcome. Here, $\hat{r}_j = h(m_1, m_2, \ldots, m_K)$, with m_j denoting N_j's message. Since the proposed model is one of incomplete information, from N_i's viewpoint, the message profile $m \in \mathcal{M}$, and consequently u_i, is in general a random variable. Therefore, it is understood that N_i is an expected-utility maximizer.

5.3 Design of a reputation mechanism

5.3.1 The punish-reward (PR) mechanism

We next set out to construct a reputation mechanism which invokes the use of both self-reports and cross-reports, and will forgo the use of taxation. As we shall see later, even though the system's own observation R_{0i} is sufficient for implementing the proposed reputation mechanism, more cross-reports can improve the performance of our proposed mechanism when used properly. In particular, we allow the system to collect cross-observations from users N_j on users N_i who are in her set of interest, i.e., $i \in \mathcal{I}_j$. We will assume that a user N_j has observations of a user N_i if and only if this user is within her set of interest.

We first introduce a simple, benchmark mechanism, referred to as the *simple averaging mechanism*, where the reputation agent solicits cross-reports x_{ji} for $j, i \in \mathcal{I}_j$, and computes the estimate $\hat{r}_i$ as the average of these x_{ji} and its own observation R_{0i}. This is the basic mechanism used in many existing online systems, e.g., Amazon and Epinions [65]. The following proposition shows that for this mechanism, N_j will choose to participate, and truthfully disclose her observation R_{ji}. The proof is given in Appendix E.

Proposition 5.1. *Under the simple averaging mechanism, truthful revelation of the observation R_{ji}, by N_j, $j : i \in \mathcal{I}_j$ is a Bayesian Nash equilibrium. In addition, this mechanism is individually rational.*

In fact, if the estimates R_{ji}, for $j,\ i \in \mathcal{I}_j$, are unbiased, then $\hat{r}_i$ can be made arbitrarily close to r_{ii} as the number of participants increases. It is not hard to see that under this mechanism, if asked, N_i will always report $x_{ii} = 1$, and thus the self-reports will bear no information.

Alternatively, we could seek to build a mechanism that incentivizes N_i to provide a *useful* self-report even if it is not the precise truth r_{ii}. With this in mind, a good mechanism might on one hand convince N_i that she can help contribute to a desired, high estimate $\hat{r}_i$ by supplying input x_{ii}, while on the other hand try to use the cross-reports, which are estimates of the truth r_{ii}, to assess N_i's self-report and threaten

with punishment if it is judged to be overly misleading.

Furthermore, we design the mechanism such that N_i's cross-reports are not used in deriving her own reputation. By doing so, we ensure that the cross-reports of N_i affect only her truth elements $f_i(|\hat{r}_j - r_{jj}|)$. It is worth mentioning that in another class of reputation systems, N_i could exploit the *indirect* effect of her cross-report by badmouthing other users so as to improve its *relative* position in the system, i.e., make herself look better by comparison. Here however, there is no clear incentive for N_i to do so, since the current model is one of absolute, rather than proportional/relative reputations. If our proposed mechanism is to be used in a relative reputation setting, the system should not collect self-reports, and instead operate based on its own observations R_{0i} alone.

Consider the following way of computing the reputation index $\hat{r}_i$ for N_i. The system uses its own observation R_{0i}, along with the received cross-reports R_{ji}, for $j,\ i \in \mathcal{I}_j$, to judge N_i's self-report. In the simplest case, the system can take the average of all these estimations to get $\bar{x}_{0i} := \frac{\Sigma_{j:i\in\mathcal{I}_j} x_{ji} + R_{0i}}{T_i+1}$, where $T_i := \|\{k : i \in \mathcal{I}_k\}\|$, and derive $\hat{r}_i$ using:

$$\hat{r}_i(x_{ii}, \bar{x}_{0i}) = \begin{cases} \frac{\bar{x}_{0i}+x_{ii}}{2} & \text{if } x_{ii} \in [\bar{x}_{0i} - \epsilon, \bar{x}_{0i} + \epsilon], \\ \bar{x}_{0i} - |\bar{x}_{0i} - x_{ii}| & \text{if } x_{ii} \notin [\bar{x}_{0i} - \epsilon, \bar{x}_{0i} + \epsilon]. \end{cases} \tag{5.3}$$

where ϵ is a fixed and known constant. In words, the reputation system takes the average of the self-report x_{ii} and the *aggregate cross-report* $\bar{x}_{0i}$ if the two are sufficiently close, or else punishes N_i for reporting significantly differently. We refer to this mechanism as the *punish-reward* mechanism. There are other ways to convey the same idea of weighing between averaging and punishing, such as punishing only when the self-report exceeds the cross-report. We analyze this simple variant in this chapter.

Next we examine the strategic behavior of users when playing the induced game. We will start by assuming that all cross-observations are unbiased and are reported truthfully, i.e., $x_{ji} \sim \mathcal{N}(r_{ii}, \sigma^2),\ j,\ i \in \mathcal{I}_j$ and $R_{0i} \sim \mathcal{N}(r_{ii}, \sigma^2)$.[3] This will allow us

[3]Note that we are assuming σ is common, and known by the reputation system as well as the

to focus on the strategic choice of the self-reports r_{ii}. We return to characterizing the mutual best-response strategies for all participants in Section 5.3.3, where we allow users to strategically bias their cross-reports.

5.3.2 Choice of self-reports in the PR mechanism

As N_i knows the distribution of the observations R_{ji}, she is best-responding to an aggregate cross-report which is a sample of a distribution $\mathcal{N}(\mu, \sigma'^2)$, with $\mu = r_{ii}$ and $\sigma'^2 = \frac{\sigma^2}{T_i+1}$. The choice of self-report x_{ii} is then determined by the solution to the optimization problem $\max_{x_{ii}} E[\hat{r}_i]$.

Using (5.3), $E[\hat{r}_i]$ eventually simplifies to (with $F(\cdot)$ and $f(\cdot)$ denoting the CDF and PDF of $\bar{x}_{0i}$, respectively):

$$E[\hat{r}_i] = \quad x_{ii} + \tfrac{\epsilon}{2}(F(x_{ii}+\epsilon) - 3F(x_{ii}-\epsilon)) - \tfrac{1}{2}\int_{x_{ii}-\epsilon}^{x_{ii}+\epsilon} F(x)\mathrm{d}x - 2\int_{-\infty}^{x_{ii}-\epsilon} F(x)\mathrm{d}x \ . \tag{5.4}$$

Taking the derivative with respect to x_{ii}, we get:

$$\frac{\mathrm{d}E[\hat{r}_i]}{\mathrm{d}x_{ii}} = 1 + \frac{\epsilon}{2}[f(x_{ii}+\epsilon) - 3f(x_{ii}-\epsilon)] - \frac{1}{2}[F(x_{ii}+\epsilon) + 3F(x_{ii}-\epsilon)]. \tag{5.5}$$

We next re-write $\epsilon = a\sigma'$; this expression of ϵ reflects how the reputation system can limit the variation in the self-report using its knowledge of this variation σ'. Replacing $\epsilon = a\sigma'$ and $\bar{x}_{0i} \sim \mathcal{N}(\mu, \sigma'^2)$ in (5.5), and making the change of variable $y := \frac{x_{ii}-\mu}{a\sigma'}$ results in:

$$\frac{a}{\sqrt{2\pi}}(e^{-(\frac{a(y+1)}{\sqrt{2}})^2} - 3e^{-(\frac{a(y-1)}{\sqrt{2}})^2}) - \frac{1}{2}(\mathrm{erf}(\frac{a(y+1)}{\sqrt{2}}) + 3\mathrm{erf}(\frac{a(y-1)}{\sqrt{2}})) = \quad 0 \ . \tag{5.6}$$

Therefore, if y solves (5.6) for a given a, the optimal value for x_{ii} would be $x^*_{ii} = \mu + a\sigma' y$. Equation (5.6) can be solved numerically for a, resulting in Figure 5.1.

participants. This standard deviation σ can be thought of as a measure of the variation of the estimates on N_i, which depends on the nature of the observations and the algorithm used for the estimate.

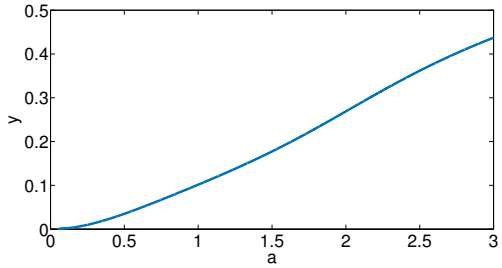

Figure 5.1: Solution of (5.6): y vs. a

Two interesting observations can be made from Figure 5.1: (1) $0 < y < 1$, and (2) as a consequence $\mu < x^*_{ii} < \mu + \epsilon$. This means that N_i chooses to inflate her self-report in hope of inflating $\hat{r}^A_i$, while trying to stay within her prediction of the acceptable range.

5.3.2.1 Properties of the PR mechanism

We first compare the performance of (5.3) to the simple averaging mechanism. Define $e_m := E[|\hat{r}_i - r_{ii}|]$ as the mean absolute error (MAE) of the mechanism described in (5.3) with $\epsilon = a\sigma'$. Assuming the optimal self-report x^*_{ii}, and unbiased, truthful cross-reports, it is possible to find the expression for e_m as a function of the parameter a. We can thus optimize the choice of a by solving the problem $\min_a e_m$. Taking the derivative of e_m we get:

$$\begin{aligned}\frac{\mathrm{d}e_m}{\mathrm{d}a} = \frac{\sigma'}{2}\Big(&\frac{a}{\sqrt{2\pi}}(e^{-\frac{(a(y+1))^2}{2}} - 3e^{-\frac{(a(y-1))^2}{2}}) \\ &+ (ay + y')\big(\mathrm{erf}(\tfrac{ay}{\sqrt{2}}) - \frac{1}{2}(\mathrm{erf}(\tfrac{a(y+1)}{\sqrt{2}}) - 3\mathrm{erf}(\tfrac{a(y-1)}{\sqrt{2}}))\\ &+ \frac{a}{\sqrt{2\pi}}(e^{-\frac{(a(y+1))^2}{2}} + 3e^{-\frac{(a(y-1))^2}{2}}) + 2\big)\Big) = 0\ . \end{aligned} \tag{5.7}$$

As seen in (5.7), the optimal choice of a does not depend on the specific values of μ and σ'. Therefore, the same mechanism can be used for any set of users. Equation (5.7) can be solved numerically, with a result that the minimum error is achieved at $a \approx 1.7$. This can be seen from Figure 5.2, which shows the MAE of

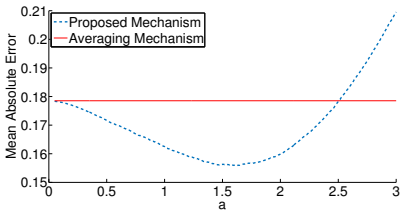

Figure 5.2: Mean absolute error of the PR mechanism vs. simple averaging

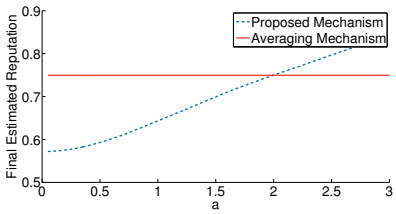

Figure 5.3: Expected reputation of a user in the PR mechanism

the PR mechanism compared to that of the averaging mechanism. Under the simple averaging mechanism the MAE is $E[|\bar{x}_{0i} - r_{ii}|] = \sqrt{\frac{2}{\pi}}\sigma'$. We see that for a large range of a values, the PR mechanism given in (5.3) results in smaller estimation error. This suggests that N_i's self-report can significantly benefit the system, as well as all users other than N_i.

We have now verified the PR mechanism as a suboptimal solution to the problem of determining users' reputations; when designed appropriately, it can outperform the simple averaging mechanism. It is clearly budget balanced as no taxation is imposed. We next check whether there is incentive for N_i to provide her self-report, i.e., does this benefit N_i herself? Figure 5.3 compares N_i's estimated reputation $\hat{r}_i$ under the proposed mechanism to that under the averaging mechanism[4]; the latter is simply the average of all observations on N_i, and is $E[\bar{x}_{0i}] = r_{ii}$ when cross-reports are unbiased.

Taking Figs. 5.2 and 5.3 together, we see that there is a region, $a \in [2, 2.5]$ in which the presence of the self-report helps N_i obtain a higher reputation index, while helping the system reduce its estimation error on N_i. This is a region that is mutually beneficial to both N_i and the system, and N_i clearly has an incentive to participate and provide her self-report.

[4]In calculating the reserved utility, we have assumed that in the event N_i chooses to stay out of the game, the reputation system will use simple averaging on the gathered cross-observations to estimate N_i's reputation.

5.3.3 Choice of cross-reports in the PR mechanism

As shown in Section 5.3.2, each user will introduce a positive bias in her self-report. Predicting this, one may expect other users (and the reputation system) to decrease the effect of this positive bias, by say, negatively biasing their submitted cross-reports. In this section, we formalize this intuition by allowing users to strategically choose their cross-reports. We will model this strategic behavior by the addition of a bias term to the initial observation, so that $x_{ji} = R_{ji} + b_{ji}$. As all users N_j providing cross-reports on N_i are assumed symmetric, they will choose the same bias term, denoted b_{ii}.

As a result, the new aggregate cross-report to be used in the PR mechanism is given by $\bar{x}_{0i} \sim \mathcal{N}(\mu, \sigma'^2)$, where $\mu = r_{ii} + b_{ii}$ and $\sigma'^2 = \frac{\sigma^2}{K}$. The argument to derive the optimal self-report will proceed exactly as in Section 5.3.2, so that N_i will accordingly adapt her self-report to be $x^*_{ii} = r_{ii} + b_{ii} + ay\sigma'$, given other users' strategic behavior.

We next find an expression for e_m, the MAE of the PR mechanism, given the parameter a.[5]

$$\begin{aligned} e_m = &\frac{1}{2}\int_{2r_{ii}-\mu-ay\sigma'}^{\mu+a(y+1)\sigma'} xf(x)\mathrm{d}x - \frac{1}{2}\int_{\mu+a(y-1)\sigma'}^{2r_{ii}-\mu-ay\sigma'} xf(x)\mathrm{d}x - 2\int_{-\infty}^{\mu+a(y-1)\sigma'} xf(x)\mathrm{d}x \\ &+ \mu - r_{ii} + ay\sigma' + (\mu + ay\sigma')(\frac{3}{2}F(\mu + a(y-1)\sigma') - \frac{1}{2}F(\mu + a(y+1)\sigma')) \\ &+ (2r_{ii} - \mu - ay\sigma')\ F(2r_{ii} - \mu - ay\sigma')\ . \end{aligned} \tag{5.8}$$

where $F(\cdot)$ and $f(\cdot)$ are the CDF and PDF of $\bar{x}_{0i}$. To find the value of b at which the error is minimized, we take the derivative of (5.8), resulting in:

$$\frac{\mathrm{d}e_m}{\mathrm{d}\mu} = 1 - 2F(2r_{ii} - \mu - ay\sigma')\ . \tag{5.9}$$

Setting (5.9) equal to zero for a given a, we find that the MAE is minimized at $b^*_{ii} = -\frac{ay\sigma'}{2}$. As a result, each N_j will choose the cross-report $x^*_{ji} = R_{ji} - \frac{ay\sigma'}{2}$.

[5]The following calculations are for moderate values of bias $b_{ii} \in [-ay\sigma', -ay\sigma' + \frac{a\sigma'}{2}]$.

Similarly, the reputation system will choose to input $x_{0i} = R_{0i} - \frac{ay\sigma'}{2}$. By predicting this behavior, N_i will choose the self-report $x^*_{ii} = r_{ii} + \frac{ay\sigma'}{2}$. Therefore, given the mechanism, N_j can strategically choose her cross-report to decrease the estimation error on N_i.

5.4 Discussion and conclusion

As mentioned earlier, several variations on the proposed punish-reward mechanism are possible, and can potentially improve the performance of the reputation mechanism. One straightforward variation would be to use the weighted mean of the cross-reports, instead of a simple average, to generate the aggregate cross-report

$$\bar{x}_{0i} := \frac{\sum_{j:i\in\mathcal{I}_j} w_j x_{ji}}{\sum_{j:i\in\mathcal{I}_j} w_j} \ , \tag{5.10}$$

where $\underline{w} := (w_j)_{j:i\in\mathcal{I}_j}$ is a vector of weights, also specified by the reputation system. One reasonable choice for $\underline{w}$ could be a vector of previously computed reputations $\hat{r}_j$, with the intention of allowing the more reputable users to have a higher influence on the estimates. Similar ideas are commonly used in rating/ranking systems. We proceed by analyzing the performance of this alternative mechanism.

Assume $x_{ji} \sim N(r_{ii}, \sigma^2_{ji})$, i.e., all interested users have an unbiased view of N_i, but with potentially different accuracy as reflected by different values of σ_{ji}, with smaller variances corresponding to more precise estimates. In the special case $\sigma_{ji} = \sigma$, $\forall j,\ i \in \mathcal{I}_j$, it can be shown that the weighted average will (regardless of the choice of $\underline{w}$) *increase* the variance of the aggregated cross-report, and thus the estimation error. What this implies is that users with equally accurate views should be given the same power to affect the outcome.

On the other hand, if σ_{ji}'s are different across users, then choosing $\underline{w}$ such that $\sum_{j:i\in\mathcal{I}_j} {w_j}^2\sigma^2_{ji} \leq \sum_{j:i\in\mathcal{I}_j} \frac{1}{T_i^2}\sigma^2_{ji}$ leads to a lower variance, and thus a lower estimation error. This rearrangement shows clearly that for the inequality to hold, it suffices to put more weight on the smaller σ_{ji}'s, i.e., more weight on those with more accurate observations. Technically this result is to be expected. However, in our context it

points to the following interesting interpretation: more reputable users (higher $\hat{r}_j$) should only be given higher weights if they also have more accurate observations (smaller σ_{ji}), which may or may not be the case. This is a scenario where reputation itself should not carry more voting power. Otherwise the system is better off assigning equal weights to all.

We end this chapter by noting that the punish-reward mechanism is only one possible mechanism to achieve a sub-optimal solution to problem (5.1). We have chosen it for its simplicity and effectiveness. An additional advantage of the PR mechanism is that its design (specifically the choice of a) is independent of the other model parameters (e.g., the true reputations and the number of users), with appropriate choices of its parameter a resulting in a mutually beneficial region for the users and the system, regardless of the problem instance. Most importantly, the PR mechanism only uses the commodity of interest to shape users' incentives, allowing us to forgo the issue of modeling users' valuation of monetary taxation/rewards. It remains an interesting and challenging problem to find a mechanism that results in the *smallest* performance gap, if it exists, compared to the solution to the centralized problem (5.1).

Chapter 6

Conclusion

6.1 A brief review

In this thesis, we investigated the theoretical underpinnings of users' incentives in exerting effort towards the provision of public goods on networks. We studied incentive mechanisms for implementing the socially optimal outcomes in these environments, with a focus on the issue of voluntary participation. We show the importance of accounting for users' outside options when providing non-excludable goods, such as security, as the non-excludability of the good presents additional challenges that are not present in previously studied private or excludable public good provision problems. In particular, we propose the notion of exit equilibrium to describe users' outside options from mechanisms for incentivizing the provision of non-excludable public goods. We use this notion to show the negative result that there exists no tax-based incentive mechanism which can implement the socially optimal effort profile, while guaranteeing voluntary participation and maintaining weak budget balance, in all instances of games of provision of non-excludable goods. By extending this result to risk-averse users purchasing cyber-insurance contracts, we highlight the limitations of using cyber-insurance as a method for improving the state of cyber security.

We then considered a particular class of public good provision games played on networks. We provide necessary and sufficient conditions for the uniqueness of

Nash equilibria of these games, based solely on the network structure. We show that existing results in the literature can be recovered as special cases of our result. We further identify necessary and sufficient conditions for the existence of Nash equilibria for the family of games at the two extremes of our model, namely games of strategic complements and games of strategic substitutes. We establish a connection between agents' efforts at different outcomes of these games, namely Nash equilibria, Pareto efficient outcomes, and semi-cooperative equilibria, and their positions in the interdependence network. We separate the effects of agents' dependencies (outgoing edges in the network) and influences (incoming edges in the network) on their effort decisions, and uncover an alternating effect over walks of different length in the network.

We studied the design of inter-temporal incentives in a particular class of security games, namely security information sharing agreements. We show that if participants condition their future cooperation on the history of their past interactions, where the history consists of announcements by a public monitor who provides (imperfect) assessments of participants' cooperation, then it is possible for firms to play equilibria in which they always truthfully disclose their security information. We show that in the absence of public monitoring, the same result can be attained if firms are provided with a platform to communicate their private assessments of others' adherence to the agreement.

6.2 Future directions

We first discuss additional directions for alleviating the negative result on a simultaneous guarantee on social optimality, voluntary participation, and weak budget balance. One alternative is to consider restricted utility functions and/or network structures, and leverage the additional information in the design of incentive mechanisms. We have considered this alternative by finding families of positive instances for the weighted effort models in Section 2.4.2. Chapter 3 provides additional tools for this analysis. An interesting direction of future work using the framework and results of Chapter 3 is to identify conditions on the network structure under which the

aforementioned requirements on the performance guarantees of incentive mechanisms can be simultaneously satisfied.

Another implication of our negative result is that, when a designer lacks additional information about the specifics of the problem environment and users' preferences, she may choose to forgo the social optimality requirement, instead focusing on reliably attaining a sub-optimal solution while guaranteeing full voluntary participation and weak budget balance. Characterizing mechanisms that can provide a guarantee on the sub-optimality of the solution (e.g., ϵ-close to, or a β-fraction of, the optimal solution), as well as the best attainable sub-optimal solution, remain as directions of future work.

We further considered the alternative of designing inter-temporal incentives by studying public good provision games in a repeated framework. We illustrate the methodology in our study of repeated information sharing agreements in Chapter 4. Possible extensions of this study include the analysis of inter-temporal incentives that use both public and private assessments simultaneously, as well as the benefits of increasing the accuracy of the monitoring. It is also interesting to consider alternative monitoring technologies, in particular those in which the accuracy of the assessment of firms' truthfulness is dependent on the type of underlying (un)disclosed incident.

The inter-temporal incentives proposed for achieving cooperation in security information sharing agreements are to be sustained *among peers*, that is, a perceived deviation by a firm leads her peers to modify their future actions. As a direction of future work, we also propose the design of such inter-temporal incentives in a *principal-agent* framework. Consider a cyber-insurer, who proposes bundles of contracts to the insured. A bundle of contracts purchased by an insured consists of sub-contracts, one of which is *activated*, for the duration of a pre-agreed fraction of the contract term, depending on the (perceived) effort exerted by the insured. For example, when an ongoing threat is discovered, an insured who is perceived to have patched her system against this threat will activate a contract with lower premium and/or higher coverage for that month. An interesting direction of future work is studying the design of these bundles of contracts, such that the insured has incentives to exert effort throughout the term of the contract, especially as the threat

landscape changes.

Finally, we propose possible directions of future work for the study of public good provision games on networks. First, we are interested in the generalization of the current results to other games played on networks, in particular, games in which agents' best replies are non-linear in others' actions. The other interesting direction is using the centrality-effort characterization for conducting comparative statics (e.g., the effects of adding/removing links), as well as for the design of targeted tax/subsidy policies that can incentivize the improved provision of the public good.

Appendices

Appendix A

The Pivotal Mechanism: Social optimality and voluntary participation

We present two propositions to illustrate the main properties of the Pivotal mechanism, namely social optimality and voluntary participation. The proofs follow directly from the classical literature on VCG mechanisms and are included for completeness.

Proposition A.1. *In the Pivotal mechanism with taxes given by* (2.6)*, reporting the true type, i.e., the true utility function $u_i(\cdot)$, is a dominant strategy for all users i. Therefore, the socially optimal solution is implemented.*

Proof. The total utility of user i when reporting $\tilde{u}_i(\cdot)$, while others report $\tilde{u}_j(\cdot)$, $j \neq i$, is given by

$$v_i(\tilde{\mathbf{x}}, t_i) = u_i(\tilde{\mathbf{x}}) + \sum_{j \neq i} \tilde{u}_j(\tilde{\mathbf{x}}) - \sum_{j \neq i} \tilde{u}_j(\hat{\mathbf{x}}^i) \ ,$$

where $\tilde{\mathbf{x}} = \arg\max_{\mathbf{x} \succeq 0} \sum_{k=1}^{N} \tilde{u}_k(\mathbf{x})$ is the allocation that is optimal given the reported types $\tilde{u}_k(\cdot), \forall k$. We first note that the last term is independent of user i's report. Then, as the allocation $\tilde{\mathbf{x}}$ is chosen according to the optimization problem $\arg\max_{\mathbf{x} \succeq 0} \sum_{k=1}^{N} \tilde{u}_k(\mathbf{x})$ over the reported types, the sum of the first and second terms is maximized at $\tilde{u}_i(\cdot) = u_i(\cdot)$. Therefore, users will reveal their true preferences, irrespective of other users' reports. Consequently, the socially optimal investment profile will be prescribed by the mechanism designer. □

Proposition A.2. *The Pivotal mechanism with taxes given by* (2.6) *satisfies voluntary participation.*

Proof. The change in the utility of a user i when staying in vs. opting out of the mechanism is given by

$$\begin{aligned} & v_i(\mathbf{x}^*, t_i) - u_i(\hat{\mathbf{x}}^i) = \\ & \quad u_i(\mathbf{x}^*) + \sum_{j \neq i} u_j(\mathbf{x}^*) - \sum_{j \neq i} u_j(\hat{\mathbf{x}}^i) - u_i(\hat{\mathbf{x}}^i) = \\ & \quad \sum_j u_j(\mathbf{x}^*) - \sum_j u_j(\hat{\mathbf{x}}^i) \geq 0 \ . \end{aligned}$$

The inequality is due to the fact that $\mathbf{x}^*$ is the socially optimal solution given by the maximizer of the sum of all users' utilities. We conclude that it is in the best interest of users to participate in the Pivotal mechanism with the given taxes. □

Appendix B

The Externality Mechanism: Social optimality and budget balance

We present two propositions to illustrate the main properties of the Externality mechanism, namely social optimality and budget balance.

Proposition B.1. *In the Externality mechanism with taxes given by* (2.7)*, investing the socially optimal security effort* $\mathbf{x}^*$ *is a Nash equilibrium.*

Proof. First, note that the taxes assigned to users at a vector of investments $\mathbf{x}$ (possibly off equilibrium) is given by

$$t_i^E(\mathbf{x}) = -\sum_{j=1}^{N} x_j L_i \frac{\partial f_i}{\partial x_j}(\mathbf{x}^*) - x_i \frac{\partial h_i}{\partial x_i}(x_i^*) \ .$$

Now, assume all users other than i are investing at the socially optimal level $\mathbf{x}_{-i}^*$. Then, user i's total utility is

$$\begin{aligned} v_i(x_i, \mathbf{x}_{-i}^*, t_i) = W_i - L_i f_i(x_i, \mathbf{x}_{-i}^*) - h_i(x_i) \\ + x_i L_i \frac{\partial f_i}{\partial x_j}(x^*) + \sum_{j \neq i}^{N} x_j^* L_i \frac{\partial f_i}{\partial x_j}(\mathbf{x}^*) + x_i \frac{\partial h_i}{\partial x_i}(x_i^*) \ . \end{aligned}$$

The first derivative of i's utility with respect to x_i is given by

$$\begin{aligned}\frac{\partial v_i(x_i, \mathbf{x}^*_{-i}, t_i)}{\partial x_i} &= -L_i \frac{\partial f_i}{\partial x_j}(x_i, \mathbf{x}^*_{-i}) - \frac{\partial h_i}{\partial x_i}(x_i) \\ &\quad + L_i \frac{\partial f_i}{\partial x_j}(\mathbf{x}^*) + \frac{\partial h_i}{\partial x_i}(x_i^*) \ .\end{aligned}$$

We conclude that x_i^* is a best response for user i, and hence the socially optimal effort profile $\mathbf{x}^*$ is a Nash equilibrium given the Externality taxes (2.7). □

Proposition B.2. *The Externality mechanism with taxes given by* (2.7) *has strong budget balance.*

Proof. The sum of taxes in (2.7) is given by

$$\begin{aligned}\sum_{i=1}^N t_i^E(\mathbf{x}^*) &= -\sum_{i=1}^N \sum_{j=1}^N x_j^* L_i \frac{\partial f_i}{\partial x_j}(\mathbf{x}^*) - \sum_{i=1}^N x_i^* \frac{\partial h_i}{\partial x_i}(x_i^*) \\ &= \sum_{i=1}^N x_i^* \left(-\sum_{j=1}^N L_j \frac{\partial f_j}{\partial x_i}(\mathbf{x}^*) - \frac{\partial h_i}{\partial x_i}(x_i^*) \right) = 0 \ . \qquad \text{(B.1)}\end{aligned}$$

The last line follows from the observation that, using (2.2), at the socially optimal solution, either x_i^* or the term inside the parentheses in (B.1) is zero. Hence, the Externality mechanism guarantees (strong) budget balance. □

Appendix C

Effects of self-dependence in weighted effort games

In this appendix, we consider the security game where users' utilities are given by (2.8) and interdependence matrix (2.9). We solve for the socially optimal investment profile, and identify the possible exit equilibria, and parameter conditions under which each equilibrium is possible.

The socially optimal investment profile in this game will be given by

$$x_i^* = \frac{1}{a+N-1} \ln \frac{a+N-1}{c}, \forall i \ .$$

To find the exit equilibrium for user i, $\hat{\mathbf{x}}^i$, we write the first order conditions on (2.4). To simplify notation, denote $x := \hat{x}_i^i$ and $y := \hat{x}_j^i, \forall j \neq i$. The system of equation determining x and y is given by

$$\begin{aligned} -a\exp(-ax-(N-1)y)+c \geq 0 \ , \\ -(a+N-2)\exp(-x-(a+N-2)y)+c \geq 0 \ . \end{aligned} \tag{C.1}$$

There are four possible exit equilibria, depending on whether x and/or y are non-zero. We look at each case separately.

Exit equilibria with $x > 0, y > 0$ Intuitively, when user i steps out, both sides continue to invest in security, perhaps at reduced levels, but no user is fully free-

riding. We would need the following to hold simultaneously:

$$\begin{aligned} -a\exp(-ax-(N-1)y)+c=0\ , \\ -(a+N-2)\exp(-x-(a+N-2)y)+c=0\ . \end{aligned}$$

Solving for x, y leads to

$$\begin{aligned} x &= \frac{1}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)}\ , \\ y &= \frac{1}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{a}\ . \end{aligned}$$

To find the range of parameters for which the above holds, we need to ensure that x, y are indeed positive.

- If $a > 1$, then $y > 0$. For $x > 0$, we need

$$(\frac{a}{c})^{a-1} > (1+\frac{N-2}{a})^{N-1}\ .$$

- If $a < 1$, then $x > 0$. For $y > 0$, we need

$$(1+\frac{N-2}{a})^{a} < (\frac{a}{c})^{1-a}\ .$$

Exit equilibria with $x > 0, y = 0$ In this case, the participating users revert to investing zero, so that the outlier is forced to increase her investment:

$$\begin{aligned} -a\exp(-ax)+c=0\ , \\ -(a+N-2)\exp(-x)+c>0\ . \end{aligned}$$

As a result, we get $x = \frac{1}{a}\ln\frac{a}{c}$. For this to be consistent with the second condition, we require

$$(1+\frac{N-2}{a})^{a} < (\frac{a}{c})^{1-a}\ .$$

The above always fails to hold for $a > 1$, as the LHS is always more than 1, while the RHS is surely less than 1 by the assumption $a > c$. Intuitively, when self-dependence is higher than co-dependence on the outlier, the remaining users will not rely solely on externalities, and continue investing even when user i steps out.

For $a < 1$ on the other hand, for a small enough c (which leads to higher investment x be the outlier), the equation can hold.

Exit equilibria with $x = 0, y > 0$ This means that the loner free-rides, so that we have

$$\begin{aligned} -a\exp(-(N-1)y) + c > 0 \ , \\ -(a+N-2)\exp(-(a+N-2)y) + c = 0 \ . \end{aligned}$$

As a result, we get $y = \frac{1}{a+N-2}\ln\frac{a+N-2}{c}$. For this to be consistent with the first condition, we need

$$(1+\frac{N-2}{a})^{N-1} > (\frac{a}{c})^{a-1} \ .$$

Note that this always hold for $a < 1$, but not necessarily for $a > 1$.

Exit equilibria with $x = 0, y = 0$ We would need the following to hold simultaneously:

$$\begin{aligned} -a + c > 0 \ , \\ -(a+N-2) + c > 0 \ , \end{aligned}$$

which will never hold, as we initially required that $c < a$.

Weak budget balance and voluntary participation constraints We now separately analyze each of the possible cases identified in the previous section, summarized in Table 2.1. Specifically, we are interested in the weak budget balance condition under the Pivotal mechanism, and users' participation incentives in the Externality mechanism.

Case α In this case, the underlying parameters satisfy $a > 1$ and $(1 + \frac{N-2}{a})^{N-1} > (\frac{a}{c})^{a-1}$. As a result, the exit equilibrium (EE) is such that $x = 0$, and $y = \frac{1}{a+N-2} \ln \frac{a+N-2}{c}$. Therefore, the utilities of users at the SO and EE are given by

$$
\begin{aligned}
u_j(\mathbf{x}^*) &= W - \frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}), \forall j \\
u_j(\hat{\mathbf{x}}^i) &= W - \frac{c}{a+N-2}(1 + \ln \frac{a+N-2}{c}), \forall j \neq i \\
u_i(\hat{\mathbf{x}}^i) &= W - \frac{c}{a+N-2}^{\frac{N-1}{a+N-2}} .
\end{aligned}
$$

Weak Budget Balance in the Pivotal mechanism Note that $\frac{1+\ln z}{z}$ is a decreasing function of z. Thus, $u_j(\hat{\mathbf{x}}^i) > u_j(\mathbf{x}^*)$ for all j, resulting in $t_i^P < 0$, indicating rewards to all users i, and thus a budget deficit in all scenarios. Intuitively, although when a user i steps out, other users have to invest less in security (thus decreasing their direct investment costs), still their overall security costs go up as a result of the increased risks. Consequently, each user i should be payed a reward to be kept in the mechanism, resulting in a budget deficit.

Voluntary Participation in the Externality mechanism Voluntary participation will hold if and only if $u_i(\hat{\mathbf{x}}^i) \leq v_i(\mathbf{x}^*, t_i^E)$, that is,

$$
\begin{aligned}
& \frac{c}{a+N-2}^{\frac{N-1}{a+N-2}} \geq \frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}) \\
\Leftrightarrow & \frac{c}{a+N-2}^{N-1} \geq (\frac{c}{a+N-1})^{a-1+N-1}(1 + \ln \frac{a+N-1}{c})^{a-1+N-1} \\
\Leftrightarrow & (\frac{a+N-1}{a+N-2})^{N-1}(1 + \frac{N-1}{a})^{a-1}(\frac{a}{c})^{a-1} - (1 + \ln \frac{a}{c}(1 + \frac{N-1}{a}))^{a+N-2} \geq 0
\end{aligned}
$$

Based on the last inequality, define the function $g(z) := \kappa_1 z^{a-1} - (1 + \ln z)^{a+N-2}$. This function is increasing in z. As a result, it obtains its maximum when z reaches its maximum value, which by the initial condition is given by $\frac{a}{c} = (1 + \frac{N-2}{a})^{\frac{N-1}{a-1}}$.

Thus,

$$\begin{aligned} g_{max} &= (\frac{a+N-1}{a+N-2})^{N-1}(1+\frac{N-1}{a})^{a-1}(1+\frac{N-2}{a})^{N-1} \\ &\qquad - (1+\ln(1+\frac{N-2}{a})^{\frac{N-1}{a-1}}(1+\frac{N-1}{a}))^{a+N-2} \\ &\leq (1+\frac{N-1}{a})^{a+N-2} - (1+\ln(1+\frac{N-1}{a})+\frac{N-1}{a-1}\ln(1+\frac{N-2}{a}))^{a+N-2} \\ &\leq (1+\frac{N-1}{a})^{a+N-2} - (1+\ln(1+\frac{N-1}{a})+\frac{N-1}{a}\ln(1+\frac{N-2}{a}))^{a+N-2} \end{aligned}$$

Let $z := \frac{N-1}{a}$, and define $f(z) := \ln(1+z) + z\ln(1+z-\frac{1}{a}) - z$ (i.e., we are assuming a fixed a). The derivative of this function wrt z is given by

$$\frac{1}{1+z} + \ln(1+z-\frac{1}{a}) + \frac{z}{1+z-\frac{1}{a}} - 1 = \ln(1+z-\frac{1}{a}) + \frac{\frac{1}{a}z}{(1+z)(1-\frac{1}{a}+z)} \ .$$

As the above is positive for all $a > 1$, we conclude that $f(z)$ is an increasing function in z. Furthermore, $\lim_{z\to 0} f(z) = 0$, which in turn means that $f(z) \geq 0, \forall z \geq 0$, and therefore, g_{max} is always non-positive. This in turn means that the voluntary participation condition can never be satisfied.

Case β For this case, the underlying parameters satisfy $a > 1$ and $(1+\frac{N-2}{a})^{N-1} < (\frac{a}{c})^{a-1}$. As a result, the exit equilibrium (EE) is such that $x > 0, y > 0$, and are given by $x = \frac{1}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)}$ and $y = \frac{1}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{a}$. Therefore, the utilities of users at the SO and EE are given by:

$$\begin{aligned} u_j(\mathbf{x}^*) &= W - \frac{c}{a+N-1}(1+\ln\frac{a+N-1}{c}), \forall j \\ u_j(\hat{\mathbf{x}}^i) &= W - \frac{c}{a+N-2} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{a}, \forall j \neq i \\ u_i(\hat{\mathbf{x}}^i) &= W - \frac{c}{a} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)} \ . \end{aligned}$$

Weak Budget Balance in the Pivotal mechanism If $u_j(\hat{\mathbf{x}}^i) \leq u_j(\mathbf{x}^*)$, the mechanism would always have a budget deficit. This holds if and only if

$$\begin{aligned}
&\frac{c}{a+N-1}(1+\ln\frac{a+N-1}{c}) \leq \frac{c}{a+N-2} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^a \\
\Leftrightarrow &1+\ln\frac{a}{c}(1+\frac{N-1}{a}) \leq 1+\frac{1}{a+N-2}+\ln\frac{a}{c}(1+\frac{N-2}{a})^{\frac{a}{a-1}} \\
\Leftrightarrow &\ln(1+\frac{N-1}{a}) \leq \frac{1}{a+N-2}+\frac{a}{a-1}\ln(1+\frac{N-2}{a}) \\
\Leftarrow &\ln(1+\frac{N-1}{a}) \leq \frac{1}{a+N-2}+\ln(1+\frac{N-2}{a}) \\
\Leftrightarrow &\ln(1+\frac{1}{a+N-2}) \leq \frac{1}{a+N-2}
\end{aligned}$$

The last line is true because $\ln(1+x) \leq x$, for all $x > 0$. Therefore, the mechanism always carries a budget deficit.

Voluntary Participation in the Externality mechanism The mechanism fails voluntary participation if and only if

$$\begin{aligned}
&\frac{c}{a+N-1}(1+\ln\frac{a+N-1}{c}) \geq \frac{c}{a} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)} \\
\Leftrightarrow &1+\ln\frac{a}{c}(1+\frac{N-1}{a}) \geq 1+\frac{N-1}{a}+\ln\frac{a}{c}(1+\frac{N-2}{a})^{\frac{-(N-1)}{a-1}} \\
\Leftrightarrow &\ln(1+\frac{N-1}{a})+\frac{N-1}{a-1}\ln(1+\frac{N-2}{a}) \geq \frac{N-1}{a} \\
\Leftarrow &\ln(1+\frac{N-1}{a})+\frac{N-1}{a}\ln(1+\frac{N-1}{a}-\frac{1}{a}) \geq \frac{N-1}{a}
\end{aligned}$$

Let $z := \frac{N-1}{a}$, and define $f(z) := \ln(1+z)+z\ln(1+z-\frac{1}{a})-z$ (i.e., we are assuming a fixed a). The derivative of this function wrt z is given by $\ln(1+z-\frac{1}{a})+\frac{\frac{1}{a}z}{(1+z)(1-\frac{1}{a}+z)}$. As this is positive for all $a > 1$, we conclude that $f(z)$ is an increasing function in z. Furthermore, $\lim_{z\to 0} f(z) = 0$, which in turn means that $f(z) \geq 0, \forall z \geq 0$, and therefore, that the voluntary participation condition always fails to hold under these parameter settings.

Case γ Here, we only require that $a < 1$, and all other values of N or c will guarantee the existence of an equilibrium $x = 0$ and $y = \frac{1}{a+N-2} \ln \frac{a+N-2}{c}$. This is thus parallel with Case α. Users' utilities in the SO and EE are similarly given by

$$
\begin{aligned}
u_j(\mathbf{x}^*) &= W - \frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}), \forall j \\
u_j(\hat{\mathbf{x}}^i) &= W - \frac{c}{a+N-2}(1 + \ln \frac{a+N-2}{c}), \forall j \neq i \\
u_i(\hat{\mathbf{x}}^i) &= W - \frac{c}{a+N-2}^{\frac{N-1}{a+N-2}} .
\end{aligned}
$$

Weak Budget Balance in the Pivotal mechanism Note that $\frac{1+\ln z}{z}$ is a decreasing function of z. Thus, $u_j(\hat{\mathbf{x}}^i) < u_j(\mathbf{x}^*)$ for all j, resulting in $t_i^P < 0$, indicating rewards to all users i, and thus a budget deficit in all scenarios (exactly similar to case α).

Voluntary Participation in the Externality mechanism Voluntary participation will fail to hold if and only if $u_i(\hat{\mathbf{x}}^i) \geq v_i(\mathbf{x}^*, t_i^E)$, that is,

$$
\begin{aligned}
& \frac{c}{a+N-2}^{\frac{N-1}{a+N-2}} \leq \frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}) \\
\Leftrightarrow & \frac{c}{a+N-2}^{N-1} \leq (\frac{c}{a+N-1})^{a-1+N-1}(1 + \ln \frac{a+N-1}{c})^{a-1+N-1} \\
\Leftrightarrow & (\frac{\frac{a}{c}(1 + \frac{N-1}{a})}{1 + \ln \frac{a}{c}(1 + \frac{N-1}{a})})^{a-1} \geq (\frac{1 + \ln \frac{a}{c}(1 + \frac{N-1}{a})}{1 + \frac{1}{a+N-2}})^{N-1}
\end{aligned}
$$

First, we note that the RHS is always greater than 1, as $1 + \ln x \leq x$. On the other hand, since $a < 1$, $\frac{1}{a+N-2} < \ln \frac{a}{c}(1 + \frac{N-1}{a})$ holds for all $N \geq 3$, so that the LHS will be less than 1. Therefore, the voluntary participation condition always fails.

Case ζ This case has equilibrium investments similar to case β, but under parameter conditions $a < 1$, and $(1 + \frac{N-2}{a})^a < (\frac{a}{c})^{1-a}$. Therefore, we have

$$u_j(\mathbf{x}^*) = W - \frac{c}{a+N-1}(1 + \ln\frac{a+N-1}{c}), \forall j$$
$$u_j(\hat{\mathbf{x}}^i) = W - \frac{c}{a+N-2} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^a, \forall j \neq i$$
$$u_i(\hat{\mathbf{x}}^i) = W - \frac{c}{a} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)} .$$

Weak Budget Balance in the Pivotal mechanism If $u_j(\hat{\mathbf{x}}^i) \geq u_j(\mathbf{x}^*)$, we would always have weak budget balance. This holds if and only if

$$\frac{c}{a+N-1}(1+\ln\frac{a+N-1}{c}) \geq \frac{c}{a+N-2} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^a$$
$$\Leftrightarrow 1 + \ln\frac{a}{c}(1+\frac{N-1}{a}) \geq 1 + \frac{1}{a+N-2} + \ln\frac{a}{c}(1+\frac{N-2}{a})^{\frac{a}{a-1}}$$
$$\Leftrightarrow \ln(1+\frac{N-1}{a}) \geq \frac{1}{a+N-2} + \frac{a}{a-1}\ln(1+\frac{N-2}{a})$$
$$\Leftarrow \ln(1+\frac{N-1}{a}) \geq \frac{1}{a+N-2}$$

The last line follows from the previous because $a < 1$, and is true because its LHS is $\geq \ln N$ and its RHS is $\leq 1/(N-1)$. Therefore, the mechanism always has weak budget balance in this scenario.

Voluntary Participation in the Externality mechanism The mechanism has voluntary participation if and only if

$$\frac{c}{a+N-1}(1+\ln\frac{a+N-1}{c}) \leq \frac{c}{a} + \frac{c}{(a-1)(a+N-1)}\ln(\frac{a}{c})^{a-1}(1+\frac{N-2}{a})^{-(N-1)}$$
$$\Leftrightarrow 1 + \ln\frac{a}{c}(1+\frac{N-1}{a}) \leq 1 + \tfrac{N-1}{a} + \ln\frac{a}{c}(1+\frac{N-2}{a})^{\frac{-(N-1)}{a-1}}$$
$$\Leftrightarrow \ln(1+\frac{N-1}{a})(1+\frac{N-2}{a})^{\frac{N-1}{a-1}} \leq \frac{N-1}{a}$$

The last statement holds because the second element in the logarithm is always less than 1, due to $a < 1$, and the result follows as $\ln(1+z) \leq z$, for all $z > 0$.

Case ω The last case emerges under parameter settings $a < 1$ and $(1 + \frac{N-2}{a})^a < (\frac{a}{c})^{1-a}$, and $x = \ln \frac{a}{c}$ and $y = 0$ is the possible exit equilibrium. The users' utilities in the SO and EE here are given by

$$\begin{aligned} u_j(\mathbf{x}^*) &= W - \frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}), \forall j \\ u_j(\hat{\mathbf{x}}^i) &= W - (\frac{c}{a})^{\frac{1}{a}}, \forall j \neq i \\ u_i(\hat{\mathbf{x}}^{\mathbf{i}}) &= W - \frac{c}{a}(1 + \ln \frac{a}{c}) \ . \end{aligned}$$

Weak Budget Balance in the Pivotal mechanism First we use $(1 + \frac{N-2}{a})^a < (\frac{a}{c})^{1-a}$ to conclude that $(\frac{c}{a})^{\frac{1}{a}} \leq \frac{c}{a+N-2}$. Now, for the mechanism to have weak budget balance it would be enough to have $u_j(\hat{\mathbf{x}}^i) \geq u_j(\mathbf{x}^*)$, which holds if and only if

$$\begin{aligned} &\frac{c}{a+N-1}(1 + \ln \frac{a+N-1}{c}) \geq (\frac{c}{a})^{\frac{1}{a}} \\ \Leftarrow &1 + \ln \frac{a}{c}(1 + \frac{N-1}{a}) \geq 1 + \frac{1}{a+N-2} \\ \Leftarrow &\ln(1 + \frac{N-1}{a}) \geq \frac{1}{a+N-2} \end{aligned}$$

where the last line line follows from the previous because $\frac{a}{c} > 1$, and is true because its LHS is $\geq \ln N$ and its RHS is $\leq 1/(N-1)$. Therefore, the mechanism always has weak budget balance in this scenario.

Voluntary Participation in the Externality mechanism As $\frac{1+\ln x}{x}$ is a decreasing function in x when $x > 1$, and $1 < \frac{a}{c} < \frac{a+N-1}{c}$, the utilities when staying out are lower for user i. Therefore voluntary participation is satisfied in the Externality mechanism in this case.

Appendix D

Effects of a dominant user in weighted effort games

In this appendix, we consider the weighted effort game with interdependence matrix (2.10), and solve for the socially optimal investment profile, and identify the possible exit equilibria, and parameter conditions under which each equilibrium is possible. It is straightforward to show that in a socially optimal investment profile $\mathbf{x}^*$, only user 1 will be exerting effort, so that

$$x_1^* = \frac{1}{a} \ln \frac{aN}{c}, \quad x_j^* = 0, \forall j = 2, \ldots, N \ .$$

We next find the exit equilibria. First, if any non-dominant user $i \neq 1$ steps out of the mechanism, user 1 will continue exerting all effort, but at a lower level given by

$$\hat{x}_1^i = \frac{1}{a} \ln \frac{a(N-1)}{c}, \quad \hat{x}_j^i = 0, \forall j = 2, \ldots, N \ .$$

Next, if user 1 steps out of the mechanism, there are two possible exit equilibria: if $a > N - 1$, there will be enough externality for users $j \neq 1$ to continue free-riding, resulting in the following equilibrium investment levels:

$$\hat{x}_1^1 = \frac{1}{a} \ln \frac{a}{c}, \quad \hat{x}_j^1 = 0, \forall j = 2, \ldots, N \ .$$

However, when $a < N - 1$, user 1 will free-ride on the externality of other users'

investments, leading to the exit equilibrium

$$\hat{x}_1^1 = 0, \quad \hat{x}_j^1 = \frac{1}{N-1} \ln \frac{N-1}{c}, \forall j = 2, \ldots, N \ .$$

Voluntary participation in the Externality mechanism We now analyze the performance of the Pivotal and Externality mechanisms, under the different exit equilibria identified in the previous section, and summarized in Table 2.2.

In the Externality mechanism, users' taxes are given by

$$t_1^E(\mathbf{x}^*) = cx_1^*(\frac{1}{N} - 1) \ ,$$
$$t_j^E(\mathbf{x}^*) = \frac{c}{N} x_1^*, \forall j = 2, \ldots, N \ .$$

For non-dominant users $i \in \{2, \ldots, N\}$ to voluntarily participate in the mechanism, we require $u_i(\hat{\mathbf{x}}^i) \leq v_i(\mathbf{x}^*, t_i^E(\mathbf{x}^*))$, which reduces to

$$\frac{c}{a(N-1)} \geq \frac{c}{aN} + \frac{c}{aN} \ln \frac{aN}{c} \quad \Leftrightarrow \quad \frac{1}{N-1} \geq \ln N + \ln \frac{a}{c} \ .$$

However, $\ln N \geq \frac{1}{N-1}, \forall N \geq 3$, and $a > c$. Therefore, the voluntary participation constraints will always fail to hold for free-riders in the Externality mechanism.

A perhaps more interesting aspect is that the voluntary participation of user 1, i.e., the main investor who is receiving a reward, may also fail to hold. Specifically, when $a < N - 1$, user 1 will participate voluntarily if and only if $u_1(\hat{\mathbf{x}}^1) \leq v_1(\mathbf{x}^*, t_1^E(\mathbf{x}^*))$, which reduces to

$$\frac{c}{N-1} \geq \frac{c}{aN} + \frac{c}{aN} \ln \frac{aN}{c} \ .$$

However, the above inequality does not necessarily hold. For example, with $N = 10$, $c = 0.45$, and $a < 5$, the above will fail to hold, indicating that the main investor will also prefer to opt out. It is also interesting to mention that when $a > N - 1$, the voluntary participation of the main investor always holds.

Weak budget balance in the Pivotal mechanism Finally, we analyze the total budget in the Pivotal mechanism. The taxes for the non-dominant users $i \neq 1$ will be given by

$$t_i^P = \frac{c}{a}(\ln \frac{N}{N-1} - 1) \ .$$

The taxes for user 1 will depend on the realized exit equilibrium. If $a < N-1$, this tax is given by

$$t_1^P = (N-1)\frac{c}{aN} - c(1 + \ln \frac{N-1}{c}) \ .$$

The sum of the Pivotal taxes under this parameter conditions will then be given by

$$\sum_i t_i^P = c(\frac{N-1}{a}(\ln \frac{N}{N-1} - 1 + \frac{1}{N}) - (1 + \ln \frac{N-1}{c}))$$

Note that $\ln z - \frac{1}{z} < 0, \forall z < \frac{3}{2}$, and therefore, with $N \geq 3$, the above sum is always negative. We conclude that the Pivotal mechanism will always carry a deficit.

On the other hand, when $a > N-1$, the tax for the dominant user is given by

$$t_1^P = (N-1)\frac{c}{aN} - (N-1)\frac{c}{a} = (N-1)\frac{c}{a}(\frac{1}{N} - 1) \ .$$

The sum of the Pivotal taxes will then be given by

$$\sum_i t_i^P = \frac{c(N-1)}{a}\left(-1 + \ln \frac{N}{N-1} - 1 + \frac{1}{N}\right)$$

By the same argument as before, the above sum will always be negative, indicating a budget deficit in the Pivotal mechanism under this scenario as well.

Appendix E

Proof of Proposition 5.1

We need to show that N_i's expected utility when choosing the message x_{ii} under any strategy profile $\{x_{jj}\}_{j\neq i}$ for all other users, is maximized at $x_{ii} = r_{ii}$. N_i's expected utility is given by

$$\begin{aligned} E\left[v_i(x_{ii}, \{x_{jj}\}_{j\neq i})\right] = & - \sum_{j\in\mathcal{I}_i} E[f_i(|\hat{r}_j - r_{jj}|)] \\ -E[\ |x_{ii} - R_{0i}|^2] & + \tfrac{1}{K-1}\sum_{j\neq i} E[|x_{jj} - R_{0j}|^2] \end{aligned} \tag{E.1}$$

It can be easily seen that N_i's report x_{ii} can only adjust the second term in (E.1) regardless of other participants' strategies. N_i is a self-utility maximizer, therefore x_{ii} is chosen so as to minimize the second term, i.e., minimize the punishment due to discrepancy with respect to the system's observation. By assumption, N_i knows that $R_{0i} \sim \mathcal{N}(r_{ii}, \sigma_0^2)$, and it is easy to see that the optimal choice is indeed $x_{ii} = r_{ii}$.

Bibliography

[1] Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, pages 1041–1063, 1990.

[2] Airmic. Airmic review of recent developments in the cyber insurance market, 2013.

[3] Ross Anderson. Why information security is hard– an economic perspective. In *The Proceedings of the 17th Annual Computer Security Applications Conference*, pages 358–365. IEEE, 2001.

[4] Christian Anschuetz. The Weakest Link Is Your Strongest Security Asset. `http://blogs.wsj.com/cio/2015/02/26/the-weakest-link-is-your-strongest-security-asset/`. Retrieved on 04-17-2016.

[5] David B. Audretsch and Maryann P. Feldman. Knowledge spillovers and the geography of innovation. *Handbook of Regional and Urban Economics*, 4:2713–2739, 2004.

[6] Australian Signals Directorate. Strategies to Mitigate Targeted Cyber Intrusions. `http://www.asd.gov.au/infosec/mitigationstrategies.htm`. Retrieved on 04-17-2016.

[7] Coralio Ballester and Antoni Calvó-Armengol. Interactions with hidden complementarities. *Regional Science and Urban Economics*, 40(6):397–406, 2010.

[8] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. Who's who in networks. wanted: the key player. *Econometrica*, 74(5):1403–1417, 2006.

[9] Coralio Ballester, Yves Zenou, and Antoni Calvó-Armengol. Delinquent networks. *Journal of the European Economic Association*, 8(1):34–61, 2010.

[10] Batterley. The Betterley report: Cyber/privacy insurance market survey, June 2012.

[11] Abraham Berman and Robert J. Plemmons. *Nonnegative matrices in the mathematical sciences*, volume 9. SIAM, 1994.

[12] Rainer Böhme and Galina Schwartz. Modeling cyber-insurance: Towards a anifying framework. In *Workshop on the Economics of Information Security (WEIS)*, 2010.

[13] Phillip Bonacich. Power and centrality: A family of measures. *American Journal of Sociology*, pages 1170–1182, 1987.

[14] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social Networks*, 23(3):191–201, 2001.

[15] Phillip Bonacich, Gerald H Shure, James P. Kahan, and Robert J Meeker. Cooperation and group size in the n-person prisoners' dilemma. *Journal of Conflict Resolution*, 20(4):687–706, 1976.

[16] Yann Bramoullé and Rachel Kranton. Public goods in networks. *Journal of Economic Theory*, 135(1):478–494, 2007.

[17] Yann Bramoullé and Rachel Kranton. Games played on networks. In *The Oxford Handbook on the Economics of Networks*. Oxford University Press, 2015.

[18] Yann Bramoullé, Rachel Kranton, and Martin D'amours. Strategic interaction and networks. *The American Economic Review*, 104(3):898–930, 2014.

[19] Ted Bridis. Sony Emails Show a Studio Ripe for Hacking. `http://www.apnewsarchive.com/2014/Sony-emails-reveal-loose-use-of-passwords-and-IDs-ripe-for-hacking/id-041c9dc46e9d408fa569ccac15c0ffe0`. Retrieved on 04-17-2016.

[20] Katherine Campbell, Lawrence A. Gordon, Martin P. Loeb, and Lei Zhou. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security*, 11(3):431–448, 2003.

[21] Huseyin Cavusoglu, Birendra Mishra, and Srinivasan 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*, 9(1):70–104, 2004.

[22] Composite Blocking List. http://cbl.abuseat.org/.

[23] Cisco. As Strong as the Weakest Link. http://www.cisco.com/cisco/web/solutions/small_business/resource_center/articles/secure_my_business/as_strong_as_the_weakest_link/index.html. Retrieved on 04-17-2016.

[24] Thomas Claburn. 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. Retrieved on 02-24-2016.

[25] Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[26] Summary description of the CNCI. https://www.whitehouse.gov/issues/foreign-policy/cybersecurity/national-initiative, 2015. Retrieved on 10-26-2015.

[27] Olivier Compte. Communication in repeated games with imperfect private monitoring. *Econometrica*, pages 597–626, 1998.

[28] Jacomo Corbo, Antoni Calvó-Armengol, and David C. Parkes. The importance of network topology in local contribution games. In *Internet and Network Economics*, pages 388–395. Springer, 2007.

[29] Robert Cordray. Top 5 cyber breaches in 2014. http://www.itbusiness.ca/blog/top-5-high-profile-cyber-security-breaches-that-have-affected-millions/52793. Retrieved on 04-17-2016.

[30] Richard W. Cottle and George B. Dantzig. Complementary pivot theory of mathematical programming. *Linear Algebra and Its Applications*, 1(1):103–125, 1968.

[31] Chandler Davis. Theory of positive linear dependence. *American Journal of Mathematics*, 76(4):733–746, 1954.

[32] Verizon Enterprise Data Breach Investigations Reports. http://www.verizonenterprise.com/DBIR/.

[33] Verizon 2015 Data Breach Investigations Report. http://www.verizonenterprise.com/DBIR/2015/. Retrieved on 04-17-2016.

[34] Cybersecurity Insurance - Homeland Security. http://www.dhs.gov/cybersecurity-insurance. Retrieved on 04-17-2016.

[35] Information sharing and analysis organizations (ISAOs). http://www.dhs.gov/isao, 2015. Retrieved on 10-26-2015.

[36] Jamie Dimon. Annual Letter to J.P. Morgan Shareholders. http://online.wsj.com/public/resources/documents/040913dimon.pdf, 2014. Retrieved on 04-17-2016.

[37] DShield. http://www.dshield.org/.

[38] Matthew Elliott and Benjamin Golub. A network approach to public goods. In *The Proceedings of the 14th ACM Conference on Electronic Commerce*, pages 377–378. ACM, 2013.

[39] Matthew Elliott and Benjamin Golub. A network approach to public goods. *Available at SSRN 2436683*, 2015.

[40] Executive order 13636: Improving critical infrastructure cybersecurity. https://www.whitehouse.gov/the-press-office/2013/02/12/executive-order-improving-critical-infrastructure-cybersecurity, 2013. Retrieved on 10-26-2015.

[41] 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. Retrieved on 10-26-2015.

[42] Fact sheet: Administration cybersecurity efforts 2015. https://www.whitehouse.gov/the-press-office/2015/07/09/fact-sheet-administration-cybersecurity-efforts-2015, 2015. Retrieved on 10-26-2015.

[43] Andrew D. Foster and Mark R. Rosenzweig. Microeconomics of technology adoption. *Annual Review of Economics*, 2, 2010.

[44] James H. Fowler and Nicholas A. Christakis. Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham Heart Study. *Bmj*, 337:a2338, 2008.

[45] Drew Fudenberg, David Levine, and Eric Maskin. The folk theorem with imperfect public information. *Econometrica*, 62(5):997–1039, 1994.

[46] Drew Fudenberg and Eric Maskin. The folk theorem in repeated games with discounting or with incomplete information. *Econometrica*, pages 533–554, 1986.

[47] Esther Gal-Or and Anindya Ghose. The economic incentives for sharing security information. *Information Systems Research*, 16(2):186–208, 2005.

[48] Dwight J. Goehring and James P. Kahan. The uniform n-person prisoner's dilemma game construction and test of an index of cooperation. *Journal of Conflict Resolution*, 20(1):111–128, 1976.

[49] Sharad Goel, Daniel M. Reeves, and David M. Pennock. Collective revelation: A mechanism for self-verified, weighted, and truthful predictions. In *The Proceedings of the 10th ACM conference on Electronic Commerce*, pages 265–274. ACM, 2009.

[50] Lawrence A. Gordon, Martin P. Loeb, and William Lucyshyn. Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy*, 22(6):461–485, 2003.

[51] Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, and Tashfeen Sohail. The impact of the Sarbanes-Oxley Act on the corporate disclosures of information security activities. *Journal of Accounting and Public Policy*, 25(5):503–530, 2006.

[52] Lawrence A. Gordon, Martin P. Loeb, William Lucyshyn, and Lei Zhou. The impact of information sharing on cybersecurity underinvestment: A real options perspective. *Journal of Accounting and Public Policy*, 34(5):509–519, 2015.

[53] Lawrence A. Gordon, Martin P. Loeb, and Tashfeen Sohail. A framework for using insurance for cyber-risk management. *Communications of the ACM*, 46(3):81–85, 2003.

[54] Jens Grossklags, Nicolas Christin, and John Chuang. Secure or insure?: A game-theoretic analysis of information security games. In *The Proceedings of the 17th International Conference on World Wide Web*, pages 209–218. ACM, 2008.

[55] Jens Grossklags, Svetlana Radosavac, Alvaro A. Cárdenas, and John Chuang. Nudge: Intermediaries' role in interdependent network security. In *Trust and Trustworthy Computing*, pages 323–336. Springer, 2010.

[56] Theodore Groves and Martin Loeb. Incentives and public inputs. *Journal of Public Economics*, 4(3):211–226, 1975.

[57] Annette Hofmann. Internalizing externalities of loss prevention through insurance monopoly: an analysis of interdependent risks. *The GENEVA Risk and Insurance Review*, 32(1):91–111, 2007.

[58] Jianwei Huang, Randall A. Berry, and Michael L. Honig. Auction-based spectrum sharing. *Mobile Networks and Applications*, 11(3):405–418, 2006.

[59] Leonid Hurwicz. Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. *The Review of Economic Studies*, pages 217–225, 1979.

[60] Cisco Systems Inc. SpamCop Blocking List - SCBL, 2011. `http://www.spamcop.net/`.

[61] Ponemon Institute. Ponemon 2015 Cost of Cyber Crime Studies. `http://www8.hp.com/us/en/software-solutions/ponemon-cyber-security-report/index.html?jumpid=va_fwvpqe387s`. Retrieved on 04-17-2016.

[62] Matthew O. Jackson and Yves Zenou. Games on networks. In *Handbook of Game Theory*, volume 4. Elsevier Science, 2014.

[63] Khyati Jain. These Top 7 Brutal Cyber Attacks Prove 'No One is Immune to Hacking'. `http://thehackernews.com/2015/09/top-cyber-attacks-1.html`. Retrieved on 04-17-2016.

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

[65] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[66] Michihiro Kandori. Introduction to repeated games with private monitoring. *Journal of Economic Theory*, 102(1):1–15, 2002.

[67] Michihiro Kandori and Hitoshi Matsushima. Private observation, communication and collusion. *Econometrica*, pages 627–652, 1998.

[68] Leo Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 1953.

[69] Jay P. Kesan, Rupterto P. Majuca, and William J. Yurcik. The economic case for cyberinsurance. *University of Illinois Law and Economics Working Papers*, 2004.

[70] Howard Kunreuther and Geoffrey Heal. Interdependent security. *Journal of Risk and Uncertainty*, 26(2-3):231–249, 2003.

[71] Aron Laszka, Mark Felegyhazi, and Levente Buttyán. A survey of interdependent security games. *CrySyS*, 2, 2012.

[72] Aron Laszka and Jens Grossklags. Should cyber-insurance providers invest in software security? In *Computer Security–ESORICS*, pages 483–502. Springer, 2015.

[73] Stefan Laube and Rainer Böhme. The economics of mandatory security breach reporting to authorities. In *Workshop on the economics of information security (WEIS)*, 2015.

[74] Marc Lelarge. Economics of malware: Epidemic risks model, network externalities and incentives. In *The 47th Annual Allerton Conference on Communication, Control, and Computing*, pages 1353–1360. IEEE, 2009.

[75] Marc Lelarge and Jean Bolot. Cyber insurance as an incentive for internet security. In *Workshop on the Economics of Information Security (WEIS)*, 2008.

[76] Marc Lelarge and Jean Bolot. Economic incentives to increase security in the internet: The case for insurance. In *IEEE International Conference on Computer Communications (INFOCOM)*, pages 1494–1502. IEEE, 2009.

[77] Yang Liu, Armin Sarabi, Jing Zhang, Parinaz Naghizadeh, Manish Karir, Michael Bailey, and Mingyan Liu. Cloudy with a chance of breach: Forecasting cyber security incidents. In *The 24th USENIX Security Symposium (USENIX Security 15)*, pages 1009–1024, 2015.

[78] Jeffrey K. MacKie-Mason and Hal R. Varian. Pricing congestible network resources. *IEEE Journal on Selected Areas in Communications*, 13(7):1141–1149, 1995.

[79] George J. Mailath and Stephen Morris. Repeated games with almost-public monitoring. *Journal of Economic Theory*, 102(1):189–228, 2002.

[80] George J. Mailath and Larry Samuelson. *Repeated Games and Reputations*. Oxford University Press, 2006.

[81] Mohammad Hossein Manshaei, Quanyan Zhu, Tansu Alpcan, Tamer Bacşar, and Jean-Pierre Hubaux. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3):25, 2013.

[82] Marsh. Benchmarking trends: More companies purchasing cyber insurance, March 2013.

[83] Marsh. Benchmarking trends: As cyber concerns broaden, insurance purchases rise, March 2015.

[84] Andreu Mas-Colell, Michael Dennis Whinston, and Jerry R. Green. *Microeconomic Theory*, volume 1. Oxford university press New York, 1995.

[85] R. Ann Miura-Ko, Benjamin Yolken, John Mitchell, and Nicholas Bambos. Security decision-making among interdependent organizations. In *The 21st IEEE Computer Security Foundations Symposium (CSF'08)*, pages 66–80. IEEE, 2008.

[86] Katta G. Murty. On the number of solutions to the complementarity problem and spanning properties of complementary cones. *Linear Algebra and Its Applications*, 5(1):65–108, 1972.

[87] Katta G. Murty and Feng-Tien Yu. *Linear complementarity, linear and nonlinear programming*. Citeseer, 1988.

[88] Parinaz Naghizadeh and Mingyan Liu. Establishing network reputation via mechanism design. In *Game Theory for Networks*, pages 47–62. Springer, 2012.

[89] Parinaz Naghizadeh and Mingyan Liu. Budget balance or voluntary participation? incentivizing investments in interdependent security games. In *The 52nd Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2014.

[90] Parinaz Naghizadeh and Mingyan Liu. Perceptions and truth: A mechanism design approach to crowd-sourcing reputation. *IEEE/ACM Transactions on Networking (TON)*, 24(1):163–176, 2014.

[91] Parinaz Naghizadeh and Mingyan Liu. Voluntary participation in cyber-insurance markets. In *Workshop on the Economics of Information Security (WEIS)*, 2014.

[92] Parinaz Naghizadeh and Mingyan Liu. Provision of non-excludable public goods on networks: From equilibrium to centrality measures. In *The 53rd Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2015.

[93] Parinaz Naghizadeh and Mingyan Liu. Exit equilibrium: Towards understanding voluntary participation in security games. In *IEEE International Conference on Computer Communications (INFOCOM)*, 2016.

[94] Parinaz Naghizadeh and Mingyan Liu. Inter-temporal incentives in security information sharing agreements. In *AAAI Workshop on Artificial Intelligence for Cyber-Security (Position Paper)*, 2016.

[95] Parinaz Naghizadeh and Mingyan Liu. On the role of public and private assessments in security information sharing agreements. *arXiv preprint arXiv:1604.04871*, 2016.

[96] Parinaz Naghizadeh and Mingyan Liu. Provision of public goods on networks: On existence, uniqueness, and centralities. *arXiv preprint arXiv:1604.08910v2*, 2016.

[97] Parinaz Naghizadeh and Mingyan Liu. Opting out of incentive mechanisms: A study of security as a non-excludable public good. *IEEE Transactions on Information Forensics and Security*, To appear.

[98] Hulisi Ogut, Nirup Menon, and Srinivasan Raghunathan. Cyber insurance and IT security investment: Impact of interdependence risk. In *Workshop on the economics of information security (WEIS)*, 2005.

[99] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. On a way to improve cyber-insurer profits when a security vendor becomes the cyber-insurer. In *IFIP Networking Conference*, pages 1–9. IEEE, 2013.

[100] Ranjan Pal, Leana Golubchik, Konstantinos Psounis, and Pan Hui. Will cyber-insurance improve network security? a market analysis. In *IEEE International Conference on Computer Communications (INFOCOM)*, pages 235–243. IEEE, 2014.

[101] Jong-Shi Pang. Hidden Z-matrices with positive principal minors. *Linear Algebra and Its Applications*, 23:201–215, 1979.

[102] Jee-Hyeong Park. Enforcing international trade agreements with imperfect private monitoring. *The Review of Economic Studies*, 78(3):1102–1134, 2011.

[103] David Christopher Parkes. *Iterative combinatorial auctions: Achieving economic and computational efficiency*. PhD thesis, University of Pennsylvania, 2001.

[104] Paulo Passeri. Hackmageddon. `http://hackmageddon.com`.

[105] PhishTank. `http://www.phishtank.com/`.

[106] Dražen Prelec. A Bayesian truth serum for subjective data. *Science*, 306(5695):462–466, 2004.

[107] Jennifer Price and Tara Javidi. Leveraging downlink for efficient uplink allocation in a single-hop wireless network. *IEEE Transactions on Information Theory*, 53(11):4330–4339, 2007.

[108] Dan Raile. Sony Hack Was Not All That Sophisticated, Cybersecurity Experts Say. `http://www.billboard.com/articles/business/6413955/sony-security-kevin-mitnick-electronic-frontier`. Retrieved on 04-17-2016.

[109] Sasha Romanosky. Comments to the department of commerce on incentives to adopt improved cybersecurity practices, April 2013.

[110] Sasha Romanosky, Rahul Telang, and Alessandro Acquisti. Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2):256–286, 2011.

[111] J. Ben Rosen. Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica*, pages 520–534, 1965.

[112] Tatsuyoshi Saijo and Takehiko Yamato. A voluntary participation game with a non-excludable public good. *Journal of Economic Theory*, 84(2):227–242, 1999.

[113] Paul A. Samuelson. The pure theory of public expenditure. *The Review of Economics and Statistics*, pages 387–389, 1954.

[114] John Schwartz. Who Needs Hackers? `http://www.nytimes.com/2007/09/12/technology/techspecial/12threat.html?ref=technology`, 2007. Retrieved on 04-17-2016.

[115] Shrutivandana Sharma and Demosthenis Teneketzis. A game-theoretic approach to decentralized optimal power allocation for cellular networks. *Telecommunication Systems*, 47(1-2):65–80, 2011.

[116] Nikhil Shetty, Galina Schwartz, Mark Felegyhazi, and Jean Walrand. Competitive cyber-insurance and internet security. In *Economics of Information Security and Privacy*, pages 229–247. Springer, 2010.

[117] Takuo Sugaya. Folk theorem in repeated games with private monitoring. *Economic Theory Center Working Paper*, 2011.

[118] Takuo Sugaya. Folk theorem in repeated games with private monitoring. *Working Paper*, 2013.

[119] The Department of Homeland Security. Enhancing resilience through cyber incident data sharing and analysis. `https://www.dhs.gov/sites/default/files/publications/Data\%20Categories\%20White\%20Paper\%20-\%20508\%20compliant.pdf`, 2015. Retrieved on 02-24-2016.

[120] ThreatTrack Security. Majority of malware analysts aware of data breaches not disclosed by their employers. `http://www.marketwired.com/press-release/majority-of-malware-analysts-aware-of-data-breaches-not-disclosed-by-their-employers-1849009.htm`, 2013. Retrieved on 02-24-2016.

[121] Hal Varian. *Microeconomic Theory*. W. W. Norton & Company, New York, 1992.

[122] Hal Varian. System reliability and free riding. *Economics of Information Security*, pages 1–15, 2004.

[123] Vocabulary for Event Recording and Incident Sharing (VERIS). `http://veriscommunity.net/index.html`.

[124] The web application security consortium's Web-Hacking-Incident-Database. `http://projects.webappsec.org/w/page/13246995/Web-Hacking-Incident-Database`.

[125] Robert L. Winkler, Javier Munoz, José L. Cervera, José M. Bernardo, Gail Blattenberger, Joseph B. Kadane, Dennis V. Lindley, Allan H. Murphy, Robert M. Oliver, and David Ríos-Insua. Scoring rules and the evaluation of probabilities. *Test*, 5(1):1–60, 1996.

[126] Yuanzhang Xiao, Jaeok Park, and Mihaela van der Schaar. Design and analysis of intervention mechanisms in power control games. In *Global Telecommunications Conference (GLOBECOM)*, pages 1–6. IEEE, 2011.

[127] Jing Zhang, Zakir Durumeric, Michael Bailey, Manish Karir, and Mingyan Liu. On the mismanagement and maliciousness of networks. In *The Proceedings of the Network and Distributed System Security Symposium (NDSS '14)*, 2014.