Cyber Risk Information Sharing with Authorities

Stefan Laube

Department of Information Systems
Westfälische Wilhelms-Universität Münster

November 2017

This dissertation is submitted for the degree of
Doctor of Economic and Political Sciences (Dr. rer. pol.)

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften durch die Wirtschaftswissenschaftliche Fakultät der Westfälischen Wilhelms-Universität Münster

Dekanin:	Prof. Dr. Theresia Theurl
Erster Gutachter:	Prof. Dr. Rainer Böhme
Zweite Gutachterin:	Prof. Dr. Theresia Theurl
Tag der mündlichen Prüfung:	14. November 2017
Tag der Promotion:	14. November 2017

To Anne-Christin

Acknowledgements

I'm grateful to my supervisor Rainer Böhme for his support on every step of the way to this dissertation. He sparked my interest in information security when I was a graduate student, taught me how to write in German and English all over again, and inspired me in manifold ways. His insights and comments always were enormously valuable, and helped me to improve the quality of my works. Rainer has been a great mentor. I think that his style of mentoring had a very positive impact on my character.

I also thank the rest of the former IT Security Research Group in Münster – now Security and Privacy Lab in Innsbruck – for their fruitful discussions. Even when all of them moved to Innsbruck, they made me feel part of the team and supported me in writing this dissertation at any time possible. In particular, I thank Markus Riek for inspiring research talks, and Pascal Schöttle for conversations on game-theoretic problems. Additionally, I'd like to thank Paulina Pesch, who teamed up with me in Münster, for providing her valuable insights into German law. Besides, I thank all members of the Practical Computer Science Group in Münster, led by Herbert Kuchen, who helped Paulina and me in many ways after our research group left to Innsbruck.

I am grateful to Terrence August, Duy Dao, and Florin Niculescu, for their cooperation and mentorship when I learned about new approaches to solve game-theoretic models during my four month stay at the University of California, San Diego, in 2017.

I thank all of my friends for their support during the time of writing this dissertation. Special thanks go to Martin Gurzinski and Yves Passlack for the interesting debates related to this work. And I'd like to thank Dominik Hamachers for listening to my research approaches, and his industry insights on cyber risk management in response.

Above all I'm grateful for the support of my family, my parents Karin and Karl-Heinz Laube, my sister Janina Laube, and my fiancée Anne-Christin Grewe. Everything that I have accomplished in my life is due to their efforts. They always encouraged me to do the things that I really liked, regardless of the difficulties that they involved. Without their support and encouragement during my studies, this dissertation could not have happened. I love you all.

Summary

Cyber risk management largely reduces to a race for information between attackers and defenders of ICT systems. Defenders can gain advantage in this race by sharing cyber risk information with each other. Yet, defenders often exchange less information than is socially desirable, as their decisions are guided by selfish reasons. This can motivate regulators to enact laws mandating defenders' information exchange. In particular in Europe, many laws oblige defenders' information sharing with authorities, who in turn can advise others to strengthen the overall defense in the economy. This dissertation sheds first light into the economics of cyber risk information sharing with authorities.

This work provides two main contributions. First, we systematize knowledge on the economics of defenders' cyber risk information sharing in a novel framework. This survey aims to contribute a consolidated understanding of defenders' incentives to exchange information privately or with the public. Second, our survey motivates us to investigate the economics of firms' mandatory ICT system breach information sharing with authorities based on two game-theoretic models. The model analyses aim to explain effects of regulators' effective law enforcement on welfare and affected firms.

The first model is used to identify conditions under which effectively enforcing laws that mandate firms' breach information sharing with authorities improves welfare. It assumes that firms have difficulties to detect breaches, and few incentives to unilaterally report them once laws are enacted. Regulators can effectively enforce laws by initiating audits at firms to detect and sanction non-reporting. Yet, audits cannot differentiate between nescience and breach concealment. Thus, the model analysis predicts that even under optimistic assumptions on authorities' success in advising others, it is difficult to adjust expected sanctions such that an effective enforcement of laws improves welfare.

The second model is used to inquire effects of effectively enforcing laws that mandate breach information sharing with authorities on affected firms' incentives to invest in breach prevention and detection. It assumes that audits and sanctions introduced by regulators work (with varying effectiveness), and influence firms' investment decisions. The model analysis suggests that enforced laws incentivize firms to under-invest in breach prevention and over-invest in detection. And if regulators increase expected sanctions, affected firms shift their investment priority from preventive to detective controls. This practice can result in resource allocations that are socially detrimental.

Keywords: Cyber risk management; information sharing; policy; game-theory.

Zusammenfassung

Cyber-Risikomanagement ist durch einen Wettlauf um sicherheitsrelevante Informationen zwischen Angreifern und Verteidigern von IKT-Systemen geprägt. Verteidiger können sich dabei einen Vorteil verschaffen, indem sie solche Informationen austauschen. Sie tauschen jedoch oft weniger Informationen aus, als es gesellschaftlich wünschenswert wäre, weil ihre Entscheidungen von egoistischen Motiven getrieben werden. Dies kann Staaten dazu motivieren, Informationsaustausch gesetzlich vorzuschreiben. Insbesondere in Europa verpflichten viele Gesetze Verteidiger zur Meldung von Informationen an Behörden, die wiederum andere Verteidiger beraten und dadurch die Sicherheit von IKT-Systemen in der Gesellschaft erhöhen können. Diese Dissertation behandelt ökonomische Aspekte des Austauschs sicherheitsrelevanter Informationen mit Behörden.

Die Arbeit liefert zwei wesentliche Beiträge. Erstens systematisiert sie Forschungsergebnisse zum Austausch sicherheitsrelevanter Informationen von Verteidigern in einem neuen Rahmenwerk. Dies führt zu einem Verständnis der Anreize von Verteidigern, solche Informationen privat oder öffentlich preiszugeben. Zweitens, motiviert aus der Systematisierung, enthält die Arbeit eine ökonomische Untersuchung der Auswirkungen von Gesetzen, die Firmen dazu verpflichten, ihre privaten Informationen über Sicherheitsvorfälle an Behörden zu melden. Die Analyse von zwei spieltheoretischen Modellen gibt Aufschluss über Effekte solcher Gesetze auf die Wohlfahrt und verpflichtete Firmen.

Mit Hilfe des ersten Modells werden Bedingungen identifiziert, unter denen eine wirksame Durchsetzung der Gesetze die Wohlfahrt steigert. Es wird angenommen, dass Firmen Schwierigkeiten haben, Vorfälle zu entdecken, und wenig Anreize, sie unilateral zu melden. Dem kann der Staat entgegenwirken, indem er IKT-Systeme der Firmen auditieren lässt und nicht gemeldete Vorfälle sanktioniert. Audits können jedoch nicht differenzieren, ob Firmen Vorfälle nicht entdeckt oder verschwiegen haben. Darum zeigt die Modellanalyse, dass selbst unter sehr optimistischen Annahmen zur Informationsverwertung der Behörden eine wirksame Durchsetzung der Gesetze oft nicht sinnvoll ist.

Mit Hilfe des zweiten Modells wird untersucht, wie die Gesetze die Anreize der Firmen verändern, in Vorfall-Prävention und Vorfall-Detektion zu investieren. Es wird angenommen, dass realisierte Audits und Sanktionen funktionieren (mit unterschiedlicher Effektivität) und die Investitionsentscheidungen der Firmen beeinflussen. Die Modellanalyse zeigt, dass von Gesetzen betroffene Firmen aus gesellschaftlicher Perspektive zu wenig in Vorfall-Prävention und zu viel in Vorfall-Detektion investieren.

Stichworte: Cyber-Risikomanagement; Informationsaustausch; Politik; Spieltheorie.

Table of contents

List of figures

List of tables

Chapter 1

Introduction

The security of information and communication technology (ICT) systems that are interconnected by physical links (e. g., wires), logical links (e. g., same vulnerabilities), or social links (e. g., trust relationships), does not only depend on the actions of single defenders, but also on actions of others. This *interdependence* substantially influences the efficiency of defenders' actions to secure ICT systems [Kunreuther and Heal, 2003].

The endeavor of securing interconnected ICT systems is often characterized as a *race for information* [Ransbotham et al., 2012]. This race usually starts with the discovery of a vulnerability that is simultaneously present in ICT systems of many defenders, e. g., a programming error in software. If an attacker discovers the vulnerability first, he can develop an exploit to attack unprotected systems, often remotely over networks. This creates *cyber risk* for defenders with affected systems. And defenders' expected cyber risk increase if the attacker leverages interconnections of breached systems to let his attacks propagate and affect others. Thus, in economic jargon, unprotected ICT systems can generate *negative externalities* affecting whole networks. Yet, if a defender discovers the vulnerability first, he may be able to effectively manage his cyber risk, e. g., by reconfiguring software to shield a programming error. And in case that his efforts are effective, they decrease expected cyber risk of other defenders with interconnected systems. Such spillover effects of effective cyber risk management efforts are referred to as *positive externalities*. Overall, externalities make the defense of ICT systems a joint effort. Evidently, defenders can exchange information with each other that may support their cyber risk management efforts. Real-world examples are: firms' exchange of risk prevention solutions in industry-based sharing centers [National Concil of ISACs, 2017]; firms' release of patches for vulnerabilities exploitable by the *Code Red worm* [Zou et al., 2002]; and defenders' public disclosure of information on the *heartbleed* vulnerability [Durumeric et al., 2014]. We call such actions *cyber risk information sharing*.

1.1 Motivation

Despite the potential benefits of defenders' cyber risk information sharing, they often share less than is socially desirable. Many scholars argue that the main obstacle is economics rather than technology [Anderson and Moore, 2006]: defenders lack incentives to exchange information. If a lack of incentives to share information jeopardizes the protection of ICT systems, thus generating negative externalities, this can justify government intervention in form of laws mandating defenders' information exchange.

One approach by regulators is to enact laws that mandate firms' breach information sharing with authorities, such as the European Union (EU) Directive 2016/1148. The objective is twofold. First, the laws aim to let affected firms internalize negative externalities from their ICT systems. Incentives for firms to prevent breaches may be created because this decreases the amount of reporting obligations along with disclosure costs, e. g., reputation damages due to authorities' release of reported breach information to the public. Also, incentives for firms to detect breaches may be created because this enables to comply with reporting obligations. Second, the laws aim to let informed authorities establish an economy-wide transparency on breaches. For instance, authorities can use received information to draw and share conclusions with others, helping them protect from propagating attacks. Yet, firms may not unilaterally report breaches in the first place due to expected disclosure costs. The previously mentioned EU directive suggests that regulators can counter such disincentives by initiating audits at firms to detect and sanction non-compliance with reporting obligations. But this harms firms which do not report breaches as they cannot detect them. Consequently, welfare not necessarily improves by an effective enforcement of breach notification laws.

In this work, we try to achieve three goals. First, we aim to provide an understanding for defenders' information sharing incentives. Second, we mean to use this understanding to identify conditions under which an effective enforcement of laws that mandate firms' breach information sharing with authorities improves welfare. Third, we intend to evaluate effects of such laws on firms' incentives to prevent and detect breaches.

1.2 Classification and scope

This work contributes to the growing stream of literature in the *security economics* domain that studies cyber risk information sharing. In the 1980s, scholars started out to inquire the economics of information sharing in non-security related domains, e. g., firms' engagement in trade associations [Kirby, 1988] and research joint ventures [Kamien

et al., 1992]. This literature was resurrected in the 2000s, when *security economics* emerged as a new research field [Anderson, 2001]. Today, hundreds of studies investigate the economics of cyber risk information sharing between defenders or attackers.

We narrow the scope of literature relevant to this dissertation on works investigating the economics of institutionalized cyber risk information sharing by benign defenders. *Institutionalized sharing* means that the information sharing activities of interest are socially recognized, formalizable in principle, and have a measurable effect on economy-wide cyber risk. The focus on works inquiring institutionalized sharing ensures that all distilled contributions are relevant to scholars and practitioners with the objective to investigate or develop impactful new cyber risk information sharing incentive mechanism. For instance, it excludes literature on defenders' ad hoc sharing, such as in purely informal exchanges or personal blog posts. In this dissertation, *defenders are benign*, protecting their ICT systems against intentional malice by attackers. Thus, we concentrate on information sharing instances that are potentially in the societal interest. For example, this focus excludes studies where defenders are allowed to share misinformation, and literature on information exchange of attackers to increase expected rewards from undermining the security of ICT systems [Hausken, 2015, 2017].

1.3 Prerequisites

We use *game theory* to theoretically study the planned and purposeful behavior of defenders, affected by exogenous threats due to actions of attackers. Specifically, we construct game-theoretic models that capture the conflicting interests between defenders, to be comprehended as players in a competitive game. The (anticipated) interests of these players determines their *strategies* to share cyber risk information, invest in security, or enforce laws. Adopting a game-theoretic terminology, a strategy refers to a player's plan of actions with the objective to maximize his own utility while taking into account or anticipating the actions of others. For readers not familiar with this concept and without a basic game-theoretic background, we recommend the following two textbooks (which are increasing in their mathematical sophistication) that provide extensive methodological introductions from an economic angle:

- Osborne, M. J. (2003). *An introduction to game theory.* Oxford University Press, Oxford
- von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior.* Princeton University Press

Specific contents presented in the above textbooks are particularly relevant to this dissertation: we solve our presented models for *pure* and *mixed* strategy Nash equilibria; moreover, our model analyses are limited to *one-shot games* with *fixed states*.

Furthermore, this dissertation is written for readers with a basic background on the *economics of information goods* and *information security*. We aim to introduce or reference the most relevant concepts of these fields of study as they arise in this work. Complementary, we recommend the following two textbooks each providing an in-depth introduction to one of the fields from a computer science perspective:

- Shapiro, C. and Varian, H. R. (1998). *Information rules: A strategic guide to the network economy*. Harvard Business Review Press
- Anderson, R. (2008). *Security engineering: A guide to building dependable distributed systems*. Wiley, 2 edition

Shapiro and Varian [1998] introduce *economics* of information goods, which are relevant to this dissertation as cyber risk information can be considered such goods. In particular, an understanding of these economics is required to comprehend the evolution of information markets, e. g., the markets for software vulnerabilities [Zhao et al., 2015], and how these markets are different from traditional ones. Differences of information markets can be traced back to the characteristics of information goods, e. g., negligible costs of reproduction, which promote for instance first mover advantages and the exploitation of network externalities. The competition between firms on information markets partly explains why nowadays the security level of ICT systems may often be suboptimal from a social welfare point of view, as revisited later in this work.

Anderson [2008] provides an understanding for the functioning of *technologies* that defenders may use to secure their ICT systems, which is required to comprehend some of the rather technical arguments in this dissertation. In particular, insight into these technologies help to recognize why systems may fail, and which cyber risk management strategies can protect from such failures. In general, many ICT systems appear to fail because they are used wrong, rather than due to weak underlying security mechanisms. For that reason, cyber risk associated with ICT systems can often be explained by economic principles. This highlights that interdisciplinary research approaches involving economic and technical knowledge as well as skills are much needed to gain insights on how to effectively and efficiently improve the security of interconnected ICT systems.

1.4 Dissertation outline

This dissertation is divided into three parts: a *systematization of knowledge* on the economics of defenders' cyber risk information sharing (Chapters 2–5), a derivation of *new results* in this context (Chapters 6–7), and a *summary* of our findings (Chapter 8).

After this introduction, in Chapter 2, we propose a *framework* that sets the stage to study the economics of cyber risk information sharing. In this framework, we define actors that may have information to share, information types, and a timing model to capture the race for information aspect that is specific to cyber risk. Also, we present a theoretical model that provides the formal basis for the rest of this dissertation.

In Chapters 3 and 4, we review literature that studies the economics of defenders' *voluntary private and public cyber risk information sharing*, respectively. Specifically, we review assumptions made in theoretical works and their implications on information sharing strategies. We contrast this with effects of information sharing reported in empirical studies, allowing us to make statements about the validity of theoretical models and implied strategies. By integrating all works into our framework, we aim to provide a consolidated understanding of defenders' incentives to voluntary share information.

In Chapter 5, we review literature inquiring the economics of defenders' *mandatory cyber risk information sharing*. We first provide a context for laws that mandate information sharing with authorities or individuals. Thereafter, we integrate works investigating these laws into our framework. This calls attention to a lack of studies on the economics of defenders' cyber risk information sharing with authorities, motivating us to make a first step towards filling this research gap within the followup chapters.

In Chapter 6, we *identify conditions under which effectively enforcing laws that mandate firms' breach information sharing with authorities improves welfare*. We first characterize effects of regulators' enforcement practices on firms' incentives to invest in breach prevention and to share breach information with authorities. Then, we present a model capturing these effects as a game between a regulator and firms. We examine the regulator's equilibrium enforcement practices and resulting investment and information sharing strategies of firms, with respect to exogenous actions of the involved authority. This enables us to determine how an effective enforcement of laws changes welfare.

In Chapter 7, we inquire the *effects of effectively enforcing laws that mandate breach information sharing with authorities on affected firms' incentives to invest in breach prevention and detection*. Therefore, we come up with a variant of the model presented in the previous chapter, allowing firms to invest limited budget in productive activity, preventive and detective security controls. We examine firms' equilibrium investment

strategies, with respect to exogenous enforcement practices of the regulator and actions of the involved authority. This enables us to determine how an effective enforcement of laws changes firms' security spendings, investment priority, and their overall profit.

Finally, Chapter 8 *concludes* this dissertation by providing a summary of our results and an outlook on cyber risk information sharing with authorities in the future. The outlook highlights some possible policy, managerial, and research implications.

1.5 Publications and collaboration

Almost all texts and concepts presented in this dissertation originate from my previous research publications, prepared and published in collaboration with Rainer Böhme.

The Chapters 2, 3, 4 and parts of Chapter 5 are based on a survey article, prepared in collaboration with Rainer Böhme who helped to structure the field:

- Laube, S. and Böhme, R. (2017). Strategic aspects of cyber risk information sharing. *ACM Computing Surveys*, 50(5) (https://doi.org/10.1145/3124398)

The idea for the cascade model of cyber risk arrival in this dissertation (cf. Figure 2.1) resulted from some brainstorming together with Rainer Böhme and Markus Riek for preparation of the following article:

- Böhme, R., Laube, S., and Riek, M. (2018). A fundamental approach to cyber risk analysis. *Variance*, 11(2)

Most of the contents presented in Chapters 5 and 6 were published in collaboration with Rainer Böhme, who had the idea that the conflict between a regulator who enforces a law and affected firms can be interpreted as a principal-agent setup:

- Laube, S. and Böhme, R. (2016). The economics of mandatory security breach reporting to authorities. *Journal of Cybersecurity*, 2(1):29–41 (https://doi.org/10.1093/cybsec/tyw002)

The above journal article is based on a preceding conference paper and a slightly extended workshop paper, both also prepared together with Rainer Böhme:

- Laube, S. and Böhme, R. (2015b). Meldepflichten für IT-Sicherheitsvorfälle: Ein Prinzipal-Agent-Ansatz. In Thomas, O. and Teuteberg, F., editors, *Tagungsband Wirtschaftsinformatik*, pages 1146–1162, Osnabrück, Germany
  (This publication was nominee for the best paper award at the "Wirtschaftsinformatik" conference in Osnabrück, Germany, 2015.)

- Laube, S. and Böhme, R. (2015c). The economics of mandatory security breach reporting to authorities. In *Workshop on the Economics of Information Security (WEIS)*, Delft University of Technology, The Netherlands

Going further back in time, I want to acknowledge that some of the ideas presented in the two papers listed above have initially been qualitatively introduced in my very first joint work together with Rainer Böhme:

- Böhme, R. and Laube, S. (2014). Das IT-Sicherheitsgesetz. In Baetge, J. and Kirsch, H.-J., editors, *Mittelstand im Blick: Compliance und Risikomanagement*, pages 17–36, Düsseldorf, Germany. IDW

The content introduced in Chapter 7 was also published in collaboration with Rainer Böhme, who encouraged me to pursue the idea for the presented model during a skiing seminar of the Department of Information Systems in 2015:

- Laube, S. and Böhme, R. (2015a). Mandatory security information sharing with authorities: Implications on investments in internal controls. In *ACM Conference on Computer and Communication Security (ACM CCS), Workshop on Information Sharing and Collaborative Security (WISCS)*, Denver, CO, USA (https://doi.org/10.1145/2808128.2808132; This publication was awarded the "Science of Risk Prize 2015" by Lloyd's of London in the "Cyber Risk" category.[1])

Besides, I have contributed to empirical work that resulted from my co-supervision of a bachelor-level seminar, which does not fit conceptually into this dissertation (further publications in the same context are planned):

- Machuletz, D., Sendt, H., Laube, S., and Böhme, R. (2016). Users protect their privacy if they can: Determinants of webcam covering behavior. In *Proceedings of the 1st European Workshop on Usable Security (EuroUSEC)*, Darmstadt, Germany

Additionally, during my semester abroad at the University of California, San Diego, I contributed to a working paper on the economics of ransomware which is not part of this dissertation (a journal version of this paper is planned):

- August, T., Dao, D., Laube, S., and Niculescu, F. (2017). Economics of ransomware attacks. In *Workshop on Information Systems and Economics (WISE)*, Seoul, South Korea

[1]For further information on this award, see the website of Lloyd's of London (https://www.lloyds.com/news-and-insight/news-and-features/lloyds-news/2015/11/from-the-physical-to-the-intangible).

Part I

Systematization of knowledge

Chapter 2

Framework

Cyber risk information sharing is an act of communication. Thus, a natural approach is to start with Lasswell's popular general model of communication [Lasswell, 1948]

"*Who* says *What* in *Which Channel* to *Whom* with *What Effect*?"

We use different components from this statement as dimensions for our framework, leveraged to review the effects of cyber risk information sharing (*What Effect*) in Chapters 3, 4, and 5. In Section 2.1, we introduce different actors (*Who* and *Whom*) that may have cyber risk information to share. These information can be of diverse types (*What*), as discussed in Section 2.2. The sharing channel (*Which Channel*) does not merit a dimension of its own, but is discussed where relevant within this dissertation. Although not explicitly mentioned in Lasswell's model, our framework considers information sharing at different points in time, as explicated in Section 2.3. We synthesize a formal model that summarizes modeling assumptions of many theoretical works on cyber risk information sharing in Section 2.4, capturing defenders' strategy space for relevant single points in time.

2.1 Actors

We broadly differentiate between two types of *actors*: attackers and defenders. *Attackers* try to strategically undermine the security of ICT systems for own economic advantages. *Defenders* try to strategically manage the cyber risk associated with their ICT systems due to attackers' actions. Defenders can be categorized in regulators, individuals, and firms. *Regulators* take measures to optimize the security of ICT systems in economies. *Individuals* are a part of ICT systems, e. g., people that use computers during their

work at firms. And *firms* are professionally managed organizations in the private or public sector that can broadly be classified into vendors, consumers, and intermediaries.

Vendors are firms belonging to the ICT industry. They develop and sell ICT products (e. g., software) or ICT services (e. g., cloud services) to generate profit. Vendors compete on markets and have to keep developing their products or services to stay competitive. Thereby, their decisions on product or service security are governed by "information rules" [Shapiro and Varian, 1998]. Following these economics of information goods, vendors trade-off investments in functional features against security. This may lead to an under-provision of security in economies [Anderson et al., 2008], enabling the formation of a security industry that sells security products to consumers.

Consumers are firms belonging to any industry. They compete on markets in their specific sector of business and require products and services of vendors to set up and operate ICT systems, which may help to stay competitive. Thereby, consumers cannot strategically differentiate themselves in the use of ICT systems [Carr, 2003]. Their cyber risk is to a large extent determined by vendors' security related decisions.

Intermediaries are firms or government authorities that moderate information sharing between all other actors on a non-profit or for-profit basis. Non-profit based intermediaries gather and share information to enhance welfare. By contrast, profit-seeking intermediaries share cyber risk information for own economic advantages.

2.2 Cyber risk information

We use a simple notion of cyber risk, which is compatible with the popular National Institute of Standards and Technology (NIST)-framework [NIST, 2012], to derive information of different types that may support defenders' cyber risk management:

$$\text{Cyber risk} = \text{Breach probability}(\textit{Attacks}, \textit{Controls}) \times \textit{Impact} \ . \tag{2.1}$$

The factors in italic map into the cascade model of cyber risk arrival depicted in Figure 2.1, refined from the NIST-framework. This model induces our classification of information types. Attacks are realizations of threats exploiting vulnerabilities in ICT systems. *Attack information*, presented in Section 2.2.1, help defenders to *identify cyber risk*. Controls can prevent or detect attacks [Cavusoglu et al., 2004a]. And *control information*, introduced in Section 2.2.2, support the *treatment of cyber risk*. Impact results from controls' failure, such that attackers can compromise assets, causing losses. *Impact information* help defenders to *assess cyber risk*, as discussed in Section 2.2.3.

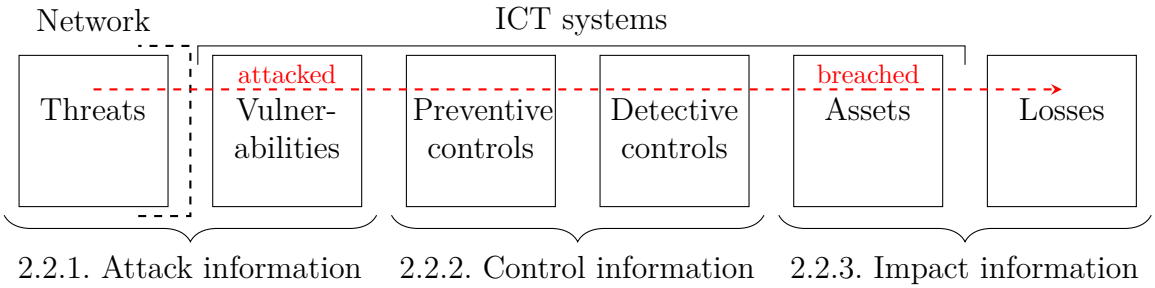


Fig. 2.1 Cascade model of cyber risk arrival.

2.2.1 Attack information

Defenders can identify cyber risk by collecting *information on the threats to their ICT systems*. Such threats can be assessed with the help of various frameworks [Caralli et al., 2007; NIST, 2012] or by modeling techniques, e. g., attack trees [Schneier, 2000]. Threats are characterized by the resources that attackers require for their realization. The threat realizations of attackers aim at vulnerabilities, which can be comprehended as master keys to affected ICT systems. Threats that get realized on vulnerabilities alter the state of ICT systems to *attacked*, as annotated in Figure 2.1.

Vulnerabilities by themselves also indicate cyber risk. Landwehr et al. [1994] offer a taxonomy for vulnerabilities, and Ozment [2007] coins the notion of a vulnerability life cycle. In this work, an undiscovered vulnerability is considered unknown. The discovery of such a vulnerability requires investment in security research and luck [Khouzani et al., 2014b]. Once some actor discovers an unknown vulnerability, it is designated secret. If this actor discloses the vulnerability to selected others, it is referred to as private. A vulnerability becomes public if it is published, e. g., via advisories or the media; or indirectly, by a vendor's release of a patch that gets reverse engineered.

The exploitation of vulnerabilities by attackers allows defenders to collect real-time attack information. This information includes, for instance, the *source of offending network packets*, identifiers of *abusive websites*, or *attack patterns*. Attack patterns can provide further insights that allow defenders to distinguish between attack types [Collins et al., 2006]: strategic attacks are targeted, characterized by attackers who exert effort to adjust their attack methods to specific ICT systems (e. g., industrial espionage); opportunistic attacks are untargeted, characterized by attackers using standardized methods until weak ICT systems are found (e. g., spam e-mails). Defenders can make use of security controls in an attempt to influence the success of attacks.

2.2.2 Control information

In order to successfully treat cyber risk, defenders require *information on effective and efficient cyber risk treatment strategies*. The canonical risk treatment strategies are: risk mitigation by investment in preventive or detective security controls, risk avoidance by abstaining from the use of exposed ICT systems altogether – thereby forgoing profit –, risk transfer to third-parties, and risk acceptance. Information on how to effectively invest in controls are of particular interest to defenders. Know-how on investment in preventive controls (e. g., packet filters, patches for vulnerabilities, or staff training to facilitate secure behavior) can help to build up a defensive shield around ICT systems. By contrast, corresponding information on detective controls (e. g., intrusion detection systems (IDS), or security audits) may support the monitoring of ICT systems for attacks that overcome preventive measures. We subsume *structured information required to run security controls effectively* under the term control information.

The effectiveness of preventive controls in mitigating cyber risk is depending on different types of information. For instance, many packet filters use *up-to-date blacklists* of attack source addresses to identify and block malicious traffic. Furthermore, patches provided by vendors, and fixes such as advisories or workarounds, come with inherent *rules that adapt ICT systems' configuration* to close or shield vulnerabilities.

Similarly, information that determine the effectiveness of detective controls are diverse. For instance, signature-based IDS require *indicators of compromise (IOCs)*, e. g., vulnerability, malware, or virus signatures, to detect malicious activity. By contrast, anomaly-based IDS compare activities on ICT systems to *normal activity profiles* and interpret every deviation as malicious activity. The performance of IDS is typically measured by their type I and type II error probabilities [Öğüt et al., 2008].

Finally, control information includes *procedures minimizing the impact of attacks*, which can only be initiated once attacks get detected. Such information may be recorded during the response to impactful attacks [Freiling and Schwittay, 2007].

2.2.3 Impact information

The *identity and value of assets that may get compromised* by attackers can help defenders to assess expected losses associated with attacks. Scholars traditionally differentiate between physical and information assets. Physical assets are tangible, e. g., hardware of ICT systems, machines, and facilities. The identification and valuation of these assets is generally considered feasible. By contrast, information assets are intangible, e. g., proprietary software, source code, or personal customer data. Their identification [NIST,

2011] and valuation [Moody and Walsh, 1999] is considered to be a hard task. If an attacker violates one or more of tangible or intangible assets' canonical protection goals (confidentiality, integrity, and availability), a breach occurs, as depicted in Figure 2.1. *Breaches* of ICT systems (also referred to as *successful attacks*) may be classified by the protection goals that they violate. Furthermore, they can be distinguished by the types of compromised assets. Scholars speak of privacy breaches if the protection of individuals' personal data fails, and of security breaches otherwise [Fischer-Hübner, 2001].

If a breach occurs, the associated *losses* that an affected defender has to assess are usually multifaceted [Romanosky, 2016] and can broadly be classified into primary and secondary [Bandyopadhyay et al., 2009]. Primary losses result from first-degree effects of the breach. An example is the destruction of a firm's valuable information assets that cannot be restored, which causes productivity losses. Secondary losses constitute second-degree effects of the same event. For instance, information regarding assets' destruction may get public, resulting in losses due to liability claims, or lost reputation and thus decreases in demand. As secondary losses arise if a breach becomes public, they may be influenced by defenders' timing of corresponding cyber risk information disclosures.

2.3 Timing model

Cyber risk information always starts out as *secret* to some actor, may then be *privately* shared, but are likely to become *public* eventually. This defines the time horizon regarded in this dissertation, and is justified by three important characteristics of information goods [Shapiro and Varian, 1998]. First, many types of cyber risk information are non-excludable in the long run. Consider attack information such as about an unknown vulnerability, which can theoretically be detected by everyone with access to vulnerable ICT systems. Second, the consumption of cyber risk information is non-rivalrous. Returning to the previous example, everyone who gets to know that a vulnerability exists can check whether his own or the ICT systems of others are also affected. Third, cyber risk information can be shared with others at negligible costs. For instance, the actor who discovers a vulnerability first can use the network to privately or publicly share corresponding information without significant effort.

A defender in possession of *secret* cyber risk information may take the strategic decision to share them with selected others. This starts the timeline depicted in Figure 2.2 and is referred to as *private* information sharing. However, before such sharing, the defender may want to make sure that the counterparties keep received information private for some time by vetting their trustworthiness. For instance,

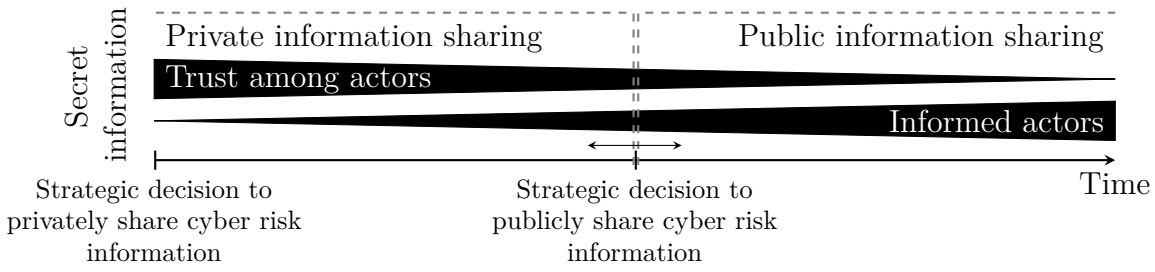


Fig. 2.2 Transformation from secret, to private, to public cyber risk information.

ex ante background checks can help to determine others' tendency to leak cyber risk information. And ex post leakage may be avoided by tailored non-disclosure agreements. In general, such vetting defines the group of recipients with whom cyber risk information get privately shared. However, after information are privately shared, the trust among informed actors may decrease over time. The reason is that the number of informed actors typically increases over time, justified for example by staff turnover or the accumulation of small leaks. In turn, the likelihood that some actor decides to intentionally publish information for own economic advantages elevates.

Informed actors can take the strategic decision to publish their secret or private cyber risk information at any point in time, referred to as *public* information sharing. With respect to Figure 2.2, the disclosure of secret information does not have a preceding timeframe where they are privately shared among actors, unlike the announcement of private information. After information get public, the trust among informed actors repeals: a disclosure not only informs all trustworthy defenders, but also attackers.

In this dissertation, we distinct between works that investigate the economics of defenders' private and those inquiring public cyber risk information sharing. We have introduced the above timing model to argue that this distinction requires a defined time horizon: in the long run, all sharing is public as cyber risk information are likely to become public eventually. This addresses the race for information aspect specific to cyber risk. With respect to this race, the timing of cyber risk information secrecy, private or public sharing is part of defenders' strategy space. The formal model presented next refines this strategy space for the relevant single point in time.

2.4 Unified formal model

All theoretical works on cyber risk information sharing presented in this dissertation interpret the conflict between defenders and attackers as a game. Thereby, most authors – just as we – do not study defenders' information sharing in isolation, but combine it

Table 2.1 Symbols: baseline for this dissertation.

Symbol	Description	Type	Constraint
x	security investment	decision variable	$x \geq 0$
s	information sharing	decision variable	$s \in [0, 1]$
p	price of ICT products	decision variable	$p \geq 0$
y	quality of shared information	product	$y = x \cdot s$
r	attack magnitude	parameter	$r \geq 0$
γ	interdependence	parameter / function	$\gamma \in [-1, 1]$
G	profit	function	
P	breach probability	function	
R	cyber risk	function	
D	demand for ICT products	function	
I	impact of breaches	function / constant	
C	costs	function / constant	

with their decision to invest in security. Yet, the existing literature is very diverse in terms of assumptions, modeling approaches, and notations. To discuss a large set of theoretical works in a unified manner, and to provide a formal basis for the rest of this dissertation, we define a parsimonious model that does not omit properties which are required to investigate defenders' strategies. Table 2.1 summarizes the notation used.

Our model assumes an economy with two defenders $i \in \{0, 1\}$ and one attacker, modeled as an exogenous threat. Each defender is rational, risk-neutral, and acts selfish but not malicious. Furthermore, he makes use of an ICT system for his own economic advantage, which is affected by attacks that the attacker realizes with some resources $r \geq 0$. The probability for a defender's ICT system to experience a breach is given by P. This probability serves as a proxy for the defender's ICT system security level, which is not directly observable. Breaches cause some impact I to the defender.

Each defender can invest in controls $x \geq 0$ to reduce the breach probability, but such investment yields decreasing marginal returns [Gordon and Loeb, 2002]. Specifically, an increase in investment decreases the probability of breaches at a defender $\partial P/\partial x < 0$, but at a decreasing rate $\partial^2 P/\partial x^2 > 0$, such that $\lim_{x\to\infty} P = 0$. Furthermore, each defender's investment in controls has increasing marginal costs C, i. e., $\partial C/\partial x > 0$.

Both defenders may also share some portion of their cyber risk information $s \in [0, 1]$ directly or via a moderator to support each other. This influences their breach probability P, the associated impact I, or investment costs C. Many scholars contingent the quality of shared information on a defender's security investment efforts $y = x \cdot s$: if a defender does not control his cyber risk, he cannot share valuable information. In general, information sharing has positive and negative effects, revisited in the next chapters.

Additionally, defenders have interdependent $\gamma \in [-1, 1]$ payoffs. Following Laszka et al. [2014], we parametrize a defender's payoff with his own and the other's decisions if such interdependence is considered. Thereby, payoff can be generate by two channels:

- *Reduced cyber risk.* Most theoretical works assume defenders to generate payoff by reducing their expected cyber risk. According to Equation (2.1), cyber risk R of a defender is defined as the product of his breach probability P and the impact I. Thus, if the two defenders are symmetric and interdependent, their objective is:

$$\arg\min_{x_i, s_i} \; R_i(r_i, r_{1-i}, x_i, x_{1-i}, s_i, s_{1-i}) + C_i(x_i, x_{1-i}, s_i, s_{1-i}) \text{ , with} \tag{2.2}$$

$$R_i(r_i, r_{1-i}, x_i, x_{1-i}, s_i, s_{1-i}) = P_i(r_i, r_{1-i}, x_i, x_{1-i}, s_i, s_{1-i}) \cdot I_i(s_i, s_{1-i}) \text{ .} \tag{2.3}$$

- *Increased profit.* Gal-Or and Ghose [2005] assume that defenders' decisions generate payoff by increasing expected profit. The authors consider two defenders in their model, acting as ICT product vendors that compete on the market. These vendors' profit G depends on product prices p, influencing the demand D. Demand also depends on vendors' decisions regarding product security, taking a form of quality. Consumers are aware of and consider this quality in purchasing decisions. Thus, symmetric and interdependent vendors' objective function is:

$$\arg\max_{x_i, s_i, p_i} \; G_i(x_i, x_{1-i}, s_i, s_{1-i}, p_i, p_{1-i}) - C_i(x_i, x_{1-i}, s_i, s_{1-i}) \text{ , with} \tag{2.4}$$

$$G_i(x_i, x_{1-i}, s_i, s_{1-i}, p_i, p_{1-i}) = p_i \cdot D_i(x_i, x_{1-i}, s_i, s_{1-i}, p_i, p_{1-i}) \text{ .} \tag{2.5}$$

The simultaneous strategic decisions – or strategies – of defenders in this unified model can be derived from their objective functions in Equation (2.2) and (2.4). The solution concept used in all theoretical studies presented in this dissertation is the Nash equilibrium [Nash, 1950]. Nash equilibria constitute mutual best responses: no defender can gain an economic advantage by unilaterally deviating from his decisions. The sum of all defenders' objective functions is commonly referred to as *social welfare*.

Defenders' strategies can be benchmarked by introducing an imaginary *social planner* to the model. This planner is assumed to coordinate some or all of the defenders' decisions to optimize social welfare. Therefore, the level of social welfare generated by a planner is referred to as *social optimum*. A socially desirable state is reached if defenders' own strategies lead to social welfare equal to the social optimum.

This completes the definition of our unified formal model. Table 2.2 classifies all subsequently investigated theoretical works based on their key model characteristics.

Table 2.2 Mapping of theoretical works inquired in this dissertation and key model characteristics. Works indicated by an asterisk (*) comprise models that capture our assumptions and can thus be represented as instances of the unified model.

Authors	objective function		captured decision variable		sharing between defenders is	
	cyber risk	profit	investment	sharing	direct	moderated
Gordon et al. [2003]*	✓		✓	✓		✓
Gal-Or and Ghose [2005]*		✓	✓	✓		✓
Kannan and Telang [2005]		✓	✓			✓
Öğüt et al. [2005]*	✓		✓			✓
Arora et al. [2006a]		✓	✓	✓	✓	
Cavusoglu et al. [2007]	✓			✓	✓	✓
Li and Rao [2007]	✓			✓	✓	✓
Hausken [2007]*	✓		✓	✓		✓
Nizovtsev and Thursby [2007]	✓		✓	✓	✓	✓
August and Tunca [2008]		✓		✓	✓	
Arora et al. [2008]	✓		✓	✓	✓	✓
Choi et al. [2010]		✓		✓	✓	
Romanosky et al. [2010]*	✓		✓		✓	
Moore et al. [2010]	✓			✓	✓	
Liu et al. [2011]*	✓		✓	✓		✓
Gordon et al. [2015]	✓		✓			✓
Laube and Böhme [2015a]*		✓	✓			✓
Naghizadeh and Liu [2016]	✓			✓		✓
Laube and Böhme [2016]*	✓		✓	✓		✓

Chapter 3

Voluntary private cyber risk information sharing

Based on the framework proposed in the last chapter, we subsequently review literature that studies the economics of defenders' voluntary private cyber risk information sharing. We revisit the practical context relating to the channel for such information sharing in Section 3.1. Thereafter, we examine the academic literature. Section 3.2 inspects theoretical works that investigate defenders' private information sharing. This is followed by an inquiry into corresponding empirical literature in Section 3.3. We distill trends in the literature and identify directions for future research in Section 3.4.

3.1 Channel

Defenders can privately share their cyber risk information directly or indirectly via third parties. We highlight channels for direct information exchange in Section 3.1.1. Selected intermediaries who moderate information sharing are presented in Section 3.1.2.

3.1.1 Direct information sharing

Information exchanged in a private sharing arrangement can only provide value if their meaning is comprehensible to the involved parties. This requires common syntax and semantics. In this context, there exist diverse standards that defenders can agree upon [Kampanakis, 2014; Skopik et al., 2016]. Two popular standards are: the "Incident Object Description Exchange Format" (IODEF), and the "Structured Threat Information eXpression" (STIX) language. IODEF provides a framework for the communication of operational and statistical cyber risk information. The standard is compatible with

the widely adopted "Intrusion Detection Message Exchange Format" (IDMEF), defining data formats and specific information exchange procedures. However, IODEF is not as expressive as the STIX language [Barnum, 2012]. STIX defines a structure for machine-processable storage, analysis, and sharing of cyber risk information. It complements other standards such as the "Cyber Observable eXpression" (CybOX) language, the "Malware Attribute Enumeration and Characterization" (MAEC) language, and the "Trusted Automated eXchange of Indicator Information" (TAXII) standard. CybOX is used to encode events that (can) occur in ICT systems. By contrast, MAEC enables to describe malware based on its attributes, e. g., attack patterns. The TAXII standard is an automated transport mechanism for cyber risk information expressed with STIX. It enables (almost) real-time machine-to-machine information transfers, which are growing in popularity [Barnum, 2012]. In practice, the implementation of such automated cyber risk information sharing between defenders is typically moderated.

3.1.2 Moderated information sharing

Prominent intermediaries who moderate all kinds of voluntary private cyber risk information sharing are non-profit, government-facilitated "Information Sharing and Analysis Centers" (ISACs) or – more generally – "Information Sharing and Analysis Organizations" (ISAOs). The concept of ISACs was first introduced in 1998 by the United States (US) Presidential Decision Directive 63, and comprised the objective to help protect the US critical infrastructure from cyber risk. Today, organizations similar to ISACs, sometimes referred to as ISAOs, exist all over the world and no longer focus exclusively on the protection of critical infrastructure. In order to follow their objectives, these organizations collect, analyze, and share cyber risk information with their members. Furthermore, they motivate members to contribute own information. However, members have a lot of leeway in deciding what to share, and often are concerned to exchange information with competitors or participating authorities [Dacey, 2003]. On this account, members' free riding behavior is conceivable [Varian, 2002].

Another type of intermediary are non-profit "Computer Emergency Response Teams" (CERTs), which traditionally act as trusted clearinghouses for all kinds of defenders' cyber risk information. Similarly to ISACs, these teams have the objective to support their constituencies in reducing cyber risk. To this end, they promote and moderate cyber risk information exchange between their constituencies, besides providing diverse other services [West-Brown et al., 2003]. CERTs can broadly be differentiated based the constituencies that they serve [Kruidhof, 2014]. Governmental CERTs provide their services to governmental staff that protects ICT system infrastructures. National

CERTs serve a broad audience ranging from firms and government organizations to households. And industry CERTs have specific industrial organizations as their constituencies. Though, for the same reasons as members of ISACs, the constituencies of CERTs may lack incentives to share their private cyber risk information with others.

Government authorities and their hubs are other types of non-profit organizations that act as intermediaries for cyber risk information. A prominent example is InfraGard, whose mission is to promote voluntary information exchange between US firms and the "Federal Bureau of Investigation" (FBI).[1] Therefore, the organization establishes mutual non-disclosure agreements between all participating firms and the FBI. This may lead to trust among members, and thus information sharing incentives. Firms' shared information then puts the FBI in a better informed position. In turn, the FBI can forward received information to overcome information asymmetries between all InfraGard members. Thus, participants may manage their cyber risk more effectively.

Additionally, the economic value of attack information led to the formation of information markets [Böhme, 2006]. For instance, some vendors from the security industry buy vulnerabilities to enhance own products and provide security advisory services to their customers [Ransbotham et al., 2012].[2] Other brokers maintain platforms for bug challenge programs that can be used by vendors to offer monetary rewards (also known as bug bounties) in return for information on product or service vulnerabilities [Schechter, 2004].[3] Finally, there are brokers who base their business model on trading vulnerabilities with high associated cyber risk.[4] All of these brokers could serve attackers, hence their formation raises ethical questions [Egelman et al., 2013] that careful readers may derive from our subsequent literature reviews.

3.2 Theoretical literature

There are two streams of theoretical literature that investigate the economics of voluntary private cyber risk information sharing. First, works analyzing attack information sharing between defenders and firms, examined in Section 3.2.1. And, second, studies on firms' general private information sharing with each other, inspected in Section 3.2.2.

[1] For more information regarding this organization, visit its website (https://www.infragard.org).
[2] Examples of such vendors are iDefense (Verisign) with its "Vulnerability Contributor Program" (http://www.verisign.com/en_US/security-services/security-intelligence/index.xhtml) and Trend Micro with the "Zero Day Initiative" (http://www.zerodayinitiative.com).
[3] For instance, HackerOne (https://hackerone.com), and BugCrowd (https://bugcrowd.com).
[4] Prominent examples are Zerodium (http://www.zerodium.com) and the HackingTeam (http://www.hackingteam.it).

3.2.1 Attack information sharing of defenders with firms

The studies reviewed in this section analyze the economics of defenders' voluntary *vulnerability information* sharing with firms: vendor disclosure [Cavusoglu et al., 2007]; responsible disclosure [Arora et al., 2008; Cavusoglu et al., 2007]; and market disclosure [Kannan and Telang, 2005; Li and Rao, 2007]. For a coherent introduction of all institutionalized vulnerability disclosure regimes, we also propose aspects of defenders' non-disclosure [Moore et al., 2010] and immediate disclosure [Cavusoglu et al., 2007; Nizovtsev and Thursby, 2007] policies here. All authors of the presented works devise specific game-theoretic models in order to examine the effect of disclosure policies on vendors' patch provisioning behavior. However, for the sake of simplicity, this aspect is not part of our unified formal model presented in Section 2.4.

Non-disclosure means that defenders keep vulnerability information secret [Moore et al., 2010]. This cause concern to intermediaries acting in the interest of nations, e. g., government authorities. If such a defender gets to know about a vulnerability, he commonly trades-off between a disclosure to the affected vendor and stockpiling [Anderson, 2001]. The former may initiate a patch release and can thus help all consumers with vulnerable systems to secure themselves. Contrarily, the latter maintains an offensive readiness as it enables exploitations of enemy networks, thereby leaving consumers' ICT systems at risk. Moore et al. [2010] find that an intermediary acting in the interest of nations often prefers to stockpile vulnerabilities over their disclosure to vendors.

Defenders may follow a *vendor disclosure* policy to stimulate the release of a patch [Cavusoglu et al., 2007]. Yet, a vendor not necessarily releases a patch upon receiving vulnerability information. The reason is that a vendor does not suffer from breaches as much as consumers, such that he does not internalize the costs of insecurity and thus may have few incentives to provide a patch: developing and testing a patch causes *patching costs* [Telang and Wattal, 2007]. While waiting for a patch, expected losses at vulnerable consumers increase over time due to breaches of their ICT systems. Still, Cavusoglu et al. [2007] find that it may be optimal for defenders to disclose a vulnerability to the affected vendor, if he otherwise does not release a patch with certainty.

In order to establish incentives for a vendor to provide a patch, defenders can follow an *immediate disclosure* policy by publishing vulnerability information [Cavusoglu et al., 2007; Nizovtsev and Thursby, 2007] – thus adhering to Kerkhoffs's principle that "there is no security through obscurity" [Swire, 2004]. This simultaneously informs all actors about the vulnerability. As attackers get informed, the policy exposes consumers to attacks. Yet, consumers may also recognize the vulnerability, enabling them to develop own fixes or demand a patch from the affected vendor. In fact, such consumer demand

may create incentives for the vendor to provide a patch: without a patch, he incurs *costs of lost sales* [Telang and Wattal, 2007] due to consumers' dissatisfaction and accompanied reputation damages. We revisit implications of these costs on a vendor's patch provisioning in Section 4.2.1, under the topic of public cyber risk information sharing.

Responsible disclosure is a trade-off between vendor and immediate disclosure. Initially, defenders share information on a vulnerability with a trusted intermediary, e. g., a CERT [Arora et al., 2008; Cavusoglu et al., 2007]. This intermediary is committed to publish received information eventually, but not before setting the affected vendor a grace period to release a patch. It determines this grace period by minimizing the associated social cost. A longer grace period exposes consumers to more cyber risk, as the probability of attackers' vulnerability exploitation increases over time. At the same time, a longer grace period decreases the vendor's patching costs by giving him more time for development. If the intermediary publishes the vulnerability, this exposes consumers to attacks unless they can find and apply a patch or fix. Simultaneously, the vendor may incur costs of lost sales, which increase as long as no patch is provided. The analysis of Cavusoglu et al. [2007] reveals that responsible disclosure ensures a patch release of the affected vendor, but not due to the costs of lost sales. In fact, if consumers can easily fix a vulnerability, providing the affected vendor with a grace period may not ensure his release of a patch: once the vulnerability gets published, consumers can fix it themselves such that the affected vendor's savings in patching costs may outweigh his costs of lost sales. This result contrasts the findings of Arora et al. [2008]. They observe that a trusted intermediary's threat to publish a vulnerability can push the affected vendor towards releasing a patch quickly. Also, if consumers can develop fixes themselves, this provides the intermediary with leverage to reduce the grace period.

In contrast to all other policies, *market disclosure* refers to defenders' vulnerability information sharing with a broker in return for some monetary reward, rather than the prospect of a patch release [Kannan and Telang, 2005; Li and Rao, 2007]. By adjusting rewards, the broker can set incentives for defenders to increase their efforts regarding vulnerability detection and information sharing. The broker then uses received information to offer protection to his customers in a second market, who pay for this service. Note that this protection is only valuable to customers while no other fixes or patches are available. Thus, the broker does not benefit from information forwarding to vendors. Instead, he could even increase the value of his service by leaking vulnerabilities to attackers, exposing non-customers and creating value for customers. Based on these rationales, Kannan and Telang [2005] explore the welfare implications of market disclosure in comparison to responsible disclosure. Their model

analysis suggests that market disclosure almost always lead to lower social welfare than responsible disclosure, as the broker has incentives to leak information. The former can only outperform the latter if the broker is not allowed to leak information, and vulnerabilities are hard to detect. In this scenario, the broker who offers monetary rewards receives more information on vulnerabilities than the trusted intermediary. In turn, customers of the broker enjoy protection while the intermediary's constituencies stay unprotected, such that market disclosure may overall lead to higher welfare than responsible disclosure. Li and Rao [2007] observe that on top of that, the broker's presence can incentivize vendors who are informed on vulnerabilities to release patches rather late: vendors know that their customers can subscribe to the broker, and may thus manage the cyber risk associated with vulnerabilities regardless of patches.

In summary, the reviewed works predict that defenders' private vulnerability disclosure policies have an impact on affected vendors' patch provisioning. Thereby, incentives for vulnerability disclosure seem to be driven by expected patch releases, besides monetary compensations. Yet, vendor disclosure does not incentivize patch provisioning. The prevailing disincentives often root in vendors' patching costs. Also, market disclosure may not lead to patch releases. The reason is that brokers have no incentives to forward acquired information, facilitating the implementation of competing countermeasures. By contrast, responsible disclosure can incentivize vendors to release patches. This is because the policy eventually leads to vulnerabilities' announcement to the public, entailing costs of lost sales for affected vendors if no patches are provided.

3.2.2 General information sharing between firms

In contrast to the works on defenders' attack information sharing with firms, reviewed before, we now examine literature using game theory to investigate the economics of firms' voluntary *cyber risk information* (i. e., as a general category) sharing with each other. To this end, some authors analyze one-shot information sharing games [Gal-Or and Ghose, 2005; Gordon et al., 2003; Hausken, 2007; Liu et al., 2011], while others consider repeated game formulations [Gordon et al., 2015; Naghizadeh and Liu, 2016]. All one-shot game models can be mapped into our unified formal model in Section 2.4.

Gordon et al. [2003] were among the first to examine the economics of two symmetric firms' cyber risk information sharing. Both firms are affected by opportunistic attacks to their ICT systems. The firms can invest in security and engage in information sharing to manage the associated cyber risk. Thereby, information sharing does not lead to costs and supports the other firm in reducing its breach probability $\partial P_i/\partial y_{1-i} < 0$. For instance, if one firm shares observed attack source addresses, the other may enhance

the effectiveness of its preventive controls. Overall, both firms' objective function is

$$\arg\min_{x_i,s_i} \; P_i(x_i, y_{1-i}) \cdot I_i + C_i(x_i) \; . \tag{3.1}$$

The analysis of this model reveals that firms' cyber risk information sharing is a strategic substitute to security investment: if information are shared between both firms, this makes their investment more effective. Consequently, firms that share information may invest less into security to keep their previous security level. However, in equilibrium, the firms' strategy is to abstain from information sharing altogether. Rather, each firm attempts to free ride on the other's actions. This leads to security under-investment at both firms, as compared to a social planner's optimal investment.

Concurrently to Gordon et al. [2003], Gal-Or and Ghose [2005] studied the effects of two vendors' cyber risk information sharing on their profit. The authors assume market reactions to decisions of vendors, hence we classify their model as public information sharing and discuss it in Section 4.2.2. However, the model also captures that vendors' private information sharing can lead to preemptive security investment cost savings, i. e., $\partial C_i / \partial y_{1-i} < 0$. But this only holds if both vendors' products are logically interdependent $\gamma_s > 0$, referring to ICT product similarity [Katti et al., 2005; Thomas et al., 2016]: if products of two vendors use similar components, and one vendor learns to secure them, this information can help the other to make more targeted security investment.

The work of Hausken [2007] is closely related to [Gordon et al., 2003]. The main difference is that ICT systems of the two symmetric firms in his model are affected by strategic attacks with resources r_i, r_{1-i}, rather than opportunistic attacks. To manage associated cyber risk, the firms can invest in security and share information. Thereby, one firm's shared information supports the other in reducing its breach probability $\partial P_i / \partial s_{1-i} < 0$. Two other effects of information sharing are conditioned on firms' interdependence $\gamma \in [-1, 1]$. Positive interdependence captures a partnership, e. g., in form of a supply chain relationship, where breaches at each firm propagate to the other. Thus, a firm may reduce its own breach probability by sharing information with its partner $\partial P_i / \partial s_i < 0$. By contrast, negative interdependence models a strategic conflict, e. g., strong market competition, where firms benefit from breaches at the other. Consequently, a firm's information sharing makes the competitor stronger, and itself becomes a more interesting target for the attacker $\partial P_i / \partial s_i > 0$. Either way, firms' shared information may leak to the public, resulting in secondary losses $\partial I_i / \partial s_i > 0$ as mentioned in Section 4.2.3. Taking this into account, both firms' objective function is

$$\arg\min_{x_i,s_i} \; P_i(r_i, r_{1-i}, x_i, x_{1-i}, s_i, s_{1-i}) \cdot I_i + I_i(s_i, s_{1-i}) + C_i(x_i) \; . \tag{3.2}$$

The analysis of this model setup reveals two equilibrium situations, rediscovered also by Khouzani et al. [2014b]. With negative interdependence, firms' strategy is to abstain from cyber risk information sharing. By contrast, if there is positive interdependence, firms shift parts of their efforts from security investment to information sharing. This shift is promoted by, among other implicit model parameters, an increase in the security investment costs and firms' interdependence.

The model of Liu et al. [2011] is closely related to the setup in [Gordon et al., 2003], too. It similarly considers two symmetric firms that are affected by opportunistic attacks on their ICT systems. Thereby, an attack on either firm's system realizes independently with probability 1/2. Firms can invest in security and share cyber risk information with each other to reduce the associated cyber risk. A firm's investment is costly, while information sharing is free and supports the other firm in reducing its breach probability $\partial P_i/\partial y_{1-i} < 0$. However, the latter effect only holds if both firms have logically interdependent ICT systems $\gamma_s \in [0, 1]$. Also, the effect of information sharing on firms' cyber risk depends on the type of information assets protected:

- *Complementary assets.* Assets are complementary if the attacker has to breach both firms' information assets and combine them to gain economic advantages. Consequently, a firm's cyber risk information sharing reduces its own cyber risk $\partial R_i/\partial y_i < 0$, which can be defined as

$$R_i(x_i, x_{1-i}, s_i, s_{1-i}) = P_i(x_i, \gamma_s \cdot y_{1-i}) \cdot P_{1-i}(x_{1-i}, \gamma_s \cdot y_i) \cdot I \ . \tag{3.3}$$

  It turns out that if firms possess complementary assets, their strategy is to fully share information. Simultaneously, both firms tend to under-invest in security.

- *Substitutable assets.* Assets are substitutable if an attacker who compromises one firm's assets breaches both firms, e. g., because firms store each other's data to create redundancy. Now, if a breached firm is required to compensate losses of the non-breached other, its cyber risk information sharing may in fact increase own cyber risk $\partial R_i/\partial y_i > 0$, which can be defined as

$$R_i(x_i, x_{1-i}, s_i, s_{1-i}) = \frac{1}{2} \cdot P_i(x_i, \gamma_s \cdot y_{1-i}) \cdot (I + (1 - P_{1-i}(x_{1-i}, \gamma_s \cdot y_i)) \cdot I) \ . \tag{3.4}$$

  The authors find that if firms possess substitutable information assets, their equilibrium strategy is to abstain from cyber risk information sharing altogether. Simultaneously, each firm initiates suboptimal security investment.

The previous one-shot game models are complemented by repeated game models, also containing only two firms [Gordon et al., 2015; Naghizadeh and Liu, 2016]. Gordon et al. [2015] propose a two-period game model in which firms' information sharing reduces the uncertainty associated with security investment. These authors' analysis indicates that if firms share information, they rather proactively invest in security (i. e., in the earlier period) than reactively (i. e., in the later period). By contrast, Naghizadeh and Liu [2016] study an infinitely repeated voluntary information sharing game model in which each firm can monitor whether the other free rides on exchanged information. This enables both firms to condition future information sharing decisions on past interactions, helping them to create some form of mutual trust over time. The analysis of this setting reveals that, depending on the efficiency of monitoring, firms may coordinate their information sharing decisions in a way that improves welfare.

In summary, the reviewed works indicate that time influences the stability of firms' strategy to voluntarily exchange cyber risk information. The analysis of one-shot game models predicts that firms' information sharing strategies are often fragile and socially suboptimal. This is because firms may intend to free ride on each others' information. By contrast, the analysis of a repeated game model reveals that stable information sharing strategies can evolve. The reason is that firms who interact repeatedly may use mechanism to monitor each others' actions, enabling to build mutual trust over time.

3.3 Empirical literature

Empirical evidence is the ultimate bar to test the validity of the theoretical works reviewed before. Two streams of empirical works investigate the effects associated with voluntary private cyber risk information sharing. The first stream sheds light into effects of defenders' attack information sharing with firms, explored in Section 3.3.1. The second stream examines firms' attack information sharing, inspected in Section 3.3.2.

3.3.1 Attack information sharing of defenders with firms

The works reviewed in this section empirically inquire the effects associated with defenders' voluntary *vulnerability information* disclosure to vendors [Finifter et al., 2013; Maillart et al., 2016; Zhao et al., 2015], brokers [Li and Rao, 2007; Ransbotham et al., 2012], and *abuse information* reporting to firms in general [Cetin et al., 2016; Vasek and Moore, 2012]. Almost all works allow to reason about the effects of defenders' attack information sharing on the cyber risk exposure of consumers.

Several authors [Finifter et al., 2013; Maillart et al., 2016; Zhao et al., 2015] explore publicly available data on third party bug challenge programs. Using a regression analysis, Zhao et al. [2015] find that the amount of defenders' vulnerability contributions to these programs significantly depends on the rewards offered by vendors. Furthermore, they observe that many reported vulnerabilities are noise (i. e., invalid) rather than signals (i. e., valid), and more productive contributors have a higher signal-to-noise ratio. Nevertheless, contributors with low signal-to-noise ratios also seem to provide value to owners of bug challenge programs, as they bring in diversity in terms of expertise and tools to discover new vulnerabilities. The regression analysis by Maillart et al. [2016] indicates that the probability for contributors to detect vulnerabilities relevant to a program decreases in the amount of already reported valid vulnerabilities. This result goes in line with the observation that cyber risk goes down as vulnerabilities are reported to vendors [Ozment and Schechter, 2006]. It also supports the finding that contributors diversify efforts among programs for which vulnerabilities are still easy to detect. Vendors may increase rewards to counter contributors' migration, but empirical evidence suggests that this can hardly stop switchers. Yet, an exploratory analysis by Finifter et al. [2013] indicates that bug challenge programs are cost-effective for vendors, as compared to hiring full-time security researchers to find bugs. By implication, these programs may induce wage pressure for security researchers, raising ethical questions regarding their fair compensation. For instance, this is brought up by Egelman et al. [2013]: "[are bug challenge programs] just a way for companies like Microsoft, Google and Apple to outsource product testing on the cheap?"

Ransbotham et al. [2012] take another perspective by analyzing how defenders' market disclosure of vulnerabilities effect the ICT system security of consumers, as compared to the effects of immediate and responsible disclosure policies. By conducting a regression analysis on proprietary IDS and publicly available vulnerability data, they confirm four hypothesis. First, market disclosure decreases the number of distinct ICT systems attacked. A possible inference is that brokers who buy vulnerabilities effectively use the acquired information to protect their customers. Second, disclosures to brokers delay the diffusion of attacks since the vulnerabilities get published later. This result indicates that brokers rarely leak information, which is consistent with the findings of Li and Rao [2007]. Third, market disclosure of vulnerabilities decreases the probability of vulnerable ICT systems to get attacked. A reason is that the longer brokers keep vulnerability information private, the more of their customers manage associated cyber risk, which possibly deters attackers. Forth and finally, disclosures to brokers reduce the overall volume of attacks based on the vulnerabilities, which is

mainly driven by exploits. Consequently, attackers seem to abstain from exploiting vulnerabilities shared with brokers, hypothetically because attacks succeed less often.

By contrast to the previous works, there are also several authors [Cetin et al., 2016; Vasek and Moore, 2012] that study the effectiveness of defenders' abuse reporting to firms. Vasek and Moore [2012] experiment with information on website malware infections obtained from a malware feed. The authors reported the information to affected website providers during their study. They observe that if the reports include details on why websites are distributing malware, this reduces the time until providers clean up their websites as compared to the case of minimal reports. In fact, minimal reports turn out to be as ineffective as no reports at all. This highlights the importance of infection details in abuse information sharing. In a comparable study, Cetin et al. [2016] identify and notify website providers whose sites have become part of a botnet. Their analysis reveals that the reputation of a notifying defender, indicated by his email address, does not seem to influence providers' response time to abuse reports.

In summary, the reviewed literature provides empirical support that defenders' attack information sharing with firms decreases cyber risk exposure of consumers. Defenders' contributions to bug challenge programs become harder in the number of already reported vulnerabilities, indicating that vendors fix their products based on received information. Also, strong empirical evidence suggests that market disclosure of vulnerabilities positively affect the security level of consumers. Similarly, if defenders share detailed abuse reports with affected firms, the latter appear to recognize and use them to cleanup compromised hosts. All of these types of cyber risk reductions seem to incentivize defenders' attack information sharing in the first place, while vulnerability markets also stimulate participation by rewarding contributions.

3.3.2 Attack information sharing between firms

This section covers works that empirically investigate firms' *abuse information* sharing [Moore and Clayton, 2008; Vasek et al., 2016]. Evidence provided by reviewed works allows to reason about indirect effects of such sharing on consumers' cyber risk.

Moore and Clayton [2008] report missed opportunities in the fight against phishing due to brand-protection firms' lack of incentives to exchange attack information. These firms base their business model on the take-down of phishing websites for customers, who are typically banks. Brand-protection firms compete on the market and expect competitive advantages from a strategic differentiation in their private phishing feeds. The authors analyze the phishing feeds of two such firms, and compare lists of their customers. This reveals a substantial overlap of both firms' feeds, though at different

points in time. A further evaluation indicates that if the firms shared private attack information, this would have measurably improved their performance in combating phishing websites and thus provided considerable monetary benefits to their customers.

In a related study, Vasek et al. [2016] evaluate how firms' reporting of blacklisted "Uniform Resource Locators" (URLs) (for distributing malware) to responsible hosting providers effects cleanup efforts. They find that such information sharing leads to a considerable number of cleanups. More importantly, the results indicate that sharing reduces the likelihood for reported URLs to get recompromised. An optimistic interpretation suggests that providers make efforts to fix the root causes of compromise. Yet, in the long run, the picture is more diverse with some providers doing better than others.

In summary, the works reviewed in this section indicate that private abuse information sharing between firms can lead to a reduction of consumers' cyber risk exposure. If firms cooperate in the fight against phishing, they can take-down more phishing websites faster. Furthermore, it turns out that firms' sharing of information on compromised websites with affected hosting providers often results in cleanups. Thus, both types of information exchange entail that adverse websites no longer expose naïve consumers on the web. However, it appears that firms only have incentives to share abuse information in the first place if this does not lead to competitive disadvantages.

3.4 Trends and research directions

Our previous reviews indicate that defenders may only have few incentives to engage in voluntary private cyber risk information sharing. The investigated instances of information sharing are primarily enabled by mechanism that help defenders to build mutual trust over time, monetary compensations by information recipients, and expected reductions in cyber risk. The prevalent barriers for information exchange are incentives to free ride on others' actions and forgone profit. However, against the backdrop that defenders' voluntary private sharing is commonly observed in the real world (cf., e. g., [National Concil of ISACs, 2017]), these results do not seem to be conclusive and thus require further investigation. This motivates us to subsequently evaluate trends in the reviewed literature and distill directions for future research. For this purpose, we use a systematization of previously examined works, depicted in Table 3.1. We first discuss works on defenders' voluntary private information sharing with firms in Section 3.4.1. Thereafter, we focus on literature regarding cyber risk information sharing between firms in Section 3.4.2.

Table 3.1 Reviewed literature on voluntary private cyber risk information sharing.

Information	... with firms	
Actor type	Theoretical works	Empirical works
Attack information sharing of ...		
defenders ...	Kannan and Telang [2005]	Li and Rao [2007]
	Cavusoglu et al. [2007]	Ransbotham et al. [2012]
	Li and Rao [2007]	Vasek and Moore [2012]
	Arora et al. [2008]	Finifter et al. [2013]
	Moore et al. [2010]	Zhao et al. [2015]
		Cetin et al. [2016]
		Maillart et al. [2016]
firms ...	Gordon et al. [2003]	Moore and Clayton [2008]
	Gal-Or and Ghose [2005]	Vasek et al. [2016]
	Hausken [2007]	
	Liu et al. [2011]	
	Gordon et al. [2015]	
	Naghizadeh and Liu [2016]	
Control information sharing of ...		
firms ...	Gordon et al. [2003]	
	Gal-Or and Ghose [2005]	
	Hausken [2007]	
	Liu et al. [2011]	
	Gordon et al. [2015]	
	Naghizadeh and Liu [2016]	
Impact information sharing of ...		
firms ...	Gordon et al. [2003]	
	Gal-Or and Ghose [2005]	
	Hausken [2007]	
	Liu et al. [2011]	
	Gordon et al. [2015]	
	Naghizadeh and Liu [2016]	

3.4.1 Information sharing of defenders with firms

Although there are plenty of theoretical and empirical studies that investigate the economics of defenders' *vulnerability* disclosure to firms, room for new works still exists. Future theoretical works may analyze how third party bug challenge platform owners can incentivize both defenders and vendors to participate. Furthermore, an investigation of these platforms' implications on the cyber risk in economies is required. Besides, the results of existing theoretical works on private vulnerability information sharing lack empirical validations. They predict that defenders should favor disclosures to trusted intermediaries rather than vendors or brokers, as the former increases welfare the most of all. However, no empirical study measures and compares changes in consumers' costs and losses accompanied by the different disclosure policies. Arguably, these costs and losses are notoriously hard to estimate, due to the private nature of the policies. Nevertheless, against the backdrop that many defenders advocate vulnerability information sharing with trusted intermediaries, such measurements are much needed.

We additionally find only few literature on the economics of defenders' voluntary *abuse information* sharing with firms. This motivates future studies on this topic which may follow the research agenda provided by Jhaveri et al. [2017].

Furthermore, our review reveals research gaps concerning works that study the economics of defenders' *ICT system crash information* (i. e., control or impact information) sharing with firms, as indicated by missing categories in Table 3.1. These gaps are surprising, as in reality, defenders frequently share ICT system crash reports with vendors [Kim et al., 2011]. An analysis of defenders' incentives to consent with such information exchange may help vendors to fill up their crash report database, which they use to infer product or service vulnerabilities. And an examination of vendors' reactions to received reports might allow defenders to make more informed reporting decisions. While empirical studies on the former topic appear to be feasible with rather moderate effort, e. g., by laboratory experiments, works on the latter are harder to undertake, e. g., based on field experiments or by cooperating with the ICT industry.

3.4.2 Information sharing between firms

Many theoretical studies investigate the economics of firms' voluntary private *cyber risk information* sharing. All of these works consider a fully transparent exchange of information between firms. We are not aware of studies inquiring if the evolution and prevalence of cyber risk information sharing technologies that preserve confidentiality or privacy alter firms' incentives to exchange information. Yet, this may well be the

case, as the use of such technologies can ease the concern of information abuse. An examination of this topic is relevant, particularly in view that related technologies already have an impact on other types of firms' incentives [Acquisti and Varian, 2005].

Furthermore, Table 3.1 illustrates that there is almost no empirical literature examining the obstacles for and effects of firms' *cyber risk information* sharing. Some practitioners particularly in the EU name data protection as a main obstacle for information sharing: if personal data are (unintentionally) shared, firms become liable to prosecution. Yet, we are not aware of studies examining this specific issue in detail. Also, the research gap regarding the effects of firms' voluntary information sharing is remarkable against the backdrop of manifold corresponding theoretical works. Of course data availability is an issue. But this could be overcome by scholars' collaboration with firms who participate in private information sharing agreements. For instance, it is reasonable to work together with firms that exchange abuse information [Jhaveri et al., 2017] or collaborate on security [Meng et al., 2015]. These firms' information sharing effectiveness may then be determined with the help of metrics, e. g., inspired by Thomas et al. [2016]. Once the right data becomes available to scholars, obstacles for and effects of firms' private information sharing should be determinable. Publicly available sources that can provide complementary data are discussed in the next chapter.

Chapter 4

Voluntary public cyber risk information sharing

We now use our framework from Chapter 2 to review works that study the economics of defenders' voluntary public cyber risk information sharing. Specifically, we provide some context by mentioning practical information on the channels that can be used to publish or receive cyber risk information in Section 4.1. Thereafter, we examine the academic literature. Section 4.2 covers theoretical literature and Section 4.3 empirical works on defenders' cyber risk information disclosures to the public. We distill trends in the reviewed literature and identify future research directions in Section 4.4.

4.1 Channel

Defenders can either directly consult public sources in order to publish or receive cyber risk information, or involve an intermediary. Prominent sources that can be used directly are presented in Section 4.1.1. We introduce selected intermediaries who moderate defenders' cyber risk information announcements to the public in Section 4.1.2.

4.1.1 Direct information sharing

Diverse sources primarily publish attack information. For example, numerous public blacklists [Metcalf and Spring, 2014] and several organizations make available up-to-date information on threats to ICT systems.[1] Furthermore, there are databases that

[1]Organizations that make available threat information include, e. g., the Honeynet Project (https://www.honeynet.org) and PhishTank (https://www.phishtank.com).

provide information on recently observed attack source addresses.[2] Also, a large amount of web resources contain information on available exploits.[3] Other sources publicly share information on vulnerabilities. For instance, the "BugTraq" mailing list was famous for defenders' immediate disclosure.[4] Most of the published vulnerabilities are then stummarized in databases.[5] It is common practice to unanimously identify these vulnerabilities by the Common Vulnerabilities and Exposures (CVE) standard.[6] And supplementary ratings regarding the cyber risk associated with public vulnerabilities are often based on the Common Vulnerability Scoring System (CVSS).[7]

Besides, various sources focus on the public disclosures of control information. A prominent example is the Common Criteria (CC), standardizing good security practices.[8] Similarly, the US Privacy Rights Clearinghouse (PRC) informs defenders on best practices to strengthen their privacy.[9] Additionally, defenders may query the database of the US Securities and Exchange Commission (SEC) to get informed about self-reported security controls implemented by firms listed on US stock exchanges.

Other sources specialize on impact information announcements to the public. For instance, historical information on all kinds of security and privacy breaches are accessible via diverse news archives.[10] Furthermore, a considerable number of breaches published by the press is summarized in various breach databases and reports.[11]

4.1.2 Moderated information sharing

Defenders' voluntary disclosures to the public are often moderated by intermediaries that act as trusted clearinghouses for cyber risk information [Koivunen, 2012]. These clearinghouses regularly publish aggregated reports, checklists, and datasets that help others to manage cyber risk. The most prominent of such intermediaries are national CERTs. They allow defenders to responsibly disclose vulnerabilities, as explained in

[2] A prominent example is the DShield database (http://dshield.org).
[3] For example the websites of Packetstorm (http://packetstormsecurity.org), SecurityVulns (http://securityvulns.com), and Metasploit (http://www.metasploit.com).
[4] Today, "BugTraq" is complemented by other lists (http://seclists.org).
[5] Prominent vulnerability databases are the United States National Vulnerability Database (NVD) (http://nvd.nist.gov) and SecurityTracker (http://securitytracker.com). Their entries often contain links to advisories and patches by vendors.
[6] Information on this standard are accessible at the CVE's website (http://cve.mitre.org).
[7] These scorings are supervised by the FIRST organization (https://www.first.org/cvss).
[8] See the CC's website for further details (https://www.commoncriteriaportal.org/cc/).
[9] Further information can be found on the PRC's website (www.privacyrights.org).
[10] E. g., LexisNexis (https://www.lexisnexis.com) and ProQuest (http://www.proquest.com).
[11] For instance, databases of the PRC (https://www.privacyrights.org/data-breaches), DataLossDB (http://datalossdb.org), the VERIS Community (http://vcdb.org), or the Identity Theft Resource Center (ITRC) breach report (http://www.idtheftcenter.org).

Section 3.2.1. In this context, CERTs announce advisories for reported vulnerabilities, which may help consumers to safeguard their ICT systems despite no patches being available. Furthermore, when the grace period given to vendors elapses, CERTs are committed to announce vulnerabilities. In addition to CERTs, there are numerous other organizations that act as intermediaries moderating cyber risk information disclosures.[12]

Additionally, some specialized firms maintain one-to-many and many-to-many broadcasting platforms that enable attack information sharing of defenders [Thomas et al., 2016].[13] The attack information shared on these platforms includes, but is not limited to, attack source addresses and URLs of malicious websites.

Another type of intermediary are firms who offer services to support vendors' control information sharing with actors. For instance, many firms act as certification organizations. They offer certifications for vendors' ICT products or services, applying more or less comprehensive security quality metrics. These organizations can be categorized based on the certification standards that they use.[14] In an ideal case, issued certificates and seals allow vendors to signal security quality of their products or services to defenders. Finally, another set of specialized firms offers patch management tools that automate patch information sharing between vendors and their customers.[15]

4.2 Theoretical literature

Three streams of theoretical literature investigate the economics of public cyber risk information sharing. First, there are works that analyze defenders' attack information disclosures to the public, reviewed in Section 4.2.1. Second, some scholars examine firms' control information sharing with actors, inquired in Section 4.2.2. And, third, there

[12] Among others, the Anti-PhishingWorking Group (APWG) (http://www.antiphishing.org), an industry association, and the Shadowserver Foundation (https://www.shadowserver.org).

[13] Platforms enabling one-to-many broadcasting are provided by Spamhouse's "Spamhouse" (https://www.spamhaus.org) and Google's "Safe Browsing" (https://www.google.com/transparencyreport/safebrowsing/?hl=en); many-to-many broadcasting is facilitated by IBM's "X-Force Exchange" (https://exchange.xforce.ibmcloud.com), Microsoft's "Interflow" (https://technet.microsoft.com/en-us/security/dn750892), Facebook's "ThreatExchange" (https://developers.facebook.com/products/threat-exchange), and the "Malware Information Sharing Platform" [Wagner et al., 2016].

[14] Prominent standards are the CC and the ISO/IEC 27000-series. Furthermore, there exists a variety of seals attesting privacy protection measures of websites, among others, TRUSTe (https://www.truste.com), BBBOnline (http://www.bbb.org), and EuroPriSe (https://www.european-privacy-seal.eu).

[15] Popular tools include IBM's "BigFix enterprise suite" (http://www-03.ibm.com/security/bigfix/), Lumension's "PatchLink" (http://www.optimal.de/produkte/lumension-patch-link/index.html), Shavlink's "Protect + Empower" (http://www.shavlik.com/de/products/protect/), Ecora's "Patch Manager" (http://www.ecora.com/Ecora/Products/PatchManager.php), and Symantec's "Patch Management Solution" (https://www.symantec.com/products/threat-protection/endpoint-management/patch-management-solution).

is some literature that provides insights into the effects of firms' impact information announcements, explored in Section 4.2.3.

4.2.1 Attack information sharing of defenders with actors

The literature examined in this section [Cavusoglu et al., 2007; Nizovtsev and Thursby, 2007] studies the economics of defenders' immediate *vulnerability* disclosure, based on the rationales given in Section 3.2.1. The works explain defenders' announcement decisions along with the affected vendor's patch release behavior. This causal link has no correspondence in our unified formal model presented in Section 2.4.

Nizovtsev and Thursby [2007] investigate factors that establish incentives and disincentives for defenders to follow an immediate disclosure policy. They predict that the probability for a vulnerability announcement increases with decreasing ease of attackers to launch attacks based on the announced information. Thus, the time it takes between disclosure and the first attacks targeting a vulnerability plays a pivotal role in defenders' announcement decision. Another finding is that the disclosure probability increases in an affected vendor's patching costs. This indicates that defenders may take the strategic decision to impose costs of lost sales on a vendor. In turn, these costs can incentivize this vendor to release a patch despite high patching costs. The authors also argue that there are circumstances where defenders may abstain from a vulnerability announcement, but take on the costs to develop and publish own fixes instead. They find that disincentives for a disclosure get promoted by a high code transparency, e. g., with "open source" software. This transparency allows defenders to become familiar with the software at hand, increasing their chance to develop a functional fix. A last and obvious prediction is that a vulnerability causing more risk is less likely to get published. Instead, defenders focus on developing fixes by themselves.

The authors of [Cavusoglu et al., 2007; Nizovtsev and Thursby, 2007] inquire effects of a vulnerability announcement on an affected vendor's patch release behavior. Their studies indicate that public disclosure can establish incentives for a fast patch release, if the affected vendor's patching costs are below his costs of lost sales. In case that the vendor decides to patch, the time it takes depends on several factors. First, it increases in patching costs. Hence, these costs establish disincentives to act on published vulnerability information. Second, the time until a release declines in the costs of lost sales. Therefore, higher reputation damages associated with a disclosure may increase the vendor's responsiveness. Third and finally, the time decreases in the cyber risk associated with the affected ICT product. Consequently, a vendor that considers his product to be of higher security is less likely to release patches fast.

In summary, the reviewed works predict that defenders' incentives for a vulnerability announcement are promoted mainly by three conditions. First, the cyber risk associated with the vulnerability is low. Second, the announcement is expected to entail only few more attacks on affected ICT systems. And, third, the disclosure leads to a faster patch release by the affected vendor than usual. However, a vendor considers to patch faster only if his patching costs are below his opportunity costs of lost sales.

4.2.2 Control information sharing of firms with actors

The works reviewed in this section introduce models that help to evaluate the economic incentives for vendors' release of *patches* [Arora et al., 2006a; August and Tunca, 2008; Choi et al., 2010], and signals of *product security quality* [Gal-Or and Ghose, 2005]. Yet, only the model of Gal-Or and Ghose [2005] can be mapped into our model from Section 2.4. A common theme among all reviewed models is that they are used to study vendors' control information sharing strategies with profit as objective function.

Arora et al. [2006a] investigate how consumers' product valuation affects a vendor's patching strategy. In their model, consumers have full information about the security of a vendor's product, and value most an early product release with few vulnerabilities. Yet, from the vendor's perspective, an early release is accompanied by more vulnerabilities than a late release, as there is less time for product development. These vulnerabilities may cause breaches, leading to opportunity costs of lost sales. In order to offset these costs, the vendor can implement and disseminate patches [Telang and Wattal, 2007]. But this comes along with patching costs. The model analysis suggests that the vendor must amortize patching costs from sales, indicating that larger markets tend to incentivize him towards shipping a product early and patching it later.

By contrast, other authors [August and Tunca, 2008; Choi et al., 2010] examine effects that a vendor's patch release has on his customers' product valuation, and whether the evolving patching strategies are socially desirable. In particular, a vendor's patch release affects customers' product valuation in three ways. First, it increases the value of a product for those who do patch. This is because the patch helps these customers to protect from breaches. Second, it decreases the value of a product for customers who do not patch, e. g., because of higher deployment costs [Cavusoglu et al., 2008]. A quite subtle reason is that a patch release may enable attackers to reverse engineer the vulnerability, which can result in breaches of unpatched ICT systems. Third, it increases the value of a product for potential new customers. The reason is that those who do patch raise the security of all interconnected ICT systems, as security is interdependent [Kunreuther and Heal, 2003], and thus make a vendor's

product more attractive. Overall, patch releases affect a vendor's profit-maximizing product price. The model of Choi et al. [2010] suggests that a vendor may release patches in a socially suboptimal way if customers' expected cyber risk associated with his product is low relative to their costs of deploying a patch. Contrarily, a vendor releases patches whenever it is socially efficient if his customers' expected cyber risk is high relative to their deployment costs. If deployment costs are sufficiently low, it may even be optimal for a vendor to let product pirates patch [August and Tunca, 2008]. This makes interconnected systems more secure than exclusive patches do. And the so increased product valuation can justify higher prices, increasing a vendor's profit.

Vendors can additionally change consumers' valuation by signaling ICT product security quality. The incentives for vendors to take and disclosure measures that may improve ICT product security are investigated by Gal-Or and Ghose [2005]. Their model considers two symmetric vendors who form an ISAC and compete on the market, as introduced in Section 2.4. Both sell ICT products that are affected by opportunistic attacks. In order to strengthen the products' security, vendors can invest in security and, independently, exchange cyber risk information with each other. The authors assume that consumers are aware of such measures, and consider them in their purchase decisions. Specifically, if a vendor invests in security, consumers perceive an increased product security quality. This leads to a higher demand for the vendor's product $\partial D_i/\partial x_i > 0$. By contrast, if a competitor invests in product security, the vendor may lose some consumers (e. g., switchers) with sensitivity $\gamma_3 \in [0, 1]$. This translates into a negative demand shock for the vendor $\partial D_i/\partial x_{1-i} < 0$. Furthermore, if vendors share cyber risk information with each other, they experience a positive effect on their security investment cost, as discussed in Section 3.2.2. Besides, this action lets consumers perceive a higher security quality of vendors' products with sensitivity $\gamma_2 \in [0, 1]$. The latter translates into an increase of demand $\partial D_i/\partial y_{1-i} > 0$. However, a vendor's sensitive cyber risk information shared in the ISAC may leak to the public, resulting in secondary losses $I_i(s_i, s_{1-i})$, as mentioned in Section 4.2.3. This causes a negative demand shock $\partial D_i/\partial s_i < 0$. All of these effects affect a vendor's ICT product pricing decision $p_i \geq 0$, such that he maximize his profit according to

$$\arg\max_{x_i, s_i, p_i} \; p_i \cdot D_i(x_i, x_{1-i}, s_i, s_{1-i}, p_i, p_{1-i}) - C_i(x_i, \gamma_s \cdot y_{1-i}) \text{ , with} \tag{4.1}$$

$$D_i(x_i, x_{1-i}, s_i, s_{1-i}, p_i, p_{1-i}) = \gamma_2 \cdot y_{1-i} + x_i - \gamma_3 \cdot x_{1-i} + W_i(p_i, p_{1-i}) - I_i(s_i, s_{1-i}) \text{ .} \tag{4.2}$$

Here, the initial demand intercept is at $W_i(p_i, p_{1-i})$. Gal-Or and Ghose [2005] find that vendors have natural incentives to share cyber risk information with each other, if this has large and positive effects on product demand. Also, vendors' cyber risk information sharing and security investment act as strategic complements: investment or information sharing by one vendor can induce the other to increases his investment or information sharing level. Yet, the amount of exchanged information and investment in a market equilibrium falls short of the social optimum. An improvement over the situation in the equilibrium can be induced by the ISAC, if it sets rules for its members to coordinate on information sharing and investment.

In summary, the reviewed works anticipate that vendors have strong incentives to release patches and signal product security quality in case that these actions increase profit. Natural incentives to release patches may exist if this sufficiently raises customers' product valuation. Similarly, vendors engage in measures that improve product security and disclose corresponding information if this increases their demand. Profit considerations may also explain vendors' reluctance to publish impact information, as indicated by the literature on firms' impact information disclosure reviewed next.

4.2.3 Impact information sharing of firms with actors

The literature reviewed in this section introduces assumptions on effects of firms' unintentional private *breach information* disclosure to the public [Gal-Or and Ghose, 2005; Hausken, 2007; Naghizadeh and Liu, 2016]. These assumptions can all be mapped into our unified formal model that we proposed in Section 2.4.

Many authors [Gal-Or and Ghose, 2005; Hausken, 2007; Naghizadeh and Liu, 2016] assume that firms' privately shared breach information can leak to the public, which may result in secondary losses $I_i(s_i, s_{1-i})$ for the affected firm. Thereby, more information sharing elevates the probability of leakage, thus increasing a firm's expected losses $\partial I_i/\partial s_i > 0$. Gal-Or and Ghose [2005] additionally argue that if the firms under consideration are vendors competing on the same market, leakage of their privately shared information causes diverse externalities. In their two-vendor model, one vendor's leakage leads some consumers to switch to the competitor's product, offsetting the competitor's own losses from leakage $\partial I_{1-i}/\partial s_i < 0$. Furthermore, a vendor's marginal secondary losses due to leakage decrease if his competitor intensifies breach information sharing and is thus affected by more leaks himself $\partial^2 I_i/\partial s_i \partial s_{1-i} \leq 0$. A final assumption is that leakage leads to ripple effects on consumers' confidence and trust, captured by $\partial^2 I_i/\partial s_i^2 > 0$ and $\partial^2 I_i/\partial s_{1-i}^2 < 0$. Overall, expected cyber risk due to information leakage can create disincentives for firms to share their private breach information.

In summary, the reviewed works predict that firms have natural disincentives to disclose their breach information to the public because of expected additional cyber risk. Furthermore, firms may not be willing to share their breach information privately with selected others, if there is a chance that this information leaks to the public.

4.3 Empirical literature

We identify four streams of empirical works that investigate the effects associated with public cyber risk information sharing, providing some evidence for assumptions made in the theoretical studies introduced before. First, literature on defenders' attack information disclosures to the public, reviewed in Section 4.3.1. Second, works on firms' control information sharing with actors, inspected in Section 4.3.2. Third, there is a considerable amount of studies on firms' disclosure of impact information, inquired in Section 4.3.3. And fourth, works of authors that investigate the effects of firms' cyber risk information announcements to the public in general, examined in Section 4.3.4.

4.3.1 Attack information sharing of defenders with actors

The literature in this section concerns effects of defenders' *abuse information* [Moore and Clayton, 2011; Tang et al., 2013] and *vulnerability information* [Arora et al., 2010b, 2006b; Frei et al., 2010; Mitra and Ransbotham, 2015; Telang and Wattal, 2007] announcements. Reviewed works point out reactions of vendors affected by disclosures.

Several authors [Moore and Clayton, 2011; Tang et al., 2013] study defenders' disclosure of abuse information to investigate affected vendors' long-term response in cleaning up their compromised hosts. Moore and Clayton [2011] use a hazard model to analyze the lifetime of phishing websites that are either known to the public or closed communities only. Their results indicate that if compromised websites are made public on blacklists, they get re-compromised less often than phishing websites known to closed communities only. In a related controlled intervention study, Tang et al. [2013] compile a list of spam activity that takes place at some autonomous systems, and publish a subset of this list. Their regression analysis suggests that publicly shamed autonomous systems reduce their outgoing spam by about 16 %. With respect to both works, it seems that if defenders publish attack information, this creates incentives for affected vendors to make efforts towards mitigating cyber risk of their products.

While the previously introduced studies focus on longer-term effects, Telang and Wattal [2007] inquire the short-term consequences of vulnerability disclosure to the

public on the affected vendor's stock market value. Therefore, they apply the event study methodology [MacKinlay, 1997], which can be used to measure the impact of published events on stock market listed firms in terms of short-term *Cumulative Abnormal Returns* (CAR). The results suggests that announcements of vulnerabilities lead to a significant negative market value reaction on the day of disclosure. This negative effect is stronger if the affected vendor is small or the market he serves more competitive. Furthermore, disclosures of more severe vulnerabilities have a significantly greater negative impact. However, a vendor's simultaneous release of a patch for a published vulnerability can offset the adverse effect to a large extent. Overall, negative market value effects due to announcements of vulnerabilities may establish incentives for a vendor to develop more secure products, and provide patches in a timely manner.

Effects associated with disclosures of vulnerabilities on the time that affected vendors take to release a patch are investigated in [Arora et al., 2010b; Frei et al., 2010; Shahzad et al., 2012]. Arora et al. [2010b] collect data from diverse channels that defenders use for vulnerability disclosure to the public, and analyze it with a hazard model. They estimate that if disclosures do not take place before a patch is released, the patch comes in (on average) 63 days. Immediate disclosure establishes incentives for vendors to release a patch in (on average) 28 days. And responsible disclosure can lead to an even faster patch release. Also, it seems that vendors' responsiveness to published vulnerabilities increases with their associated cyber risk. This is successfully replicated by Shahzad et al. [2012], though the results of their exploratory study have to be treated with caution as they base on vulnerability and patch data from a sponsored database only. Despite the responsiveness of vendors, an exploratory vulnerability life cycle analysis by Frei et al. [2010] suggest that, in the years between 2000–2007, the release of patches for vulnerabilities was on average slower than the availability of exploits to attackers.

Mitra and Ransbotham [2015] investigate how defenders' immediate disclosure of vulnerabilities effect the ICT system security of consumers, as compared to effects of responsible or market disclosure policies. By conducting a regression analysis on vulnerability and IDS data, they confirm four hypothesis. First, immediate disclosure increases the number of distinct ICT systems that get attacked. This is explained with consumers requiring more time to patch or fix published vulnerabilities. Second, consistent with Arora et al. [2006b], disclosures to the public accelerate the diffusion of attacks. The reason is that announcements instantly provide attackers with master keys to ICT systems. Third, immediate disclosure increases the probability of vulnerable systems to get attacked for the first time, and this probability increases if more vulnerabilities are published simultaneously. The result suggests that attackers perceive

vulnerable ICT systems as attractive targets, and empirically supports the theory that strategic attackers may take advantage of periods where defenders are busy fixing multiple vulnerabilities at once. Forth and finally, defenders' disclosures only marginally affect the overall volume of attacks based on the vulnerabilities, which is mainly driven by exploits. This indicates that announcements of vulnerabilities do not provide strong incentives for attackers to develop and make use of corresponding exploits.

In summary, there is evidence that defenders' attack information disclosures to the public incentivize affected vendors to mitigate cyber risk. Disclosures of abuse information seem to shame vendors, such that they clean up their compromised hosts. Similarly, announcements of vulnerabilities entail short-term market value losses at affected vendors, which can incentivize them to develop more secure products and release patches. Both of these vendor reactions may encourage defenders to publish attack information in the first place. Yet, disclosures of such information can also aid attackers, justifying restricted announcement policies especially if security controls do not exist.

4.3.2 Control information sharing of firms with actors

This section covers works investigating effects of vendors' *patch* releases [Arora et al., 2010a, 2006b; Durumeric et al., 2014; Rescorla, 2003] and *product security information* disclosures [Edelman, 2011; Moores, 2005] on their market position and consumers.

Arora et al. [2006b] investigate how vendors' patch releases affect attack frequencies observed by consumers. They run a regression analysis on publicly available vulnerability data, and proprietary attack data collected from honeypots. The results of this analysis indicate that a patch release for an already published vulnerability reduces the accompanying frequency of attacks. One explanation for this is that ICT product consumers start to protect themselves [Durumeric et al., 2014; Rescorla, 2003], such that attackers anticipate a lower attack success probability. However, after the initial decline, the attack frequency gradually increases again. Another finding is that if a vulnerability and a patch are published simultaneously, the attack frequency ramps up as compared to the situation where both pieces of information stay private. The main reasons behind this is that consumers patch slowly. After the initial rise, the attack frequency increases only gradually over time. Overall, the results indicate that vendors' patch releases may provide new information that help attackers to develop exploits.

In another study, Arora et al. [2010a] examine the patch release behavior of competing vendors. They argue that an increase in the number of competitors, offering ICT products or services that have the same vulnerability, can lead to competition and disclosure effects. Competition effects occur if customers compare the vendors' patch

release times and penalize laggards, e. g., by switching to competitors. At the same time, disclosure effects influence patch release times indirectly: if one vendor releases a patch, thus implicitly publishing a vulnerability, this leads to losses of customer at the other vendors [Gal-Or and Ghose, 2005]. Based on a regression analysis of vulnerability and market data, the authors find that every additional competitor reduces a vendor's patch release time by – on average – 8–10 days. And the accompanying increase in vulnerability disclosure probability decreases the patching time by about 7 extra days.

Other than releasing patches fast, vendors may signal ICT product security quality to strengthen their market positions. In this context, the authors of [Edelman, 2011; Moores, 2005] investigate the effects that website privacy seals have on defenders – in particular individuals. Their examined seals are issued by certification organizations, and shall signal website trustworthiness. Moores [2005] explores if individuals understand or care about such signals. He finds that they have an understanding of the connection between seals and trust. However, the study indicates an ignorance of individuals about seals' form and functioning. Edelman [2011] investigates if firms exploit this ignorance, and observes adverse selection [Akerlof, 1970] in the market for website privacy seals. His results suggests that firms with rather trustworthy websites tend to buy seals from organizations applying strict certification rules. By contrast, firms with little trustworthy websites seek and obtain seals from organizations with rather lax standards. Therefore, the latter firms seem to use certifications as fig leafs.

In summary, the works reviewed in this section provide empirical support for the hypothesis that vendors publish control information to strengthen their own market positions. Specifically, a driving force behind vendors' patch release times appears to be competition on the market. Furthermore, vendors seem to signal website security quality as a means to gain trust of consumers, potentially strengthening the position on the market. Yet, vendors have to respect that the release of some control information may aids attackers in breaching customers, potentially leading to secondary losses.

4.3.3 Impact information sharing of firms with actors

The literature reviewed in this section cover the effects associated with firms' *breach information* disclosures to the public. Most works examine the accompanying short-term effects on firms' stock market values by applying the event study methodology [MacKinlay, 1997]. A summary of these works and their results is provided in Table 4.1. Other reviewed literature [Gordon et al., 2010; Ko and Dorantes, 2006] uses diverse research methods that allow to speculate on long-term effects of firms' breach announcements.

Table 4.1 Short-term effects of breach announcements on firms' stock market values, measured by cumulative abnormal returns (CAR) within a defined period – in days – around the disclosure day (i. e., period "$0 \to 0$").

Authors	Parameters			Effect	
Information type	Time-period	# Events	Period	Affected	CAR
Campbell et al. [2003]					
All kinds of breaches	1995–2000	$n = 43$	$-1 \to 1$	Firms	$-1.88\,\%$
Confidentiality breaches	1995–2000	$n = 11$	$-1 \to 1$	Firms	$-5.46\,\%^{**}$
Cavusoglu et al. [2004b]					
All kinds of breaches	1996–2001	$n = 66$	$0 \to 1$	Firms	$-2.09\,\%^{**}$
All kinds of breaches	1996–2001	$n = 66$	$0 \to 1$	Security industry	$1.36\,\%^{**}$
Acquisti et al. [2006]					
Privacy breaches	2000–2005	$n = 79$	$0 \to 1$	Firms	$-0.58\,\%^{*}$
Ishiguro et al. [2006]					
All kinds of breaches	2002–2005	$n = 70$	$-1 \to 10$	Firms	$-1.89\,\%'$
Confidentiality breaches	2002–2005	$n = 28$	$-1 \to 10$	Firms	$-2.25\,\%^{*}$
Gatzlaff and McCullough [2010]					
Privacy breaches	2004–2006	$n = 77$	$0 \to 1$	Firms	$-0.84\,\%^{*}$
Gordon et al. [2011]					
All kinds of breaches	1995–2007	$n = 121$	$-1 \to 1$	Firms	$-1.36\,\%^{**}$
Availability breaches	1995–2007	$n = 70$	$-1 \to 1$	Firms	$-1.97\,\%^{**}$
All kinds of breaches	1995–2001	$n = 60$	$-1 \to 1$	Firms	$-2.40\,\%^{**}$
All kinds of breaches	2002–2007	$n = 61$	$-1 \to 1$	Firms	$-0.34\,\%$
Wang et al. [2013]					
All kinds of breaches	1997–2008	$n = 101$	$-1 \to 1$	Firms	$-0.15\,\%'$
Gay [2016]					
Privacy breaches	2005–2014	$n = 542$	$0 \to 0$	Firms	$-0.27\,\%^{*}$

Statistical significance thresholds: $' = p < 0.1$; $^{*} = p < 0.05$; $^{**} = p < 0.01$

Scholars report mixed findings regarding the effects of firms' breach disclosures to the public on their stock market values, if breaches are treated generically. The authors of [Campbell et al., 2003; Hovav and D'Arcy, 2004; Kannan et al., 2007] observe that disclosures of firms listed on the US stock market do not lead to statistically significant negative market value effects. Contrarily, the studies in [Cavusoglu et al., 2004b; Gordon et al., 2011; Wang et al., 2013] discover significant negative effects around the day of firms' breach announcements. This also seems to hold for firms listed on the Japanese stock market [Ishiguro et al., 2006], where the effects are observable 10 days after disclosures – which is a result requiring further investigation. Cavusoglu et al. [2004b] additionally find that breach disclosures to the public come along with significant positive effects on market values of firms from the security industry.

Different study periods may partly explain these mixed results. The authors of [Gatzlaff and McCullough, 2010; Gordon et al., 2011] investigate effects associated with breach announcements of firms listed on the US stock market, thereby respecting different time spans. Gordon et al. [2011] find evidence that all kinds of breach disclosures in the period between 2002–2007 led to less devastating negative effects on firms' market values as compared to disclosures prior to 9/11/2001. They argue that this may be due to investors' perception of an increase in firms' efficiency to recover from successful attacks, or because of consumers' improved tolerance for breach disclosures. These results are contradicted by the findings of Gatzlaff and McCullough [2010], who examine the effects of firms' privacy breach announcements in particular. These authors identify that the negative effects on firms' market values were more significant and disastrous for breach disclosures observed in recent periods prior to the year 2006. An explanation can be the introduction of US breach notification laws during this time, which may have reinforced defenders' perceived cyber risk. With respect to all studies in Table 4.1, it seems that the negative short-term effects of breach disclosures got weaker over the years.

A differentiation between firms' breach types may explain some variance. The authors of [Cavusoglu et al., 2004b; Kannan et al., 2007] do not find that disclosures of specific types of breaches to the public lead to significant negative effects on firms' market values. Other scholars [Campbell et al., 2003; Ishiguro et al., 2006] report that announcements of confidentiality breaches result in statistically significant negative effects. Yet, such breaches only seem to have little long-term impacts, as observed by Ko and Dorantes [2006] based on a matched-sample comparison using firms' performance data. The studies in [Acquisti et al., 2006; Gatzlaff and McCullough, 2010; Gay, 2016] investigate a specific type of confidentiality breaches: those affecting defenders' privacy. Their results indicate that privacy breach announcements are associated with statistically

significant negative market value effects. Gatzlaff and McCullough [2010] additionally discover that firms which are less forthcoming about breach details experience more severe negative effects. As a countermeasure, some firms apparently bundle privacy breach disclosures with positive news reports [Gay, 2016]. The work of Gordon et al. [2011] complements pervious studies. They discover that neither the negative effects of confidentiality nor integrity breach disclosures are as devastating as those associated with availability breach announcements. Yet, respecting all studies in Table 4.1, confidentiality breach disclosures seem to have the strongest negative effects on firms.

Firms' size and type are other parameters that may explain the different effects of breach announcements on market values. Kannan et al. [2007] do not find that the size of firms has a significant impact on effects associated with their breach disclosures. Contrarily, the authors of [Cavusoglu et al., 2004b; Gatzlaff and McCullough, 2010] observe that smaller firms' disclosures are penalized stronger by the market than those of larger firms. Cavusoglu et al. [2004b] additionally detect that Internet firms, with a business model relying on e-commerce activity, suffer more from breach announcements than others. Taking also the work of Gordon et al. [2010] into consideration, it seems that Internet firms' market values are particularly sensitive to information disclosures.

In summary, the reviewed literature provides empirical support for the theoretical prediction that firms have natural disincentives to publicly disclose breach information. Nevertheless, many breaches get announced, presumably because firms are affected by enacted breach notification laws, reviewed in Section 5.1. Overall, breach disclosures have significant negative short-term effects on firms' stock market values. By contrast, little but speculation is known about the long-term effects of breach announcements.

4.3.4 General information sharing of firms with actors

The works reviewed in this section shed light into short- and long-term effects associated with firms' disclosure of *cyber risk information* in annual reports with the SEC on their stock market values [Gordon et al., 2010; Wang et al., 2013].

Gordon et al. [2010] investigate the long-term consequences of firms' cyber risk information disclosures on their market values. By making use of the value relevance methodology, the authors reveal that such disclosures come along with significant positive market value effects in the long-term. In particular, firms' announcements of implemented preventive controls, signaling engagement in cyber risk mitigation, lead to these positive effects. Yet, it seems that firms of different industries experience diverse aftermath from disclosures. Internet firms appear to benefit notably. Contrarily, firms in the financial sector do not experience significant benefits. The latter finding is

attributed to existing stringent cyber risk management regulations in this sector: firms affected by these regulations cannot strategically differentiate themselves over security measures. This line of thought is related to those presented by Carr [2003], arguing that ICT systems loose their strategic importance once they become a commodity.

Contrarily to Gordon et al. [2010], Wang et al. [2013] investigate short-term effects of firms' cyber risk information disclosures in annual reports with the SEC on their market values. Therefore, they utilize the event study methodology. Their results indicate that firms' disclosures do not lead to significant short-term effects. But there is an associations of disclosures and future breach announcements. Specifically, the authors find early evidence that if the textual content of disclosures comprise security investment themes, then firms are less likely to announce breaches in the future.

In summary, the reviewed literature documents empirical evidence that firms' cyber risk information announcements lead to significant and positive long-term and no significant short-term market value effects. The positive long-term effects may incentivize firms' information disclosures in the first place. Especially Internet firms appear to benefit from disclosures in the long run, such that they presumably have stronger incentives to announce their cyber risk information than other types of firms.

4.4 Trends and research directions

Our previous reviews shed light into the incentives of defenders to voluntarily publish cyber risk information, and highlight barriers that may hinder socially desirable disclosure strategies. We observe that the predominant enabler for information disclosures is expected cyber risk reduction, while firms' announcements in particular seem to be driven by the prospect of higher profit. The prevalent barrier for announcements is an expected increase in cyber risk, because publicly disseminated information may support attackers. Indeed, our review indicates that voluntary disclosures can substantially increase defenders' cyber risk, such that further investigations on the economics of this type of cyber risk information sharing are desirable. This motivates us to subsequently evaluate trends in the reviewed literature and distill future research directions. To this end, we use a systematization of works examined in this chapter that is depicted in Table 4.2. Our focus is on works inquiring defenders' announcements of cyber risk information in Section 4.4.1. Thereafter, we consider literature on firms' information disclosures to the public in Section 4.4.2.

Table 4.2 Reviewed literature on voluntary public cyber risk information sharing.

Information	... with all actors	
Actor type	Theoretical works	Empirical works
Attack information sharing of ...		
defenders ...	Nizovtsev and Thursby [2007]	Arora et al. [2006b]
	Cavusoglu et al. [2007]	Telang and Wattal [2007]
		Frei et al. [2010]
		Arora et al. [2010b]
		Moore and Clayton [2011]
		Ransbotham and Mitra [2013]
		Tang et al. [2013]
firms ...		Gordon et al. [2010]
		Wang et al. [2013]
Control information sharing of ...		
firms ...	Gal-Or and Ghose [2005]	Rescorla [2003]
	Arora et al. [2006a]	Moores [2005]
	August and Tunca [2008]	Arora et al. [2006b]
	Choi et al. [2010]	Arora et al. [2010a]
		Gordon et al. [2010]
		Edelman [2011]
		Wang et al. [2013]
		Durumeric et al. [2014]
Impact information sharing of ...		
firms ...	Gal-Or and Ghose [2005]	Campbell et al. [2003]
	Hausken [2007]	Hovav and D'Arcy [2004]
	Naghizadeh and Liu [2016]	Cavusoglu et al. [2004b]
		Acquisti et al. [2006]
		Ko and Dorantes [2006]
		Ishiguro et al. [2006]
		Kannan et al. [2007]
		Gatzlaff and McCullough [2010]
		Gordon et al. [2010]
		Gordon et al. [2011]
		Wang et al. [2013]
		Gay [2016]

4.4.1 Information sharing of defenders with actors

While theoretical work on the economics associated with defenders' disclosures of *vulnerabilities* to the public appears to be extensive, some inherent assumptions require further empirical underpinning. For instance, there are no studies that measure consumer costs associated with such disclosures. However, it is probable that (the threat of) announcements lead to substantial costs [Ransbotham and Mitra, 2013]. The reason is that publicly disclosed vulnerabilities are likely to get exploited fast and thus require an urgent response by affected consumers. In order to identify disclosed vulnerabilities, consumers need regularly check for news. It is for instance conceivable that they implement costly software that systematically gathers information on their used ICT system versions, and then automatically checks these versions regarding new vulnerabilities and patches by querying publicly available databases. Once consumers identify a disclosed vulnerability, different types of costs may arise depending on whether a corresponding patch gets available simultaneously. If a new patch is freshly released, the costs for its deployment are particularly high [Beattie et al., 2002]. And if no patch is available, costs may arise as routine processes no longer apply such that consumers respond in a haste to manage their cyber risk. It is reasonable but not easy for scholars to measure all of these costs at consumers, feeding the debate whether disclosures of vulnerabilities to the public are socially desirable or not.

Compared to the number of studies on defenders‘ disclosures of vulnerabilities to the public, only few works empirically investigate effects of *abuse information* announcements. The reviewed works indicate that such announcements are beneficial to society, as they create incentives for affected vendors to take mitigating actions. However, there are plenty of reasons for defenders to abstain from abuse information disclosures [Jhaveri et al., 2017]. For instance, the disclosure of information on abused ICT systems can trigger additional malicious activity: it hints attackers to particularly weak targets [Moore and Clayton, 2011]. More research that quantifies the extend to which attackers take advantage of public information on abused ICT systems is much needed and seems feasible, e. g., by monitoring these ICT systems' activities over time.

Another finding from our review is that the economics of defenders' *control and impact information* disclosures to the public have not been studied until now, as indicated by the missing categories in Table 4.2. A possible reason is the absence of institutionalized ways for defenders to announce corresponding cyber risk information. Consequently, as a first step to close this research gap, it is conceivable for scholars to come up with theoretical studies that shed light into the opportunities of institutionalized control and impact information announcements.

4.4.2 Information sharing of firms with actors

Table 4.2 also shows that there is a shortage of theoretical works that investigate the economics of firms' *attack information* announcements to the public, and accompanying empirical validations. An evident opportunity for research on this topic is to examine the effects associated with firms' public disclosures of acknowledged security threats in reportings to the SEC [Wang et al., 2013] on their own or the cyber risk of others. Yet, it supposedly will be hard to empirically quantify these effects, as firms may only have few natural incentives to publish their attack information [Gordon et al., 2010].

Contrarily, plenty of works investigate the economics associated with vendors' *patch* releases. However, this topic provides manifold future research opportunities, motivated by the evolution of the *Internet of Things* (IoT). IoT products are everyday objects embedded with software and network connectivity to enhance the consumer experience. A new risk is that vendors with little security expertise start to sell these networked devices. Thus, their products are often vulnerable, and not always patchable [Bertino et al., 2016]. A lack of vendors' expertise also leads to the issue that patches may not be provided at all. This motivates theoretical inquiries into how the market can fix patching strategies of IoT vendors, and if some government intervention is required.

The review additionally shows that there is relevant theoretical work investigating the incentives of firms to signal *product security quality*, contrasted by little empirical validations. Future empirical studies need investigate to what extent consumers act (approximately) rational when processing product security quality signals by vendors, as regularly assumed in the theoretical literature. In particular, vendors might be interested in which consumer segments are rational. Empirical studies on this topic may be inspired by anecdotal investigations, e. g., the study of Moores [2005].

Finally, there are several theoretical studies that investigate the economics of firms' *breach information* disclosures to the public, which lack some empirical validations. All theoretical studies assume that announcements of successful attacks result in negative long-term effects for firms. However, most empirical works in this context only provide rather noisy estimates for negative short-term effects that apply to firms listed on stock exchanges. Better estimates in the future might be inspired by the few notable exceptions in the literature [Ko and Dorantes, 2006; Kwon and Johnson, 2015], measuring long-term effects associated with breach disclosures of firms that do not have to be listed on stock markets. Overall, empirical studies that report concise and up-to-date estimates for short- and long-term effects of firms' breach announcements to the public are much needed, and appear to be practically feasible by scholars.

Chapter 5

Mandatory cyber risk information sharing

The reviews in the last two chapters indicate that defenders may often engage less in voluntary cyber risk information sharing than is socially desirable. This means that additional information sharing would lead to costs at defenders falling below associated economic benefits: shared information may help others to protect their systems and thus reduce negative externalities. By implication, negative externalities generated from unprotected ICT systems may justify government intervention in form of laws that oblige defenders to exchange information. Subsequently, we use our framework in Chapter 2 to examine the economics of mandatory cyber risk information sharing. We provide a context for laws that mandate defenders' information exchange in Section 5.1. Thereafter, we inquire the academic literature investigating corresponding laws. Specifically, Sections 5.2 covers theoretical studies and Section 5.3 empirical works. We distill trends in the literature and identify new research directions in Section 5.4.

5.1 Context

In this section, we review laws that mandate defenders' cyber risk information sharing, implemented or discussed in the EU and US – deeming this a representative sample for particularly impactful laws just like Hiller and Russell [2013] –, along with their mechanisms to incentivize compliance. We introduce EU laws in Section 5.1.1, and present US laws in Sections 5.1.2. Table 5.1 summarizes the key characteristics of selected EU and US notification laws. This table indicates that most disclosure regimes stipulate *breach information* sharing, which leads to *disclosure costs* at affected firms:

Table 5.1 Characteristics of selected EU and US breach notification laws.

Region	Law	Obliged	Report	Address	Objective	Effect
EU	Telecoms Package	Firms in the telecoms sector	SB&PB	A or A&I	IP&S or IP&S&R	Sa or Sa&D
EU	Directive 2015/2366	Payment service providers	SB&PB	A or A&I	IP&S or IP&S&R	Sa or Sa&D
EU	Regulation 2016/679	Data controllers and processors	PB	A or A&I	IP&S or IP&S&R	Sa or Sa&D
EU	Directive 2016/1148	Market operators	SB&PB	A	IP&S or IP&S&R	Sa or Sa&D
US	State Laws	Firms controlling personal data	PB	I or A&I	IP&R	D or Sa&D
US	HIPAA & HITECH	Firms in the health care sector	PB	A&I	IP&R	Sa&D
US	GLBA	Firms in the financial sector	PB	I or A&I	IP&R	Sa&D

SB	Security breaches	IP	Incentivize firms to take precautions
PB	Privacy breaches	S	Draw and share conclusions with others defenders
A	Authorities	R	Improve rights of affected individuals
I	Affected individuals	Sa	Sanctions
		D	Disclosure costs

expenses resulting from bureaucratic burdens, such as obligations to document and exchange breach information, and secondary losses as shared information likely become public eventually (cf. Sections 4.3.3 and 2.3). The firms affected by laws can reduce these disclosure costs if they prevent successful attacks a priori, and thereby evade reporting obligations. Indeed, from a study of legal texts and official justifications we conclude that *the main objective of notification laws is to improve social welfare by incentivizing affected firms to internalize negative externalities from their ICT systems (without setting out details on how cyber risk must be managed), besides promoting cyber risk information symmetry in the economy* (cf. also [Romanosky and Acquisti, 2009]).

5.1.1 Situation in the EU

In the EU, union laws (i. e., directives that have to be transpose into national law, and regulations which are are directly binding) and national laws of member states predominantly mandate firms to privately share breach information with national "competent authorities" in the first instance. The declared objective is to establish an economy-wide

transparency on breaches [Dekker et al., 2012]: informed authorities may be able to effectively draw and share conclusions from reported breach information with other actors in the economy, e. g., privately disseminate advise how to protect against attacks or detect breaches, publicly disclose breaches to warn regarding propagating attacks, or release guidelines that help defenders with breached ICT systems to minimize impacts. Enacted union laws that mandate breach reporting mainly affect the telecoms sector. Most of them were introduced with the "Telecoms Package" in 2009. A prominent example is Directive 2009/136/EC, amending Directive 2002/58/EC. It has the objective to protect the privacy of individuals' data handled by electronic communications service providers. The directive obliges providers to report occurred breaches to authorities *and*, under specific circumstances, also notify affected individuals. Besides such union law, some EU member states have enacted national notification laws. A non-exhaustive list of these laws is reported in [ENISA, 2015]. Our review indicates that almost all national and union breach notification laws use sanctions to incentivize affected firms' compliance with reporting obligations despite associated disclosure costs.

Three complementary union laws soon expand existing regulations and directives:

- Directive (EU) 2015/2366, referred to as the "Revised Directive on Payment Services" (PSD2). Its declared objective is to protect defenders when they pay online. It requires "payment service providers" to report breaches to authorities *and*, under some circumstances, payment service users. Authorities may then forward information to other institutions, promoting the protection of payment systems in the economy. The Directive (EU) 2015/2366 has entered into force in November 2015, and EU member states have two years to transpose it into national law.

- Regulation (EU) 2016/679, referred to as the "General Data Protection Regulation" (GDPR), accompanied with Directive (EU) 2016/680. This regulation aims to harmonize and unify existing EU privacy breach reporting obligations. It requires "data controllers and processors" in the EU to report privacy breaches to authorities *and sometimes* the affected individuals, e. g., when a breach is likely to violate rights and freedoms. (The GDPR will also apply to firms based outside the EU who process personal data of Europeans.) Authorities may then promote information symmetry in the economy. Regulation (EU) 2016/679 entered into force in May 2016, and repeals Directive 95/46/EC in May 2018.

- Directive (EU) 2016/1148, referred to as the "Network and Information Security" (NIS) Directive. Its declared objective is to establish a high level of ICT system security in the EU. Therefore, selected "operators of essential services

and digital service providers" (we use "firms" as shorthand for this legal term) will have to report breaches to authorities *only*. In turn, the authorities can advise other defenders in the economy. The NIS Directive has entered into force in August 2016. EU member states have 21 month to transpose the directive into national law, and an additional 6 month to identify the firms affected.

In the tradition of other union breach notification laws, the PSD2, GDPR, and the NIS Directive provide for sanctions to incentivize firms' compliance. For instance, a German initiative anticipating the NIS Directive imposes sanctions of up to 100 000 € for firms who fail to report breaches of their ICT systems [Deutscher Bundestag, 2015].

5.1.2 Situation in the US

By contrast to laws in the EU, in the US, state and federal laws mandate firms to share breach information with also affected individuals in the first instance, e. g., by the use of notification letters [Bisogni, 2016]. The first implemented state breach notification law was the California Civil Code Section § 1798.29. It obliges private and public firms conducting business in California to report breaches of personal data to affected individuals. Additionally, the law stipulates breach reporting to authorities if more than 500 of a firm's data records are affected. The intention of this law is twofold: first, informing individuals about privacy breaches enables them to take mitigating actions [Romanosky et al., 2010]; and, second, the law incentivizes firms to encrypt personal data, as only breaches of unencrypted records have to be reported. From the start, the Californian law led to a high number of privacy breach reports [Romanosky et al., 2011]. Because of this success, other US states enacted similar laws [Samuelson Law, Technology & Public Policy Clinic, 2007].[1] Besides these state laws, there are two prominent federal breach notification laws in the US. They are formalized in the "Health Insurance Portability and Accountability Act" (HIPAA) – amended by the "Health Information Technology for Economic and Clinical Health Act" (HITECH) – and the "Gramm–Leach–Bliley Act" (GLBA), respectively. The HIPAA mandates firms in the health care sector to report breaches of health information to affected individuals, the Department of Health and Human Services and, under some circumstances, the media. The GLBA differs from the HIPAA as it obliges firms in the financial sector to inform their primary federal regulator on privacy breaches, and in some cases notify affected individuals. Our review indicates that most of the state and federal

[1] A list of all active US state breach notification laws can be accessed on the website of the National Conference of State Legislatures (NCSL) (http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx).

breach notification laws provide for sanctions in cases of violations to incentivize firms' compliance with reporting obligations despite associated disclosure costs.

The 114th Congress has introduced new federal legislation on defenders' cyber risk information sharing. The "House of Representatives" (H.R.) proposed "H.R.1770 – Data Security and Breach Notification Act of 2015," which intends to replace the existing patchwork of state breach notification laws. Also, former US President Barack Obama submitted an "Executive Order 13691, Promoting Private Sector Cybersecurity Information Sharing" with the objective to improve cyber risk information sharing within the private sector and between the private sector and the government in 2015. This shall pave the way for new legislation on information sharing, formalized in different bills: the H.R. introduced "H.R.234 – Cyber Intelligence Sharing and Protection Act," and passed "H.R.1560 – Protecting Cyber Networks Act" in the Congress; the Senate (S.) submitted "S.456 – Cyber Threat Sharing Act of 2015," and passed "S.754 – Cybersecurity Information Sharing Act." Most of this legislation limits the liability of firms if they do not only share their private cyber risk information with each other, but also an authority, e. g., the Department of Homeland Security (DHS).

5.2 Theoretical literature

The economics of previously reviewed laws that mandate firms' cyber risk information sharing with authorities or individuals should be rigorously analyzed by scholars to inform regulators regarding effects of particular approaches on social welfare. However, we only observe two theoretical studies that shed some light into the economics of these laws. The study reviewed in Section 5.2.1 provides first insights into effects of firms' mandatory (private) cyber risk information sharing with authorities. By contrast, the theoretical work that we examine in Section 5.2.2 inquires the economics of firms' mandatory (public) privacy breach information sharing with affected individuals.

5.2.1 Mandatory information sharing with authorities

We are not aware of any literature that theoretically analyzes the economics of firms' mandatory *cyber risk information* sharing with authorities in detail. However, the model presented in [Öğüt et al., 2005] captures effects associated with such information exchange. This model maps into our unified formal model proposed in Section 2.4

Öğüt et al. [2005] present a model to investigate – besides other things – effects of two symmetric firms' security investments and cyber risk information sharing with an

authority in the presence of interdependent ICT system security. The firms represent an economy and make use of ICT systems that are exposed to opportunistic attacks. A successful attack at either firm affects both of them, if ICT system security is interdependent $\gamma \in [0, 1]$. To counter attacks, each firm can invest in security and engage in cyber risk information sharing with the authority. Investment increases a firm's ICT system security level and may thus reduce economy-wide cyber risk, but is costly C. By contrast, a firm's information exchange with the authority is assumed to be costless, and also potentially reduces economy-wide cyber risk: the authority may effectively draw conclusions from received information $s_i, s_{1-i} > 0$, and share them with other defenders to support their cyber risk management efforts. The authors propose two modeling approaches that capture associated positive effects:

- *Reduced breach probability.* Firms can use the information from the authority to enhance the effectiveness (or efficiency) of their security investments $x_i, x_{1-i} > 0$. This leads to lower breach probabilities, i. e., $\partial P_i / \partial s_{1-i} < 0$ and $\partial P_{1-i} / \partial s_i < 0$, ceteris paribus. With respect to this assumption, the cyber risk of the symmetric firms can be denoted as

$$R_i(x_i, x_{1-i}, s_i, s_{1-i}) = (1 - (1 - P_i(x_i, s_{1-i})) \cdot (1 - \gamma \cdot P_{1-i}(x_{1-i}, s_i))) \cdot I_i \ . \quad (5.1)$$

- *Reduced interdependence.* Firms can use the information from the authority to better protect from propagating attacks. This leads to a reduction in interdependence $\partial \gamma / \partial s_{1-i} < 0$ of their ICT system security, ceteris paribus. With respect to this assumption, the cyber risk of the symmetric firms can be denoted as

$$R_i(x_i, x_{1-i}, s_{1-i}) = (1 - (1 - P_i(x_i)) \cdot (1 - \gamma(s_{1-i}) \cdot P_{1-i}(x_{1-i}))) \cdot I_i \ . \quad (5.2)$$

The analyses by Öğüt et al. [2005] of the above model setups suggest that in general, interdependence reduces incentives at firms to invest in ICT system security. This leads to overall suboptimal ICT system security levels, and thus more cyber risk in the economy. However, cyber risk information exchange between firms and the authority can oppose the negative effect of interdependence. The reason is that the authority may be able to draw conclusions from reported information, and effectively share them with other defenders. In turn, the defenders can use received information to better manage their cyber risk, which leads to positive externalities in the economy.

In summary, the review in this section indicates that firms' mandatory cyber risk information sharing with authorities potentially creates a positive effect on welfare.

5.2.2 Mandatory information sharing with individuals

To the best of our knowledge, the authors of [Romanosky et al., 2010] are the only ones that theoretically analyze the economics of firms' mandatory cyber risk information sharing with individuals. Specifically, their theoretical model inquires conditions under which mandating firms' *privacy breach* reporting to also affected customers is social welfare improving. This model can be mapped into our formal model in Section 2.4.

The model of Romanosky et al. [2010] assumes one (representative) firm that uses an ICT system to store personal data belonging to its only (representative) customer – which is an individual. This ICT system is affected by opportunistic attacks. The firm can invest in security to mitigate associated cyber risk, but this is costly C. Furthermore, regardless of the firm's security investment, privacy breaches may occur [Maillart and Sornette, 2010]. These breaches result in impact at the firm I_i, and impose a negative externality on its customer I_{1-i}. The customer can limit impact on his own (self-protection), but only if the firm informs him about the breach. Yet, this only happens if the firm is affected by an enacted privacy breach notification law:

- *No breach notification law.* Without a law, the firm abstains from information sharing. This leads to the following risk at the firm and customer, respectively:

$$R_i(x_i) = P_i(x_i) \cdot I_i \text{ , and} \tag{5.3}$$
$$R_{1-i}(x_i) = P_i(x_i) \cdot I_{1-i} \text{ .} \tag{5.4}$$

- *Breach notification law.* A notification law obliges the firm to notify the customer in the event of a privacy breach. Such notifications entail disclosure costs for the firm $I_i(s_i) \geq 0$ (here assumed to exclude losses due to liability claims), but enable the customer to self-protect. Thereby, the customer may under- or overreact to privacy breach notifications. His cost function $I_{1-i}(x_{1-i})$ captures that the marginal costs of own actions increase, while marginal benefits decrease, i. e., $\partial^2 I_{1-i}/\partial x_{1-i}^2 > 0$ and $\lim_{x_{1-i}\to\infty} I_{1-i}(x_{1-i}) = \infty$. Besides, breach notifications allow the customer to make the firm liable for some portion $\gamma \in [0, 1]$ of his impact. This leads to the following cyber risk at the firm and customer, respectively:

$$R_i(x_i, x_{1-i}, s_i) = P_i(x_i) \cdot [I_i + I_i(s_i) + \gamma \cdot I_{1-i}(x_{1-i})] \text{ , and} \tag{5.5}$$
$$R_{1-i}(x_i, x_{1-i}) = P_i(x_i) \cdot (1-\gamma) \cdot I_{1-i}(x_{1-i}) \text{ .} \tag{5.6}$$

The analysis of this model suggests that an enacted privacy breach notification law can be social welfare improving, as it leads to two effects. First, it creates incentives for firms to invest more in security as compared to the scenario without a law, because they want to mitigate cyber risk due to disclosure costs and liability claims. Second, breach notifications give customers a chance to self-protect. Overall, Romanosky et al. [2010] find that a breach notification law always leads to social benefits if affected firms' disclosure costs are smaller or equal to the benefits from customers' self-protection.

In summary, our review indicates that enacted laws mandating firms' privacy breach information sharing with individuals may lead to higher social welfare as compared to the scenario without enacted privacy breach notification laws.

5.3 Empirical literature

This section covers empirical studies [Ablon et al., 2016; Bisogni et al., 2017; Choi and Johnson, 2017; Kwon and Johnson, 2015; Romanosky et al., 2011] investigating effects of firms' mandatory *breach information* sharing with individuals. Bisogni et al. [2017] shed light into determinants that make corresponding laws ineffective, while all other works evaluate effects of effective laws on defenders' adaption of security practices.

Bisogni et al. [2017] try to infer from disclosed breach information the number of successful attacks at US firms that are affected by notification laws. They reveal that a high amount of all breaches announced in the US is to be attributed to firms hosted by only four US states, indicating economy-wide underreporting of successful attacks despite existing notification laws. To inquire this underreporting, the authors conduct regression analyses on data from the ITRC and firms' notification letters, with respect to elements of laws. They identify provisions in laws that hinder breach notifications to become public, and reveal which exemptions in laws may disincentivize firms to send out notifications in the first place. Also, an analysis of notifications' contents and rates (number of reported breaches per state-sector, divided by the number of firms in the state-sector) provides early evidence that a high amount of breaches at firms stays undetected. Overall, the results indicate that more than 46 % of all successful attacks at US firms who are affected by breach notification laws remain unknown to the public.

Romanosky et al. [2011] analyze whether US breach notification laws reduce identity theft. To this end, the authors use a state and time fixed effect regression analysis on publicly available identity theft panel data covering the period in between 2002–2009. Their analysis reveals that the adoption of notification laws reduced the number of identity thefts in the US by – on average – about 6.1%. This percentage is an equivalent

to about 800 consumer records lost per breach at a US firm. The reduction in identity theft is explained by notification laws incentivizing affected firms to invest in security, and enabling informed individuals to take mitigating action. But even though identity theft appears to get reduced, welfare does not necessarily improve as notification laws come along with diverse costs for affected defenders [Romanosky and Acquisti, 2009].

The authors of [Ablon et al., 2016; Choi and Johnson, 2017; Kwon and Johnson, 2015] study effects of privacy breach disclosures due to corresponding notification laws on defenders. Ablon et al. [2016] survey US consumers, investigating their response to breach notifications received from firms. Of consumers who received a notification, most accepted offers for assistance provided by firms to evade getting sued, e. g., identity theft or credit monitoring services [Romanosky et al., 2014]. Still, 11 % of the respondents stopped dealing with breached firms altogether. This goes in line with results of Kwon and Johnson [2015], who analyze data about hospitals and their disclosed breaches by means of a propensity score matching technique. The findings of this analysis indicate that hospitals' announcements lead to significant decreases of outpatient visits and admissions in the long run. In a related study, Choi and Johnson [2017] examine the association between hospitals' breach disclosures and mortality rates by means of a multivariate regression model. The authors find early evidence for an increase in mortality rates at hospitals during the days after breach announcements, but it remains an open question if this increase is due to direct or indirect (resources diverted from patient care to notification obligations) effects of the successful attacks.

In summary, the reviewed works provide evidence that enacted breach notification laws not always incentivize affected firms' disclosure of successful attacks, but still may come along with positive economic effects. In particular, some provisions and exemptions in these laws seem to reduce affected firms' incentives to announce successful attacks. These formalities essentially give firms a chance to evade disclosure costs. However, there is empirical evidence that enacted laws reduce identity theft in economies. This indicates that the laws achieve their objectives at least to some extend.

5.4 Trends and research directions

Table 5.2 systematizes the works on mandatory cyber risk information sharing reviewed in this chapter, which are all motivated by legal requirements introduced in Section 5.1. These requirements almost exclusively oblige firms to share their impact information with other defenders. Thus, this trend in the reviewed literature is unsurprising. Next, we evaluate additional trends and motivate new research directions. We first discuss

Table 5.2 Reviewed literature on mandatory cyber risk information sharing.

Information	... with authorities	... with individuals	
Actor type	Theoretical works	Theoretical works	Empirical works
Attack information sharing of ...			
firms ...	Öğüt et al. [2005]		
Control information sharing of ...			
firms ...	Öğüt et al. [2005]		
Impact information sharing of ...			
firms ...	Öğüt et al. [2005]	Romanosky et al. [2010]	Romanosky et al. [2011] Kwon and Johnson [2015] Ablon et al. [2016] Choi and Johnson [2017] Bisogni et al. [2017]

works on mandatory information sharing with individuals in Section 5.4.1. Then, we examine literature on mandatory information sharing with authorities in Section 5.4.2.

5.4.1 Mandatory information sharing with individuals

A considerable amount of studies reviewed in this chapter inquire laws that mandate firms' *breach information* sharing with individuals. Specifically, we observe extensive empirical work on this topic, contrasted by only one corresponding theoretical study. The majority of empirical works base their analysis on publicly available breach information, originating from firms affected by enacted laws: if these firms notify individuals, this enables the latter to forward received information to the press. Against the backdrop that such actions lead to secondary losses at firms (cf. Section 4.3.3), it is surprising that no empirical study investigates which type of law enforcement mechanism best incentivizes compliance with reporting obligations. As a first step to close this research gap, it is conceivable that scholars quantify the effect of sanction levels suggested in different laws on firms' engagement in breach information sharing despite associated disclosure costs. In the same context, the one reviewed theoretical study by Romanosky et al. [2010] obtains its findings based on the unrealistic assumption that once breach notification laws are enacted, affected firms comply despite associated disclosure costs. Consequently, the authors disregard that regulators may require enforcement mechanism to incentivize firms' compliance. In turn, it is reasonable that followup studies examine effects of law enforcement mechanism on firms' incentives to share their private breach information with individuals, and social welfare by extension.

5.4.2 Mandatory information sharing with authorities

We observe a high discrepancy between the number of studies on firms' mandatory breach information sharing with individuals, and the number of inquiries into firms' mandatory *cyber risk information* exchange with authorities. In fact, while the former topic is rather extensively examined, there is no empirical and only one theoretical investigation on the latter topic. And this theoretical study by Öğüt et al. [2005] does not analyze the economics of corresponding disclosure regimes, but only provides first predictions on the effects associated with information sharing activities of involved parties. This research gap is striking against the backdrop that EU regulators regularly enact laws that mandate firms' breach information sharing with authorities, as pointed out in Section 5.1. In the rest of this dissertation, we aim to make a first step towards closing the identified research gap by theoretically studying the economics of mandating firms' breach information sharing with authorities.

Part II

New results

Chapter 6

Mandatory breach information sharing with authorities

The last chapter indicates a research gap regarding studies on the economics of firms' mandatory breach information sharing with authorities. In particular, there is no work that identifies conditions under which an effective enforcement of corresponding laws improves welfare. We outline effects of enforcing laws such as the NIS Directive in Section 6.1, motivating the research question in this chapter. To tackle this question, we devise a game-theoretic model in Section 6.2. We derive and analyze the model's social optima and Nash equilibria in Section 6.3. Finally, we present inferences from our analysis, discuss possible implications, and highlight model limitations in Section 6.4.

6.1 Motivation

The effectiveness of laws that mandate firms' ICT system breach information sharing with authorities depends on the configuration of enforcement mechanism. According to Winn [2009], effective breach notification laws likely require some form of direct regulation providing for ex post public enforcement – the use of mechanism to detect and penalize violations of stipulated rules [Polinsky and Shavell, 2007]. The reason is that such mechanism can overcome firms' disincentives to share breach information, which root in *disclosure costs*: expenses emerging from bureaucratic burdens, such as obligations to document and report breaches, and secondary losses due to authorities' disclosure or leak of received breach information to the public. Against this background, many of the laws reviewed in Section 5.1 might be ineffective as they do not directly regulate how affected firms' compliance with reporting obligations is verified. But

national implementations of the NIS Directive may change the situation. This directive mentions security audits to detect and sanction non-compliance. However, many practical questions still remain open. Our working hypothesis in this dissertation is that firms undergo spontaneous security audits of their ICT systems, initiated by regulators. In case that auditors discover unreported breaches, they inform authorities, and sanctions get imposed. By implication, if the expected sanctions for non-compliance with reporting obligations are high enough, firms will rather share than conceal private information on detected breaches with authorities in spite of disclosure costs. Thus, regulators may effectively enforce laws that mandate firms' breach information sharing with authorities by configuring the probability of security audits and the sanction level.

An effective enforcement of laws that mandate breach information sharing with authorities likely incentivizes affected firms to invest in security despite associated costs. The laws may incentivize firms to invest in the prevention of ICT system breaches, as this can decrease costs associated with the disclosure regime. Effective investment also leads to positive externalities, if the *security of ICT systems is interdependent*. However, regardless of preventive measures, firms' systems can get breached. This may incentivize firms affected by laws to invest in breach detection, required to meet reporting obligations and thus comply. Yet, firms are not able to detect all breaches of their systems, as detective controls are notoriously inaccurate [Cavusoglu et al., 2004a]: these controls can lead to alerts in cases where there is no breach; or absence of alerts although a successful attack has taken place. Therefore, even if security investments are very high, firms have to expect sanctions in case that regulators initiate audits.

An effective enforcement of laws that mandate firms' breach information sharing with authorities can benefit all defenders in economies, if the authorities promote information symmetry. According to the NIS Directive, authorities have the task to *effectively draw and share conclusions from firms' reported information*: primarily, authorities may privately share conclusions with other firms to help them prevent or detect successful attacks; secondarily, authorities can disclose breaches or related information to warn the public. Overall, conclusions shared by authorities can support cyber risk management efforts of recipients, such that they may reach higher security levels at lower costs.

Authorities' disclosures to the public also alter societal attitudes, affecting firms indirectly via expected disclosure costs. For instance, it is conceivable that authorities disclose breach related cyber risk information to educate defenders about how to protect from attacks. In turn, society may grow tolerant of breaches, decreasing firms' expected secondary losses if successful attacks to their ICT systems become public. On the other hand, authorities can publicly disclose breaches to "name and shame" affected

firms. Consequently, society might develop less tolerance for successful attacks, which increases expected secondary losses if breaches to firms' ICT systems get announced.

Overall, smart regulators effectively enforce firms' mandatory breach information sharing with authorities by audits and sanctions if this reduces expected social costs due to cyber risk. However, there is a conflict of interest between regulators and affected firms, making enforcement decisions difficult. This conflict can be interpreted as a principal-agent problem with moral hazard [Laffont and Martimort, 2002]: a regulator (principal) enacts a law similar to the NIS Directive. Firms (agents) affected by the law need to detect breaches of their systems as a prerequisite for compliance with information sharing obligations, but only have few incentives to unilaterally report breaches because of disclosure costs. In response, the regulator may incentivize firms' compliance by initiating audits to find and sanction unreported breaches. Yet, audits cannot differentiate between firms' nescience and concealment of breaches. Consequently, it might be difficult for the regulator to decide on the sanction level that effectively enforces the notification law. This issue along with the objectives of breach notification laws outlined in Section 5.1 motivates the research question tackled in this chapter:

> Under which conditions does an effective enforcement of laws, such as the NIS Directive, by audits and sanctions lead to
>
> a) higher investments in preventive controls at affected firms, and
> b) lower social costs due to cyber risk – as a measure for social welfare?

An answer to this question is relevant for affected firms' security managers who decide on *investment in preventive controls* and *breach reporting*. Also, it is relevant for regulators who enforce laws with direct regulation in form of *audits* and *sanctions*.

6.2 Model

In this section, we devise a game-theoretic model to tackle the research question above. Subsequent subsections include one decision variable and free model parameter for properties highlighted in *italic* before. In Section 6.2.1, we introduce a model for firms' decisions to *investment in preventive controls*, and the parameter for *interdependence* of ICT system security. Then, in Section 6.2.2, we formalize firms' decisions to comply with *breach reporting*, and propose the parameter for the authority's *effectiveness in drawing and privately sharing conclusions from reports*. We formalize the regulator's decision on the *audit* probability and introduce the parameter for firms' *disclosure costs* in Section 6.2.3. Table 6.1 refines symbols used in this work for the subsequent model.

Table 6.1 Symbols: mandatory breach information sharing with authorities.

Symbol	Type	Description	Constraint
x	decision variable	investment in preventive controls	$x \geq 0$
s	decision variable	compliance with breach reporting	$s \in [0,1]$
a	decision variable	security audit probability	$a \in [0,1]$
γ	parameter	interdependence of ICT system security	$\gamma \in [0,1]$
α	parameter	private information sharing effectiveness	$\alpha \in [0,1]$
λ_2	parameter	disclosure costs	$\lambda_2 \geq 0$
λ_1	constant	primary losses (breach)	$\lambda_1 = 1$
σ	constant	sanction level	$\sigma = 1$
ϵ	constant	error rate of detective controls	$\epsilon = .2$
θ	constant	breach prevention productivity	$\theta = 20$
n	constant	number of firms	$n = 2$
O	(objective) function	expected costs due to cyber risk	
I	function	impact of a breach	
H	function	changes in interdependence	
R	function	cyber risk	
P	function	breach probability	
B	random variable	breach	
β	realization	realization of B	$\beta \in \{0,1\}$
$\hat{B}$	random variable	breach detection	
$\hat{\beta}$	realization	realization of $\hat{B}$	$\hat{\beta} \in \{0,1\}$
Ψ	random variable	security audit	
ψ	realization	realization of Ψ	$\psi \in \{0,1\}$
ξ	realization	breach information reporting	$\xi \in \{0,1\}$

6.2.1 Investment in preventive controls and interdependence

Consider for now a single rational firm belonging to a larger economy. This firm *decides on investment in preventive controls* $x \geq 0$, which can decrease the probability P of breaches to its ICT system. We model realizations of the random variable B (breach) as $\beta \in \{0,1\}$, and follow Gordon and Loeb [2002] by characterizing the relationship between breach probability and preventive measures as $Pr(\beta = 1) = P(x)$. With an increase in investment x, the breach probability decreases $\partial P/\partial x < 0$, but at a decreasing rate $\partial^2 P/\partial x^2 > 0$, i. e., $\lim_{x\to\infty} P(x) = 0$. According to Böhme [2012], a simple way to capture this relationship in a functional form is $P(x) = \theta^{-x}$. Therein, the constant θ represents the firm's breach prevention productivity, which we subsequently assume to be "moderate," i. e., $\theta = 20$. Furthermore, we assume that each attack on an unprotected ICT system $x = 0$ results in a breach and causes the primary losses $I = \lambda_1$. We fix the primary losses associated with a breach on $\lambda_1 = 1$ to normalize the monetary

scale. Thus, the firm's expected costs because of cyber risk to its system are given by

$$O(x) = R(x) + x \text{ , with} \tag{6.1}$$

$$R(x) = P(x) \cdot I \text{ .} \tag{6.2}$$

We generalize this setup to an economy with $n = 2$ symmetric, a priori homogenous and rational firms. Both firms $i \in \{0, 1\}$ choose an investment x_i, and a breach at either may affect the other. We can capture the latter by introducing a *parameter for interdependence* $\gamma \in [0, 1]$ of firms' system security, altering their breach probability to

$$P_i(x_i, x_{1-i}) = 1 - (1 - P(x_i)) \cdot (1 - \gamma \cdot P(x_{1-i})) \text{ .} \tag{6.3}$$

The intuition of Equation (6.3) is that firm i can only evade a loss if itself does not get breached and no attack propagates from firm $1 - i$, due to interdependence [Öğüt et al., 2005]. Without interdependence, i. e., $\gamma = 0$, we find that the breach probability at firm i only depends on its own investment and is given by $Pr(\beta_i = 1) = P_i(x_i, x_{1-i}) = P(x_i)$.

We acknowledge that both firms have a self-interest in detecting breaches of their ICT systems and denote the realization of the random variable $\hat{B}$ (breach detected) as $\hat{\beta}_i \in \{0, 1\}$. The success to detect a breach that occurred is exogenously given by the probability $Pr(\hat{\beta}_i = 1 | \beta_i = 1) = 1 - \epsilon$, where ϵ is the error rate of detective controls. We assume that, as an exemplary detective control, firms use IDS. Yet, we ignore potential costs of such systems to restrict the number of variables in our model. As a further simplification, we consider that the type I error rate of the firms' IDS is 0 %. A study of Lippmann et al. [2000] shows that the best IDS detect about 80 % of attacks that have happened. Thus, we may (optimistically) fix the type II error rate at $\epsilon = 20\,\%$.

6.2.2 Compliance with the law and authority's effectiveness

A breach notification law requires the firms to decide on breach reporting $\xi_i \in \{0, 1\}$ to the authority as soon as a successful attack happens and is detected. We indicate a firm's decision to report the information that no breach has been detected as $\xi_i = 0$. Accordingly, $\xi_i = 1$ indicates that a firm reports a detected breach. Therefore, *compliance with breach reporting obligations* is $Pr(\xi_i = 1 | \hat{\beta}_i = 1) = s_i$. For the sake of simplicity, we assume that nobody has an interest in reporting breaches that did not happen.

If a firm reports breach information, the authority can use them to draw conclusions and advise other defenders in the economy with the objective to decrease social costs. We denote the *parameter for the authority's effectiveness in drawing and privately*

sharing conclusions from reports with other firms by $\alpha \in [0, 1]$. Öğüt et al. [2005] propose that if recipients effectively use such conclusions, the positive effect of information sharing may be interpreted as a reduction in breach probability or interdependence (cf. Section 5.2.1). Subsequently, we assume that if an informed authority privately and effectively shares conclusions derived from breach reports with other firms in the economy, this reduces interdependence of ICT system security according to

$$H(s) = 1 - \alpha \cdot (1 - \epsilon) \cdot s \ , \text{ such that} \tag{6.4}$$

$$P_i(x_i, x_{1-i}, s_{1-i}) = 1 - (1 - P(x_i)) \cdot (1 - \gamma \cdot H(s_{1-i}) \cdot P(x_{1-i})) \ . \tag{6.5}$$

The function in Equation (6.4) is bound to the interval $0 \leq H(s) \leq 1$. Furthermore, it is monotonically decreasing in the effectiveness of the authority and in firms' compliance, i. e., $\partial H/\partial \alpha < 0$ and $\partial H/\partial s < 0$. Based on Equation (6.5), this reflects the intuition that a firm's breach information reporting to the authority can reduce others' interdependence: if the authority effectively uses a firm's breach information to draw and share conclusions, this helps the recipients to better protect from propagating attacks. However, a firm's compliance with reporting obligations entails disclosure costs.

6.2.3 Disclosure costs and security audits

If the regulator passes a breach notification law, firms do not only consider the primary losses of a successful attack, but also disclosure costs associated with breach reporting. Subsequently, let $\lambda_2 \in [0, \infty[$ denote the *parameter for a firm's disclosure costs*. As compliance $s_i = 1$ inevitably leads to these costs, a firm's impact in case of a breach is

$$I_i(s_i) = (1 - \epsilon) \cdot s_i \cdot \lambda_2 + \lambda_1 \ . \tag{6.6}$$

Disclosure costs entail a conflict of interest between the regulator who passes a breach notification law and affected firms, which can be interpreted as a principal-agent problem with moral hazard. The regulator (principal) enacts a breach notification law. But affected firms (agents) may only have few incentives to unilaterally report detected ICT system breaches due to disclosure costs. We assume that firms only report if this does not make them worse off than concealing breaches. (Thus, firms which are indifferent to compliance with the law act law-abiding. In economic terms, one could consider such firms as "marginal risk averse.") In order to overcome a potentially evolving moral hazard problem where firms do not comply with the notification law, the regulator can initiate security audits to detect and sanction non-reported breaches.

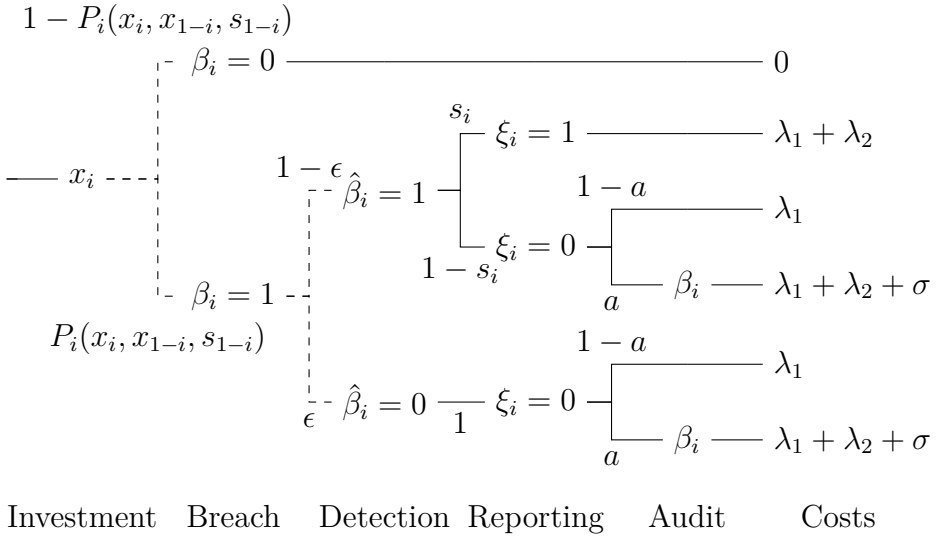


Fig. 6.1 Decisions of firm i, nature, and the regulator.

We model realizations of the random variable Ψ (security audit) as $\psi \in \{0, 1\}$. The regulator abstains from initiating security audits if a firm reports a breach, i. e., $Pr(\psi = 1|\xi_i = 1) = 0$. Else, he initiates audits with probability $Pr(\psi = 1|\xi_i = 0) = a$. We assume that realized security audits detect all breaches that have happened to an ICT system, i. e., audits are more reliable than detective controls per definition. For the sake of simplicity, we do not model the costs associated with audits. Rather, we assume that sanctions collected from non-complying firms fully compensate these costs.

The decision tree in Figure 6.1 summarizes the breach-related costs of firm i under a disclosure regime. It comprises all decisions of the firm and regulator. Dashed lines represent uncertainty because of nature's decisions. At first, the firm invests x_i in breach prevention. Then, an attack on its ICT system may take place. This attack is successful with probability $P_i(x_i, x_{1-i}, s_{1-i})$. We assume that, per period under consideration, there can at most be one breach to the firm's ICT system. A breach leads to primary losses λ_1, regardless of its detection. And every breach gets detected with probability $1 - \epsilon$. If firm i does not detect a breach, it will not report it to the authority. Else, the firm can strategically decide on reporting. In cases without breach reporting, the regulator initiates audits at random. If auditors detect an unreported breach, sanctions $\sigma \in [0, \infty[$ get imposed and the authority is notified – yet we assume that the authority does not derive and privately share conclusions from auditors' information with other firms. Once a firm's breach information are reported, they may become public due to the authority's actions. Thus, firm i expects disclosure costs λ_2.

From Figure 6.1, we can derive the expected costs due to cyber risk at firm i, conditional on the regulator's effective enforcement of a breach notification law, i. e.,

$$O_i(x_i, x_{1-i}, s_i, s_{1-i}, a) = R(x_i, x_{1-i}, s_i, s_{1-i}, a) + x_i \ , \text{ with} \tag{6.7}$$

$$R(x_i, x_{1-i}, s_i, s_{1-i}, a) = P_i(x_i, x_{1-i}, s_{1-i}) \cdot I_i(s_i, a) \ , \text{ and} \tag{6.8}$$

$$I_i(s_i, a) = (1-\epsilon) \cdot [s_i \cdot \ \lambda_2 + (1-s_i) \cdot a \cdot (\lambda_2 + \sigma)] + \epsilon \cdot a \cdot (\lambda_2 + \sigma) + \lambda_1 \ . \tag{6.9}$$

Observe from Equation (6.9) that if the regulator introduces infinitely high sanctions, and given a positive audit probability, firms always have incentives to report detected breaches. However, firms cannot find all breaches that have to be reported due to the error probability of their detective controls. Therefore, they will be burdened with sanctions eventually. In practice, this may lead to firms' bankruptcy, because unreasonably high sanctions are uncollectible. Therefore, an incentive mechanism with infinitely high sanctions is infeasible, as also acknowledged by Khouzani et al. [2014a]. We are interested in the evaluation of practically feasible incentive mechanism, and thus fix the sanctions to an assumed to be collectable level $\sigma = 1$. (Note that this level is equal to the primary losses associated with breaches $\sigma = \lambda_1 = 1$.) Consequently, in our model, the regulator only decides on the *security audit probability* $a \in [0, 1]$.

6.3 Analysis

We now analyze the previously introduced principal-agent model. Specifically, we study the model's social optima in Section 6.3.1. Thereafter, in Section 6.3.2, we determine its Nash equilibria and provide answers to our formulated research question.

6.3.1 Social optima

Social costs are defined as the sum of all firms' expected costs. A planner with control over firms' investment in preventive controls, breach reporting, and the regulator's audits, has a minimization problem based on the costs of firms in Equation (6.7), i. e.,

$$(x^*, s^*) = \arg\min_{x,s} 2 \cdot O(x, x, s, s, 0) \ . \tag{6.10}$$

Observe from this equation that a planner does not require audits: in accordance with Section 2.4, he already controls firms' investment and breach reporting and thus does not have to stimulate both. Furthermore, we may substitute x_i by x and s_i by s

because of firms' symmetry. The solution to the problem in Equation (6.10), derived in Appendix A.1, consists of extreme and boundary values that we subsequently discuss.

A social planner's optimal investment in preventive controls is

$$x^*(s^*) = -\frac{\log\left(\frac{\gamma\cdot H(s^*)+1}{4\cdot\gamma\cdot H(s^*)} - \sqrt{\frac{(\gamma\cdot H(s^*)+1)^2}{16\cdot\gamma^2\cdot H(s^*)^2} - \frac{1}{2\cdot\gamma\cdot\log(\theta)\cdot H(s^*)\cdot I(s^*,0)}}\right)}{\log(\theta)} \,. \tag{6.11}$$

Lemma 6.3.1. *For any optimal investment x^*, if $\alpha > 0$, $\gamma > 0$, $\lambda_2 > 0$, and $\epsilon > 0$, a reporting strategy $0 < s < 1$ is not socially optimal. Under these conditions, the socially optimal reporting strategy is a boundary value, i. e., $s^*(x^*) \in \{0, 1\}$.*

The proof is in Appendix A.1.2.

A social planner will either introduce reporting of all detected breaches, or abstain from reporting altogether. Therefore, his optimal breach reporting strategy is

$$s^*(x^*) = \begin{cases} 1 & \text{if } O(x^*(0), x^*(0), 0, 0, 0) \geq O(x^*(1), x^*(1), 1, 1, 0) \\ 0 & \text{otherwise .} \end{cases} \tag{6.12}$$

This case distinction can be interpreted as the enactment of mandatory breach information sharing with authorities under the assumption of firms that fully comply.

Proposition 6.3.2. *If $O(x^*(0), x^*(0), 0, 0, 0) \geq O(x^*(1), x^*(1), 1, 1, 0)$, a social planner introduces breach information reporting to the authority, and the social optimum is $(s^* = 1, x^*(1))$. Otherwise, the social optimum is $(s^* = 0, x^*(0))$.*

Proof. Follows from Lemma 6.3.1 and Equation (6.12). □

Figure 6.2 illustrates regions for social optima introduced in Proposition 6.3.2, depending on different situations in the (γ, λ_2, α)-parameter space. The three lines each starting in the origin of the coordinate system indicate a social planner's indifference in breach information reporting, which follows from Equation (6.12), for three different types of the authority's private information sharing effectiveness α. Above the lines, the social optimum is $(s^* = 0, x^*(0))$. In these parameter spaces, a social planner abstains from breach reporting as this leads to social costs lower than those from introducing firms' information exchange with the authority. By contrast, in the regions on and below the lines, reporting can be socially beneficial such that breach information sharing gets established, i. e., the social optimum is $(s^* = 1, x^*(1))$. Observe that the region below a line is larger for a more effective authority. This leads to the conclusion that a social planner's decision on the introduction of breach reporting to a large extent depends on the authority's effectiveness to privately share information.

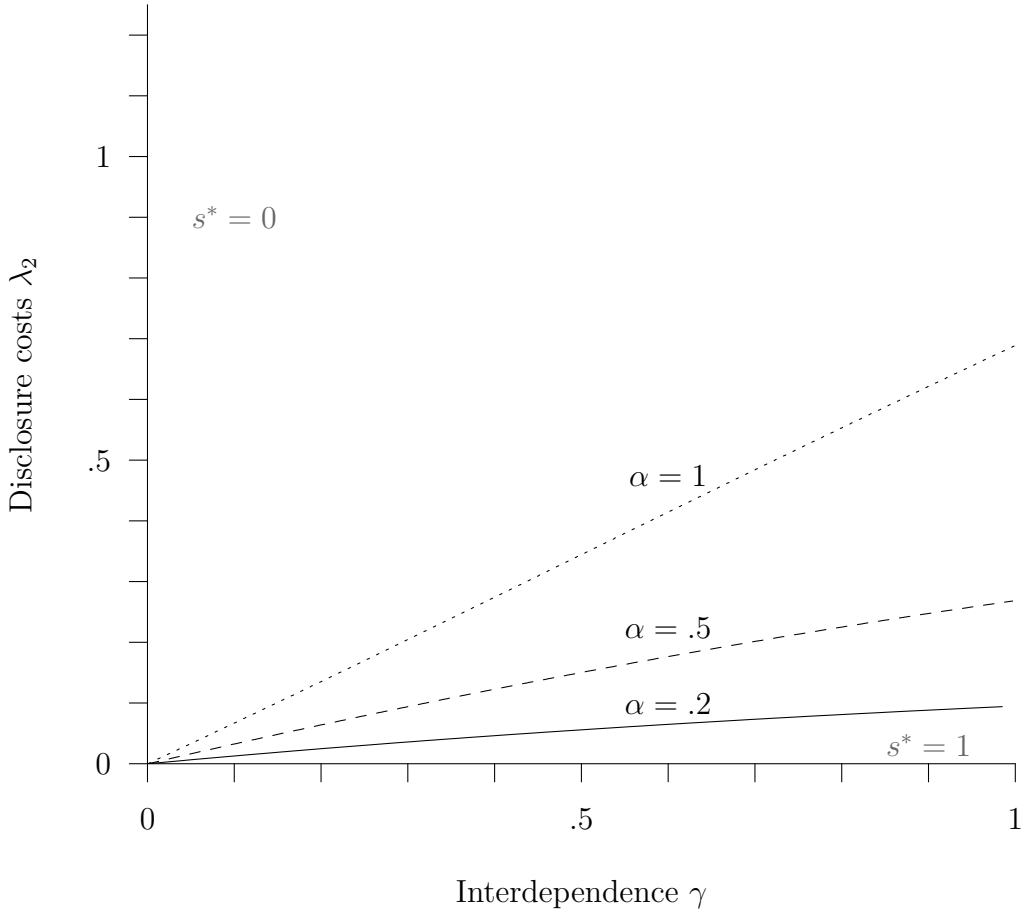


Fig. 6.2 Social planner's case distinction in (γ, λ_2, α)-parameter space.

6.3.2 Nash equilibria

In practice, however, there is no social planner and firms' strategies to minimize expected costs associated with cyber risk are determined by their incentives to invest in preventive controls and share breach information. A game-theoretic approach is needed to analyze these incentives. In what follows, we search for the Nash equilibria of the devised principal-agent game, i. e., the fixed points of the best responses of firms and the regulator, as outlined in Section 2.4. According to Macho-Stadler and Pérez-Castrillo [2009], Nash equilibria of a principal-agent game with moral hazard can be derived by the following two steps: (1) determination of the equilibria in a game between agents (firms), thereby disregarding the best response of the principal (regulator), and (2) backwards induction of derived equilibria into the objective function of the regulator (principal) to determine his best response.

Agents (firms)

If a breach notification law is enacted, firms simultaneously and independently decide on investment and breach reporting with the objective to minimize their expected costs specified in Equation (6.7). Therefore, each firm's best responses is the solution to

$$(x_i^+, s_i^+) = \arg\min_{x_i, s_i} O_i(x_i, x_{1-i}, s_i, s_{1-i}, a) \ , \quad \text{s. t. } x_i \geq 0 \ . \tag{6.13}$$

Nash equilibria follow from the mutual best responses of the two symmetric firms. The derivation of these equilibria is proposed in Appendix A.2.

Depending on the parameter setting, up to three Nash equilibria can exist simultaneously. These equilibria imply the investments in preventive controls

$$\tilde{x}_{1,2}(\tilde{s}, a) = -\frac{\log\left(\frac{1}{2\cdot\gamma\cdot H(\tilde{s})} \pm \sqrt{\frac{1}{4\cdot\gamma^2\cdot H(\tilde{s})^2} - \frac{1}{\gamma\cdot\log(\theta)\cdot H(\tilde{s})\cdot I(\tilde{s},a)}}\right)}{\log(\theta)} \ , \text{ and} \tag{6.14}$$

$$\tilde{x}_3(\tilde{s}, a) = 0 \ . \tag{6.15}$$

Lemma 6.3.3. *If the Nash equilibrium implying $\tilde{x}_1(\tilde{s}, a)$ exists, then two other equilibria that contain the investments in preventive controls $\tilde{x}_{2,3}(\tilde{s}, a)$ exist simultaneously. Thereby, it holds that $\tilde{x}_3(\tilde{s}, a) = 0 \leq \tilde{x}_1(\tilde{s}, a) \leq \tilde{x}_2(\tilde{s}, a)$. Moreover, there are settings where only the equilibrium with investment $\tilde{x}_2(\tilde{s}, a)$ or $\tilde{x}_3(\tilde{s}, a) = 0$ persist.*

The proof is in Appendix A.2.2.

Three categories of model parameter settings have to be distinguished, each resulting in equilibria that imply different investments. These categories can, for instance, be illustrated based on interdependence of firms' ICT system security, ceteris paribus. If interdependence is low, only an equilibrium where firms choose to extensively invest in breach prevention $\tilde{x}_2(\tilde{s}, a)$ exists. With moderate interdependence, two additional equilibria implying the investment strategies $\tilde{x}_{1,3}(\tilde{s}, a)$ evolve. If there is high interdependence, only the equilibrium where firms abstain from any investment $\tilde{x}_3(\tilde{s}, a) = 0$ exists. All of these equilibria also imply firms' mutual best responses in breach reporting.

Lemma 6.3.4. *If breach reporting does not associate disclosure costs $\lambda_2 = 0$, only Nash equilibria where firms voluntarily report breaches exist $\tilde{s} = 1$. Otherwise, in case that $\lambda_2 > 0$, only equilibria implying that firms do not report breaches exist $\tilde{s} = 0$, unless a security audit probability $a \geq a_{\min} = \lambda_2/(\lambda_2 + \sigma)$ is introduced.*

The proof is in Appendix A.2.3.

According to Lemma 6.3.4, Nash equilibria imply the breach reporting strategy

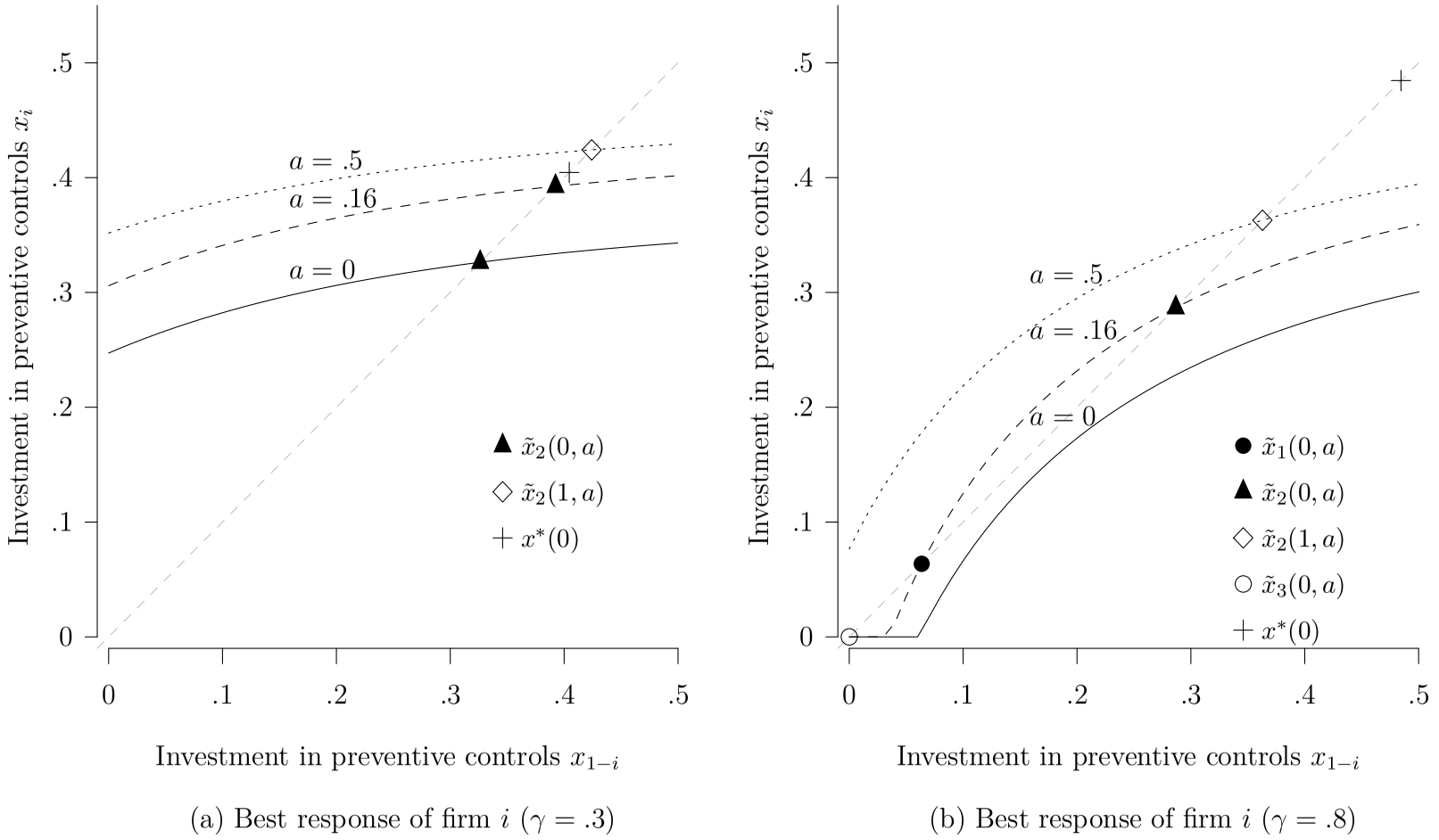
$$\tilde{s}(\tilde{x}, a) = \begin{cases} 1 & \text{if } a \geq a_{\min} \vee \lambda_2 = 0 \\ 0 & \text{otherwise .} \end{cases} \tag{6.16}$$

With respect to this case distinction, firms' reporting decisions depend on associated disclosure costs and the initiation of audits by the regulator. If there are no disclosure costs, firms will always comply with reporting obligations regardless of audits. This is because in Section 6.2.3, we assume that firms act law-abiding in cases where reporting does not make them worse off than non-reporting. Otherwise, the regulator requires security audits and sanctions to incentivize compliance. For this mechanism to work, the audit probability has to be adapted to firms' disclosure costs and the sanction level.

Figures 6.3 (a) and (b) demonstrate all interesting cases of firms' best responses to audit probabilities introduced by the regulator, assuming fixed effectiveness $\alpha > 0$, disclosure costs $\lambda_2 > 0$, and sanctions $\sigma > 0$. Both figures include the social optimum $(s^* = 0, x^*(0))$ for annotated parameter setups as a reference point (indicated by $+$).

Figures 6.3 (a) and (b) depict best responses of firms who expect security audits with probability $0 \leq a < a_{\min}$ by solid and dashed lines. (Cf. the next paragraph for a discussion of the dotted lines.) This security audit probability does not incentivize firms' compliance with breach reporting obligations. However, differences in the audit probability and interdependence alter firms' strategies to invest in preventive controls. Figure 6.3 (a) depicts a setting with low interdependence $\gamma = .3$. In this scenario, if no security audits are initiated $a = 0$, only one Nash equilibrium $(a, 0, \tilde{x}_2(0, a))$ exists. This Nash equilibrium implies investment in breach prevention that is below the socially optimal investment. With an increase in the audit probability $a \to a_{\min}$, spending at the Nash equilibrium elevates $\delta\tilde{x}_2(0, a)/\delta a > 0$. Nevertheless, this security spending never reaches the investment in preventive controls that a social planner would introduce. By contrast, Figure 6.3 (b) depicts a setting with high interdependence $\gamma = .8$. In this scenario, if no security audits are initiated $a = 0$, only the Nash equilibrium $(0, 0, \tilde{x}_3(0, 0))$ exists. At this equilibrium, firms do not invest in breach prevention at all. An increase in the security audit probability $a \to a_{\min}$ eventually leads to two additional Nash equilibria $(a, 0, \tilde{x}_{1,2}(0, a))$. Thereby, security spending at the equilibrium decreases $\delta\tilde{x}_1(0, a)/\delta a < 0$ and increases $\delta\tilde{x}_2(0, a)/\delta a > 0$ in the security audit probability, respectively. However, firms' investment in preventive controls at both Nash equilibria always stay below a social planner's optimal investment.

Figures 6.3 (a) and (b) additionally show best responses of firms who expect security audits with probability $a \geq a_{\min}$ by dotted lines. This audit probability incentivizes



(a) Best response of firm i ($\gamma = .3$)

(b) Best response of firm i ($\gamma = .8$)

Fig. 6.3 Best response in investment to prevent breaches and reporting of firm i, given decisions of firm $1 - i$ and the regulator's audit probability; Nash equilibria and social optima are depicted on the angle bisector. (Common assumptions: $\alpha = .2, \lambda_2 = .2, \sigma = 1$, leading to a minimum audit probability of $a_{\min} = .166$ required to effectively enforce reporting.)

firms' compliance with breach reporting obligations. Moreover, differences in the audit probability and interdependence alter firms' strategies to invest in preventive controls. Figure 6.3 (a) depicts a setting with low interdependence $\gamma = .3$. In this scenario, if audits are initiated with probability $a = a_{\min}$, only one Nash equilibrium $(a, 1, \tilde{x}_2(1, a))$ exists. This equilibrium implies investment in breach prevention that exceeds the socially optimal investment. And with an increase in the audit probability $a \to 1$, security spending at the equilibrium elevates further $\delta\tilde{x}_2(1, a)/\delta a > 0$. Therefore, the audit probability may create incentives at firms to substantially over-invest in preventive controls, such that the effective law enforcement potentially leads to higher social costs due to cyber risk. By contrast, Figure 6.3 (b) depicts a setting with high interdependence $\gamma = .8$. In this scenario, if security audits are initiated with probability $a = a_{\min}$, only the Nash equilibrium $(a, 1, \tilde{x}_2(1, a))$ exists. At this equilibrium, firms' investment in breach prevention is still below a social planner's optimal investment. However, with an increase in the audit probability $a \to 1$, firms introduce higher security spending $\delta\tilde{x}_2(1, a)/\delta a > 0$. Consequently, the audit probability can create incentives at firms to under- or over-invest in preventive controls, such that the effective law enforcement *may* lead to lower social costs due to cyber risk. These rationales influence the regulator's enforcement strategy.

Principal (regulator)

The regulator chooses the audit probability. In order to decide on this probability, he observes the maximum investment in breach prevention at the firms $\tilde{x}_2(\tilde{s}, a)$ as incentive compatibility constraint [Macho-Stadler and Pérez-Castrillo, 2009]. He does not have to consider any participation constraints, as breach notification laws are legally binding. Thus, the regulator's objective is to minimize social costs according to

$$\tilde{a} = \arg\min_{a} 2 \cdot O(\tilde{x}_2(\tilde{s}, a), \tilde{x}_2(\tilde{s}, a), \tilde{s}(\tilde{x}_2, a), \tilde{s}(\tilde{x}_2, a), a) \ . \tag{6.17}$$

Lemma 6.3.5. *If the sanction level is positive $\sigma > 0$, social costs always increase in the audit probability except for the case where this probability incites $\tilde{s} = 1$.*

The proof is in Appendix A.2.4.

Following Lemma 6.3.5, a high audit probability has to be avoided. However, according to Lemma 6.3.4, audits may incentivize firms' breach reporting. Consequently, the regulator introduces an audit probability which just breaks even to incentivize reporting, i. e., $a_{\min}$, and at the same time reduces the overall social costs. This leads

to the subsequent case distinction for effective enforcement of breach reporting

$$\tilde{a} = \begin{cases} a_{\min} & \text{if } O(\tilde{x}_2(0,0), \tilde{x}_2(0,0), 0, 0, 0) \geq O(\tilde{x}_2(1, a_{\min}), \tilde{x}_2(1, a_{\min}), 1, 1, a_{\min}) \\ 0 & \text{otherwise .} \end{cases} \tag{6.18}$$

Lemma 6.3.6. *If $\lambda_2 > 0$, $\sigma > 0$, and $\tilde{a} = a_{\min}$, the audit probability at all Nash equilibria $\tilde{a}$ decreases in the sanction level σ and increases in firms' disclosure costs λ_2.*

The proof is in Appendix A.2.5.

The regulator can maintain the power of his mechanism effectively enforcing firms' breach reporting to the authority by substituting the audit probability and sanction level. In general, if high disclosure costs result in disincentives against breach reporting, the regulator can counter this issue by raising the expected sanctions at firms. Therefore, he may raise the probability for security audits at firms or sanction level, ceteris paribus.

Proposition 6.3.7. *All evolving Nash equilibria imply an audit probability $\tilde{a}$ that constitutes a threshold value. If audits with the probability $\tilde{a} = a_{\min}$ decrease social costs as compared to the situation without audits $\tilde{a} = 0$, the regulator chooses the security audit probability $\tilde{a} = a_{\min}$. As a consequence of this, only the Nash equilibrium $(\tilde{a} = a_{\min}, \tilde{s} = 1, \tilde{x}_2(1, a_{\min}))$ exists. Otherwise, the regulator chooses the audit probability $\tilde{a} = 0$, and up the three equilibria $(\tilde{a} = 0, \tilde{s}, \tilde{x}_{1,2,3}(\tilde{s}, 0))$ may exist simultaneously.*

Proof. According to Equation (6.18), the regulator either effectively enforces breach reporting by an audit probability $\tilde{a} = a_{\min}$ or abstains from enforcement. If the regulator enforces reporting, he uses the incentive compatibility constrain $x = \tilde{x}_2(\tilde{s}, a)$ such that the only existing equilibrium is $(\tilde{a} = a_{\min}, \tilde{s} = 1, \tilde{x}_2(1, a_{\min}))$. Otherwise, if he abstains from enforcing the law, up to three equilibria $(\tilde{a} = 0, \tilde{s}, \tilde{x}_{1,2,3}(\tilde{s}, 0))$ may exist. At these equilibria, firms investment in breach prevention analogous to Lemma 6.3.3. Also, firms' willingness to report breaches is in accordance with Lemma 6.3.4. □

Figure 6.4 illustrates regions for the Nash equilibria introduced in Proposition 6.3.7, depending on different situations in the (γ, λ_2, α)-parameter space. The three lines each starting in the origin of the coordinate system indicate the regulator's indifference in effective enforcement of breach reporting by audits, following from Equation (6.18), for different types of the authority's effectiveness α. Above the lines, the regulator does not initiate audits. This is because their introduction increases social cost, and is thus detrimental. Consequently, up to three Nash equilibria can exist simultaneously, i. e., $(\tilde{a} = 0, \tilde{s} = 0, \tilde{x}_{1,2,3}(0, 0))$. In the regions on and below the lines, the regulator initiates

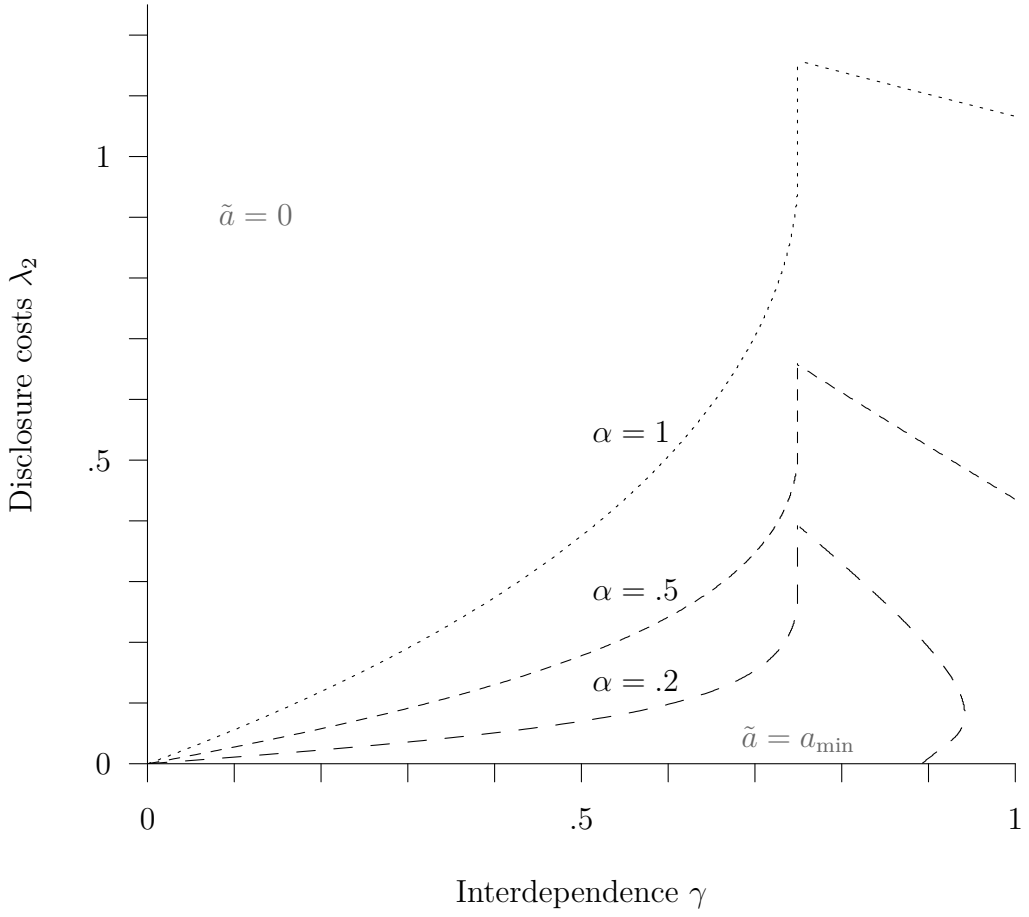


Fig. 6.4 Regulator's case distinction in (γ, λ_2, α)-parameter space.

security audits which just break even to incentivize firms' breach reporting. This decreases the social costs, and only the Nash equilibrium $(\tilde{a} = a_{\min}, \tilde{s} = 1, \tilde{x}_2(1, a_{\min}))$ exists. We conclude that *an effective enforcement of mandatory breach reporting to the authority is welfare improving if interdependence of firms' ICT system security is high, disclosure costs are low, and the sharing effectiveness of the informed authority is high.* In case that interdependence is very high, audits can even stimulate firms' investment in preventive controls (cf. the regions below the lines starting from their slopes at $\gamma = .749$ in Figure 6.4). However, if disclosure costs exceed firms' primary losses associated with successful attacks $\lambda_2 > \lambda_1 = 1$, the regulator's effective enforcement of breach information sharing with the authority is likely to be socially detrimental.

On the abscissa in Figure 6.4, we find the special case where firms always report breaches voluntarily. Thus, the regulator does not have to effectively enforce the law by security audits, and up to three Nash equilibria $(\tilde{a} = 0, \tilde{s} = 1, \tilde{x}_{1,2,3}(1, 0))$ may exist.

6.4 Discussion

Our game-theoretic model covers important characteristics of the conflicting interest between regulators who enforce laws that mandate breach information sharing with authorities and affected firms. Even though this model cannot fully represent reality, we can use results derived from its analysis to infer conditions under which effectively enforcing laws such as the NIS Directive improves welfare, presented in Section 6.4.1. These inferences may have some possible implications, proposed in Section 6.4.2. With respect to the inferences and their possible implications, a critical reflection of model limitations that call for further research is important, conducted in Section 6.4.3.

6.4.1 Inferences from our analysis

Our model analysis reveals that regulators cannot effectively enforce mandatory breach information sharing with authorities without mechanism that verify affected firms' compliance. Specifically, breach notification laws *without* security audits, regardless of the sanction level, do not incentivize breach reporting of affected firms if disclosure costs are not negligible. The reason is that firms' non-compliance with reporting obligations is undetectable and thus never leads to sanctions, which can make compliance coming along with disclosure costs the less attractive option. As an enforcement of corresponding disclosure regimes does not lead to sanctions in cases of non-compliance, firms that do not comply conduct investment in preventive controls solely based on their self-interests. This investment is, with a high probability, below the investment level that a social planner would introduce. Consequently, an enforcement of laws without mechanism that verify firms' compliance may lead to socially undesirable security levels, and therefore potentially results in high social costs due to cyber risk.

By contrast, regulators can effectively enforce mandatory breach information sharing with authorities by mechanism verifying the compliance of affected firms. Specifically, breach notification laws *with* audits *and* sanctions can incentivize breach reporting of affected firms despite disclosure costs. The reason is that firms' non-compliance with reporting obligations is detectable by audits and may thus lead to sanctions, which can make compliance coming along with disclosure costs the more attractive option. As an enforcement of corresponding disclosure regimes leads to costs regardless of compliance, firms may conduct additional investment in preventive controls to reduce breach probabilities and therefore the number of reporting obligations. Yet, if regulators misadjust the enforcement mechanism, it is conceivable that firms over-invest in breach prevention. Consequently, an effective enforcement of breach notification laws by

mechanism verifying firms' compliance may raise the overall security level in economies, and is accompanied by a decrease *or* increase in social costs due to cyber risk.

Our model predicts that regulators' effective enforcement of mandatory breach information sharing with authorities is most reasonable if affected firms have highly interdependent ICT system security, their disclosure costs are low, and the effectiveness of authorities in deriving and sharing conclusions from reported breaches is high. With our exogenous parameter choice, an effective enforcement is almost certainly socially detrimental in case that disclosure costs exceed firms' primary losses due to breaches.

6.4.2 Possible implications

Previous inferences suggests that regulators' effective enforcement of laws mandating firms' breach information sharing with authorities is reasonable only under certain parameter constellations. Specifically, there is a desirable state for each of the three parameter in our model, which at the same time has a regulatory implication. First, firms affected by breach notification laws should be in possession of ICT systems with high security interdependence to others. Consequently, it is reasonable that regulators oblige firms that control ICT systems constituting hubs in the network, e. g., providers of critical infrastructure. Second, firms affected by breach notification laws should have rather low disclosure costs, influencing regulators' adaption of law enforcement mechanism. Yet, disclosure costs at firms of different size and type likely vary. In turn, to avoid over-regulation, it is conceivable that regulators adapt different enforcement mechanism to groups of "similar" firms rather than implement a "one-size-fits-all" solution. Third, authorities that get informed as a result of breach notification laws should have a high effectiveness in deriving and privately sharing conclusions from reported information. We see this effectiveness in the responsibility of regulators who enforce laws, and thus expect them to fund research projects that shed first light into how authorities can be supported in their tasks to reduce economy-wide cyber risk.

Regulators need to carefully analyze in which situations they require what audit probability and sanction level, such that an effective enforcement of firms' breach information sharing with authorities improves welfare. The following scenario highlights practical implications involved with adjusting the enforcement mechanism of a corresponding law. Our model predicts that if a regulator imposes sanctions equal to the primary losses due to breaches at affected firms, whose disclosure costs are of the same height, the optimal audit probability to incentivize breach reporting is roughly 50 %. The situation in Germany suits to examine this scenario. In 2012, the "Statistisches Bundesamt" recorded about 80 000 German firms employing more than 50 individu-

als [Statistisches Bundesamt, 2013]. In the event that the law affects all of these firms, about 40 000 security audits are required – in a period to be defined – to incentivize compliance. And an initiation of more than 40 000 audits, or a considerable increase in the sanction level, ceteris paribus, can incentivize firms to over-invest in preventive controls. In order to effectively enforce the law in a politically feasible way, a trade-off between audits and sanctions is obvious: we can easily imagine that the regulator increases the sanction level to decrease the amount of expensive audits. But this harms firms which do not report breaches because they cannot detect them. From some firms' perspective, expected sanctions may even constitute existential threats. Thus, on the one hand, it is conceivable that these firms circumvent expected sanctions by avoiding the use of ICT systems. This potentially hinders economic growth. On the other hand, if reporting thresholds prescribed by the law are blurry, it is reasonable for the firms to go beyond minimum information sharing obligations. Such over-reporting behavior is undesirable as it harms information quality at the authority, and therefore its effectiveness.

At the same time, an effective enforcement of laws that mandate firms' breach information sharing with authorities can have several positive implications on markets in economies. The reason is that the breach notification laws strengthen awareness for cyber risk at affected firms. For instance, this awareness can incentivize firms to improve their cyber risk mitigation efforts. In turn, trade associations may develop new security standards and best practices, as also suggested in a German initiative anticipating the NIS Directive [Deutscher Bundestag, 2015]. The security industry can take advantage of this by selling products that meet these new requirements. Furthermore, it is conceivable that cyber risk awareness incentivizes risk-averse firms to engage in risk transfers. At the same time, these firms may be able to credibly signal ICT system security levels to cyber risk insurers, e. g., by providing documentations of breach notifications. In turn, such information sharing may allow the cyber insurance market to develop: insurers can use received signals to make their mathematical models of cyber risk more accurate, and therefore improve risk-adjusted premium calculations.

6.4.3 Limitations of the model

Caution is needed when transferring our previous conclusions to the real world, as they are derived from a model with three parameters that cannot cover all relevant aspects of laws mandating firms' breach reporting to authorities. Our model comprises two symmetric firms, representing one economy, that have interdependent ICT system security and are affected by a law. More realistic setups would consider more than two firms affected by the same law, all having different disclosure costs and interdependencies.

Furthermore, our model captures that the authority's shared conclusions, drawn from reported breach information of firms, help recipients to reduce their interdependencies. A modeling approach closer to reality would capture that shared conclusions help recipients to more efficiently invest in breach prevention, potentially leading to reduced breach probabilities and thus less negative externalities. In general, the productivity of preventive controls seems to drive decisions of firms in our model. Therefore, endogenizing this constant, or interpreting it as a parameter, may lead to interesting new insights into when effective enforcements of breach notification laws are socially beneficial.

We also disregard to model that regulators' effective enforcement of laws can be infeasible, e. g., as involved costs cannot be covered by collected sanctions. Specifically, we consider that the authority in our model operates at no costs at all, and regard its decisions as exogenous. Setups closer to reality would capture that the regulator has to fund the authority's actions. Also, we assume that audits at firms are costless and detect all non-reported breaches with certainty. Future models may consider that if the regulator initiate audits, prescribed thoroughness determines their accuracy and costs. Besides, such models may respect that if non-compliance gets detected, sanctions imposed by the regulator can lead to bankruptcy of firms. Overall, endogenizing the regulator's trade-off between funding the authority, audits, and adjusting the sanction level promises interesting results on the feasibility of effectively enforcing laws.

Additionally, to derive results in closed-form, we introduce simplifying assumptions on the utility functions of firms and limit their decision-making space. Specifically, our model considers risk-neutral firms only, captured by their utility function. Yet, in reality, some firms may be risk averse (and comply with reporting obligations even if expected sanctions are rather low) and others risk affine (and do not comply with laws despite high expected sanctions), e. g., justified by their market positions. Also, we assume that firms do not detect and have no interest to report breaches that did not happen. Though, breach over-reporting is conceivable and should be respected in future works: controls to detect breaches can alert firms and thus trigger reporting even though nothing happened; and against the background of the discussions in the previous section, firms may strategically over-report, too. Moreover, in our model, firms can only invest in preventive controls, and their breach detection probability is assumed to be fixed. However, we can easily imagine that firms affected by a law have strong incentives to invest in the detection of successful attacks, as this is a prerequisite for compliance with breach reporting obligations. In the next chapter, we shed first light into the effects of effectively enforcing mandatory breach information sharing with authorities on affected firms' decisions to invest in preventive *and* detective controls.

Chapter 7

Effects of mandatory information sharing on investment decisions

The model analysis in the last chapter indicates that there are conditions under which an effective enforcement of firms' breach information sharing with authorities by audits and sanctions can be socially beneficial. Yet, our previous model does not consider that an enforcement may modify affected firms' trade-off between investment in breach prevention and detection. We revisit effects of effectively enforcing laws such as the NIS Directive on firms' investment incentives in Section 7.1, motivating the research questions in this chapter. To tackle these questions, Section 7.2 introduces a variant of the model presented in the last chapter. We derive and analyze this model's social optima and Nash equilibria in Section 7.3. At last, we present inferences from our model analysis, discuss implications, and highlight model limitations in Section 7.4.

7.1 Motivation

The effective enforcement of breach notification laws by audits and sanctions influences affected firms' investment decisions. Firms have natural incentives to allocate their limited budget on productive activity, generating profit. Yet, this profit is reduced by breach-related costs associated with the disclosure regime. Investments in security controls can decrease these costs, as outlined in Section 6.1 which constitutes the common basis for the last and this chapter: investment in preventive controls helps to prevent breaches, and thus reduces the number of reporting obligations; and investment in detective controls improves information such that fewer breaches remain unnoticed, enabling compliance with reporting obligations and thus to evade *sanctions*. Overall,

firms that are affected by breach notification laws have to trade-off between allocating their limited budget on productive activity, breach prevention, and breach detection.

Smart regulators effectively enforce laws that mandate firms' breach information sharing with authorities if this is socially beneficial. Though, conflicting interest between regulators and firms make enforcement decisions difficult: a regulator enacts a law such as the NIS Directive with the objective to improve welfare by incentivizing firms' breach information reporting to an authority, enabling the latter to *effectively draw and share conclusions with others*. Affected firms need to detect breaches as a prerequisite to meet reporting obligations, but only have few incentives to unilaterally share information due to disclosure costs. In response, the regulator may incentivize compliance by initiating audits at firms to find and sanction unreported breaches. This can encourage firms to invest more in detective controls. Thus, firms potentially invest less into productive activity and breach prevention. In turn, socially undesirable resource allocations are conceivable. This reasoning motivates the research questions tackled in this chapter:

How does an effective enforcement of laws, such as the NIS Directive,

a) change the total spendings of affected firms on preventive and detective controls as compared to a social planner's optimal spendings?

b) change the investment priority of affected firms as compared to a social planner's optimal decisions?

c) change the profit of affected firms compared to a situation without the laws and to the profit at the social optimum?

Answers to these questions are relevant for affected firms' security managers who decide on investment in *preventive* and *detective controls*. Also, answers promise important insights on the incentive mechanism of breach notification laws, relevant for regulators.

7.2 Model

In this section, we devise a game-theoretic model to tackle the above research questions. Subsequent subsections either include decision variables or free model parameters for the properties highlighted in *italic* before. In Section 7.2.1, we introduce a model capturing firms' decisions to invest in *preventive* and *detective controls*. Then, in Section 7.2.2, we formalize firms' mandatory breach information sharing with the authority, and propose our two model parameters: first, the *sanctions* imposed on firms that do not comply with the law; and, second, the authority's *effectiveness to privately share information* with others. Table 7.1 refines the symbols used in this work for the model present next.

Table 7.1 Symbols: effects of mandatory information sharing on investment decisions.

Symbol	Type	Description	Constraint
x	decision variable	investment in preventive controls	$x > 0$
d	decision variable	investment in detective controls	$d > 0$
α	parameter	private information sharing effectiveness	$\alpha \in [0, 1]$
σ	parameter	sanction level	$\sigma \geq 0$
μ	constant	total budget of a firm	$\mu = 1$
θ	constant	breach prevention productivity	$\theta = 200$
ϑ	constant	breach detection productivity	$\vartheta = 250$
ρ	constant	return on investment	$\rho = 1.1$
λ_1	constant	primary loss (detected breach)	$\lambda_1 = .009$
λ_2	constant	disclosure costs	$\lambda_2 = .011$
λ_3	constant	primary loss (undetected breach)	$\lambda_3 = .5$
n	constant	number of firms	$n = 2$
O	(objective) function	profit of a firm	
I	function	impact of a breach	
R	function	cyber risk	
U	function	productive part of investments	
P	function	breach probability	
F	function	breach detection probability	
B	random variable	breach	
β	realization	realization of B	$\beta \in \{0, 1\}$
$\hat{B}$	random variable	breach detection	
$\hat{\beta}$	realization	realization of $\hat{B}$	$\hat{\beta} \in \{0, 1\}$
a	realization	security audit probability	$a \in \{0, 1\}$
ξ	realization	breach information reporting	$\xi \in \{0, 1\}$
s	realization	compliances with breach reporting	$s \in \{0, 1\}$
ϕ	reference	reference point in parameter space	

7.2.1 Investments of a firm

Consider for now a single rational and risk neutral firm in a larger economy. It has a total budget of $\mu = 1$, which can be invested in the provision of products and services $U \geq 0$ or ICT system security, i. e., *preventive controls* $x > 0$ or *detective controls* $d > 0$. Yet, every dollar invested in productive activity can no longer be spend on controls:

$$U(x, d) = \mu - x - d \ . \tag{7.1}$$

Investment in productive activity generates constant return $\rho \geq 1$. However, the profit from such investment is reduced by cyber risk $R(x, d)$ due to breaches that may

happen to the firm's ICT system. Thus, the overall profit of the firm can be denoted as

$$O(x,d) = \rho \cdot U(x,d) - R(x,d) \ . \tag{7.2}$$

The expected cyber risk $R(x,d)$ at the firm is influenced by its investments in preventive and detective controls. Investment in preventive controls reduces the breach probability of the firm's ICT system $P(x)$. And investment in detective controls increases the probability of finding breaches that have happened $F(d)$. We assume that if a firm detects a successful attack, it results in primary losses λ_1. By contrast, a breach that occurs and remains undetected entails considerably higher primary losses λ_3, as the successful attacker may exhaust vulnerabilities in the firm's ICT system over time and compromise large parts of the internal network. Thus, we may specify that $\lambda_3 \gg \lambda_1$. In turn, the overall cyber risk at the firm is given by

$$R(x,d) = P(x) \cdot I(d) \ , \text{ where} \tag{7.3}$$

$$I(d) = F(d) \cdot \lambda_1 + (1 - F(d)) \cdot \lambda_3 \ . \tag{7.4}$$

We capture the breach probability $P(x)$ by the realization $\beta \in \{0,1\}$ of the random variable B (breach), such that $Pr(\beta = 1) = P(x)$. Investment in preventive controls decreases this probability at a decreasing rate, i. e., $\partial P/\partial x < 0$, $\partial^2 P/\partial x^2 > 0$, and $\lim_{x\to\infty} P(x) = 0$. A functional form for the breach probability is $P(x) = \theta^{-x}$ [Böhme, 2012]. In this equation, the constant $\theta \geq 0$ represents productivity of investment in preventive controls. Also, observe from the breach probability function that without investment in prevention, the firm falls victim to every realized threat, i. e., $P(0) = 1$.

We capture the probability for the firm to detect a breach of its ICT system $F(d)$ by the realization $\hat{\beta} \in \{0,1\}$ of the random variable $\hat{B}$ (breach detection), such that $Pr(\hat{\beta} = 1 | \beta = 1) = F(d)$. Investment in detective controls increases this probability at a decreasing rate, i. e., $\partial F/\partial d > 0$, $\partial^2 F/\partial d^2 < 0$, and $\lim_{d\to\infty} F(d) = 1$. A functional form for the probability of breach detection that captures all of these properties is $F(d) = 1 - \vartheta^{-d}$, which matches the vulnerability discovery probability function in [Khouzani et al., 2014b]. In this equation, the constant $\vartheta \geq 0$ represents productivity of investment in detective controls. This productivity determines the controls' type II error probability $1 - F(d)$, and we abstain from modeling type I errors. Additionally, observe from the breach detection probability function that without investment in detection, no breach can be found, not even by accident, i. e., $F(0) = 0$.

A law that mandates breach information sharing with authorities may have an effect on both, the affected firm's incentives to invest in preventive and detective controls.

7.2.2 Mandatory breach information sharing

We generalize our model to $n = 2$ symmetric firms that represent an economy, indexed by $i \in \{0, 1\}$. We assume interdependence of the firms' ICT system security, but do not capture this in our model to keep analytical tractability. Thus, unprotected ICT systems can generate negative externalities, such that the regulator may justify to effectively enforce firms' mandatory breach information sharing with the authority.

We assume that a disclosure regime obliges the firms to report $\xi_i \in \{0, 1\}$ breach *and* related control information, i. e., implemented practices to prevent and detect successful attacks, which goes beyond notification requirement introduced in the last chapter but is also indicated by the NIS Directive. A firm's decision to share information that no breach has been detected is denoted as $\xi_i = 0$. Accordingly, $\xi_i = 1$ indicates that all relevant information get reported. To keep our model simple, we also assume that nobody has an interest in reporting beaches that did not happen, i. e., compliance with the law is $Pr(\xi_i = 1|\hat{\beta}_i = 1) = s_i$. Compliance entails negative and positive effects on firms.

Breach information sharing with the authority causes disclosure costs $\lambda_2 > 0$ at firms. Thus, firms may not have incentives to share their private information in the first place. We assume that – without the regulator taking measures – expected disclosure costs hinder firms' compliance. However, the regulator can effectively enforce breach reporting despite disclosure costs by initiating audits of firms' systems with probability $a \in [0, 1]$, and imposing *sanctions* to the amount of $\sigma > 0$ for detected non-compliance.

Furthermore, firms' breach information sharing with the authority can lead to economy-wide more effective security investments (cf. Section 5.2.1). This is because reported information allow the authority to draw and share conclusions on how to prevent and detect successful attacks, which may help recipients improve their cyber risk management efforts. For instance, shared conclusions can hint to measures that protect from propagating attacks, thus helping recipients to more effectively invest in preventive controls. Additionally, the authority can share information that warns concerning successful attacks, thus helping recipients to more effectively invest in detective controls. Overall, we model the positive effects resulting from an informed authority's private advice to firms as costless improvements of preventive and detective controls, such that the probability of breaches and their detection becomes, respectively:

$$P_i = P_i(x_i, x_{1-i}) = \theta^{-(x_i + \alpha \cdot s_{1-i} \cdot x_{1-i})} \text{ , and} \tag{7.5}$$

$$F_i = F_i(d_i, d_{1-i}) = 1 - \vartheta^{-(d_i + \alpha \cdot s_{1-i} \cdot d_{1-i})} \text{ ,} \tag{7.6}$$

where $\alpha \in [0, 1]$ *parameterizes the authority's private information sharing effectiveness.*

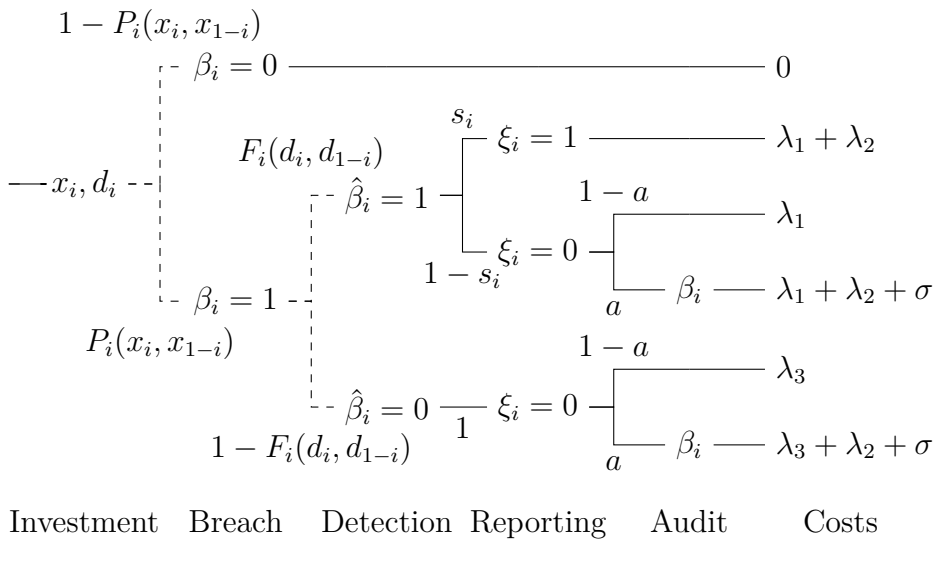


Fig. 7.1 Decisions of firm i, nature, and the regulator.

Figure 7.1 visualizes the calculation of firm i's expected cyber risk under a disclosure regime. The figure depicts all decisions of the firm and the regulator. Dashed lines represent uncertainty because of nature's decisions. Initially, firm i chooses whether or not to comply with breach information sharing obligations and invests in preventive controls x_i, detective controls d_i, as well as productive activity $U(x_i, d_i)$. Then, an attack on the firm's ICT system may take place. This attack is successful with probability $P_i(x_i, x_{1-i})$. Note that in every period under consideration, there can at most be one breach to firm i's system. Every breach causes primary losses. Yet, the amount of primary losses depends on whether the firm detects the successful attack (λ_1) or not (λ_3). Thereby, the breach detection probability is $F_i(d_i, d_{1-i})$. If the firm does not detect a successful attack, it will not report corresponding information to the authority. And in case that no breach information are reported, the regulator initiates audits at random. We assume that audits find every unreported breach and do not create false positives. Hence, they are much more reliable than detective controls at a firm. Furthermore, we ignore audit costs and assume that the regulator can pay all auditors from the sum of collected sanctions. If auditors detect an unreported breach, this results in sanctions $\sigma \in [0, \infty[$ for the firm due to non-compliance. Also, auditors notify the authority regarding the successful attack – yet we assume that the authority does not derive and privately share conclusions from auditors' information with other firms. Once a firm's breach information are reported, they may become public because of the authority's actions. Therefore, firm i has to expect disclosure costs λ_2.

In order to effectively enforce firms' breach information sharing despite disclosure costs, the regulator can adjust the audit probability a and sanction level σ based on his own cost structure. For simplicity, we subsequently assume that information sharing is always enforced with an audit probability $a = 1$ and a collectable sanction level $\sigma > 0$. At the same time, an enforced disclosure regime incentivizes firms to fully share their private breach information ($s = 1$). Contrarily, we obtain the scenario without mandatory information sharing by setting the audit probability to $a = 0$. In turn, firms do not report breach information ($s = 0$). We can derive firms' expected cyber risk without $R_i^{s=0}$ and with $R_i^{s=1}$ enforced breach information sharing from Figure 7.1, i. e.,

$$R_i^0(x_i, x_{1-i}, d_i, d_{1-i}, 0) = P_i(x_i, x_{1-i}) \cdot I_i^0(d_i, d_{1-i}, 0) \text{ , with} \tag{7.7}$$

$$I_i^0(d_i, d_{1-i}, 0) = F_i \cdot \lambda_1 + (1 - F_i) \cdot \lambda_3 \text{ ; and} \tag{7.8}$$

$$R_i^1(x_i, x_{1-i}, d_i, d_{1-i}, a) = P_i(x_i, x_{1-i}) \cdot I_i^1(d_i, d_{1-i}, a) \text{ , with} \tag{7.9}$$

$$I_i^1(d_i, d_{1-i}, a) = F_i \cdot (\lambda_1 + \lambda_2) + (1 - F_i) \cdot [(1 - a) \cdot \lambda_3 + a \cdot (\lambda_3 + \lambda_2 + \sigma)] \text{ .} \tag{7.10}$$

With respect to the above equations that capture cyber risk under different disclosure regimes, firms' expected profit in Equation (7.2) expand to either

$$O_i^0(x_i, x_{1-i}, d_i, d_{1-i}, 0) = \rho \cdot U(x_i, d_i) - R_i^0(x_i, x_{1-i}, d_i, d_{1-i}, 0) \text{ , or} \tag{7.11}$$

$$O_i^1(x_i, x_{1-i}, d_i, d_{1-i}, a) = \rho \cdot U(x_i, d_i) - R_i^1(x_i, x_{1-i}, d_i, d_{1-i}, a) \text{ .} \tag{7.12}$$

7.3 Analysis

In this section, we analyze the previously introduced game-theoretic model by making use of the solution concepts proposed in Section 2.4. We derive the model's social optima in Section 7.3.1, and its Nash equilibria in Section 7.3.2. Thereafter, we inquire both the social optima and Nash equilibria by comparing different hypothetical scenarios in the parameter space. We set and justify constants that we assume within these scenarios in Section 7.3.3. Thereafter, in Section 7.3.4, derived social optima are examined. We inspect Nash equilibria in Section 7.3.5. And finally, in Section 7.3.6, we respond to the formulated research questions by comparing the social optima and Nash equilibria that evolve in the different hypothetical scenarios with each other.

7.3.1 Social optima

The social optimum maximizes the sum of both firms' profit. A social planner has full control over the symmetric firms' decisions, i. e., $x_i = x_{1-i} = x$, and $d_i = d_{1-i} = d$, and can thus adapt them to reach the maximum. Implicitly, he does not need incentivize investment and breach information sharing by audits and sanctions, i. e., $a = \sigma = 0$.

In case that a planner does not introduce breach information sharing $s = 0$, he maximizes firms' profit based on Equation (7.11), i. e.,

$$(x^*, d^*) = \arg\max_{x,d} 2 \cdot O^0(x, x, d, d, 0) \ . \tag{7.13}$$

The solution to the problem in Equation (7.13) is given in Appendix B.1.1. At the social optimum without breach information sharing, security investment in preventive and detective controls is, respectively,

$$x_1^* = \frac{\log\left(-\frac{\lambda_1 \cdot \log(\theta) \cdot \log(\vartheta)}{\rho \cdot \log(\theta) - \rho \cdot \log(\vartheta)}\right)}{\log(\theta)} \quad \text{, and} \tag{7.14}$$

$$d_1^* = \frac{\log\left(\frac{(\lambda_1 - \lambda_3) \cdot (\log(\theta) - \log(\vartheta))}{\lambda_1 \cdot \log(\theta)}\right)}{\log(\vartheta)} \ . \tag{7.15}$$

On the other hand, if a planner introduces breach information sharing $s = 1$, he maximizes firms' profit based on Equation (7.12), i. e.,

$$(x^*, d^*) = \arg\max_{x,d} 2 \cdot O^1(x, x, d, d, 0) \ . \tag{7.16}$$

The solution to the problem in Equation (7.16) is given in Appendix B.1.2. At the social optimum with breach information sharing, security investment in preventive and detective controls is, respectively,

$$x_2^* = \frac{\log\left(-\frac{(\alpha+1) \cdot \log(\theta) \cdot \log(\vartheta) \cdot (\lambda_1 + \lambda_2)}{\rho \cdot (\log(\theta) - \log(\vartheta))}\right)}{(1+\alpha) \cdot \log(\theta)} \quad \text{, and} \tag{7.17}$$

$$d_2^* = \frac{\log\left(\frac{(\log(\theta) - \log(\vartheta)) \cdot (\lambda_1 + \lambda_2 - \lambda_3)}{\log(\theta) \cdot (\lambda_1 + \lambda_2)}\right)}{(\alpha+1) \cdot \log(\vartheta)} \ . \tag{7.18}$$

Proposition 7.3.1. *If $O^0(x_1^*, x_1^*, d_1^*, d_1^*, 0) \leq O^1(x_2^*, x_2^*, d_2^*, d_2^*, 0)$, a planner introduced breach information sharing $s = 1$ as this can increases the profit of both firms. In this case, the only social optimum is (x_2^*, d_2^*). Else, the only social optimum is (x_1^*, d_1^*).*

Proof. Follows from $s \in \{0, 1\}$ and Equations (7.17), (7.18), (7.14), (7.15). □

7.3.2 Nash equilibria

In practice, each firm's individual profit expectation determines its willingness to invest in security controls. Thereby, one firm's actions affect the other's outcomes, requiring a game-theoretic approach. We use Nash equilibria as solution concept and analyze their existence and location depending to whether the regulator does not ($a = 0$) or does ($a = 1$) effectively enforce firms' breach information sharing with the authority.

Without enforced information sharing, firm i maximizes Equation (7.11), i. e.,

$$(x_i^+, d_i^+) = \arg\max_{x_i, d_i} O_i^0(x_i, x_{1-i}, d_i, d_{1-i}, 0) \ . \tag{7.19}$$

The solution to Equation (7.19) is the best response of firm i in a regime where the regulator does not enforce information sharing by initiating audits ($a = 0$) and with respect to the decisions of firm $1 - i$. Nash equilibria follow from fixed points of both firms' mutual best responses. We derive these equilibria in Appendix B.2.1. If the regulator does not enforces information sharing, at the Nash equilibrium, firms' investment in preventive and detective controls is, respectively,

$$\tilde{x}_1 = \frac{\log\left(-\frac{\lambda_1 \cdot \log(\theta) \cdot \log(\vartheta)}{\rho \cdot \log(\theta) - \rho \cdot \log(\vartheta)}\right)}{\log(\theta)} \quad \text{, and} \tag{7.20}$$

$$\tilde{d}_1 = \frac{\log\left(\frac{(\lambda_1 - \lambda_3) \cdot (\log(\theta) - \log(\vartheta))}{\lambda_1 \cdot \log(\theta)}\right)}{\log(\vartheta)} \ . \tag{7.21}$$

With enforced information sharing, firm i maximizes Equation (7.12), i. e.,

$$(x_i^+, d_i^+) = \arg\max_{x_i, d_i} O_i^1(x_i, x_{1-i}, d_i, d_{1-i}, 1) \ . \tag{7.22}$$

The solution to Equation (7.22) is the best response of firm i in a regime where the regulator effectively enforces information sharing by initiating audits ($a = 1$) and with respect to the decisions of firm $1 - i$. Fixed points of both firms' mutual best responses are derived in Appendix B.2.2. If the regulator effectively enforces the law, at the Nash equilibrium, firms' investment in preventive and detective controls is, respectively,

$$\tilde{x}_2 = \frac{\log\left(-\frac{\log(\theta) \cdot \log(\vartheta) \cdot (\lambda_1 + \lambda_2)}{\rho \cdot (\log(\theta) - \log(\vartheta))}\right)}{(\alpha + 1) \cdot \log(\theta)} \quad \text{, and} \tag{7.23}$$

$$\tilde{d}_2 = \frac{\log\left(\frac{(\log(\theta) - \log(\vartheta)) \cdot (\lambda_1 - \lambda_3 - \sigma)}{\log(\theta) \cdot (\lambda_1 + \lambda_2)}\right)}{(\alpha + 1) \cdot \log(\vartheta)} \ . \tag{7.24}$$

Proposition 7.3.2. *If* $O_i^0(\tilde{x}_1, \tilde{x}_1, \tilde{d}_1, \tilde{d}_1, 0) \leq O_i^1(\tilde{x}_2, \tilde{x}_2, \tilde{d}_2, \tilde{d}_2, 1)$*, a smart regulator uses audits* $a = 1$ *and sanctions* $\sigma > 0$ *to effectively enforce breach information sharing* $s = 1$ *as this can improve the profit of both firms. In this case, the only Nash equilibrium is* $(\tilde{x}_2, \tilde{d}_2)$*. Otherwise, the only Nash equilibrium is* $(\tilde{x}_1, \tilde{d}_1)$*.*

Proof. Follows from $a \in \{0, 1\}, a = s$, and Equations (7.23), (7.24), (7.20), (7.21). □

7.3.3 Scenario setup

In order to analyze the social optima and Nash equilibria of Propositions 7.3.1 and 7.3.2 in parameter space, we first have to set and justify the variables that are unspecified until know. Specifically, for the subsequent numerical analysis, we specify all exogenous model variables as constants relative to the investment budget $\mu = 1$ of each firm. A typical order of magnitude for our unit μ would be US$ 1 billion in the real world.

Return on investment

Firms can spend their budget on productive activity, preventive, or detective controls. To contextualize all of these investments, we first fix the return on investment in productive activity at $\rho = 1.1$. This value constitutes the 10 year average of the "Dow Jones Industrial Average" – which was 8.36 % as of July 2005 – rounded to 10 %. Yet, budget allocations on productive activity, that generate returns, have to be traded of against investment in preventive or detective controls, reducing breach related costs.

Costs of detected breaches

We take into account the "Target breach" that has happened at the end of the year 2013 to estimate the costs of detected breaches at firms. The Target Corporation is a firm that had a total equity of US$ 14 billion in the financial year 2014. This total equity can be used as an estimate for the budget of Target. The breach at Target resulted in costs of about US$ 1 billion.[1] By attributing all of these costs to the financial year 2014, we find that detected breaches at firms who have more than US$ 1 billion as total budget can result in costs of $\lambda_1 + \lambda_2 = 1/14 = .07$, relative to our model. Yet, as the Target breach belongs to the worst breaches of all time, it is reasonable to assume that the majority of successful attacks in economies are not that devastating.[2] Thus, we fix

[1] Further information on this figure can be found on a website of the New York Times (http://www.nytimes.com/2014/08/06/business/target-puts-data-breach-costs-at-148-million.html?_r=0).

[2] The breach at Target is put into context, e. g., on the technology news website "tomsguide" (http://www.tomsguide.com/us/pictures-story/872-worst-data-breaches.html).

the costs of detected breaches at $\lambda_1 + \lambda_2 = .02$, assuming primary losses of $\lambda_1 = .009$ and disclosure costs of $\lambda_2 = .011$. This cost ratio goes in line with previous research concluding that, if breaches become public, an affected firm's primary losses are lower than its disclosure costs – specifically secondary losses [Cavusoglu et al., 2004b].

Costs of undetected breaches

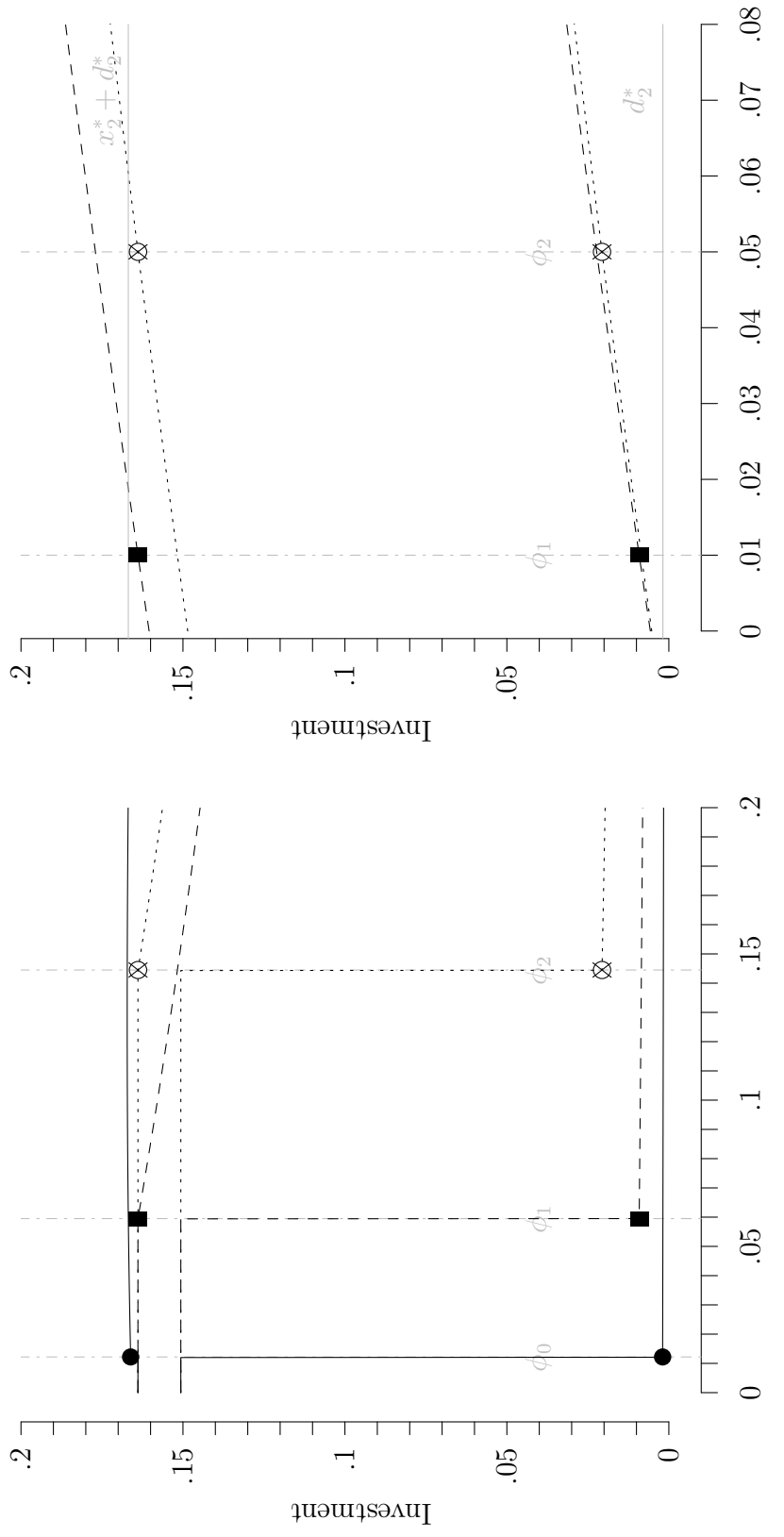
We assume that breaches which remain undetected by firms for a long time are more severe than detected breaches. The reason is that attackers have time to compromise large parts of the firms' internal network. However, we do not find empirical studies supporting any particular cost level of such breaches. In our model, we fix the costs of undetected breaches at $\lambda_3 = .5 \gg \lambda_1 = .009$, assuming that firms face an existential threat if they do not invest in detective controls and thus overlook all successful attacks.

Productivity of investments

Security investments may help firms to decrease breach related costs. Yet, it is notoriously hard to calibrate investment productivity parameters in analytical models. Acknowledging the uncertainty, we fix the productivity of investment in preventive controls at $\theta = 200$. This level can be considered "high," as compared to the productivity specified in the previous chapter. Furthermore, we fix the productivity of investment in detective controls at $\vartheta = 250$. This efficiency is considerably higher than the productivity of investment in vulnerability detection specified by Khouzani et al. [2014b]. Therefore, in our model, we consider breaches less costly to detect than vulnerabilities.

7.3.4 Decisions of a social planner

The previous setup allows us to analyze a social planner's decisions. The two solid lines in Figure 7.2 (a) show his investment decisions as a function of the authority's information sharing effectiveness. The lowermost solid line describes optimal investment in detective controls d^*, while the uppermost solid line sketches the sum of optimal investments in controls $x^* + d^*$. Furthermore, we introduce a reference point ϕ_0 in Figure 7.2 (a) that restricts the interval of low information sharing effectiveness from above, i. e., the equation in Proposition 7.3.1 is not fulfilled for an effectiveness in $0 \leq \alpha < \phi_0$. If the authority's effectiveness is below the reference point ϕ_0, a social planner does not introduce information sharing, and the social optimum is (x_1^*, d_1^*). By contrast, if the authority's effectiveness is in $\phi_0 \leq \alpha \leq 1$, this justifies a planner's introduction of breach information sharing, and the social optimum is (x_2^*, d_2^*).



(a) $\sigma = .01$ (dashed); $\sigma = .05$ (dotted)

(b) $\alpha = .06$ (dashed); $\alpha = .14$ (dotted)

Fig. 7.2 Socially optimal investments (solid lines) and investments at the Nash equilibrium (dashed and dotted lines); lowermost lines: investment in detective controls d; uppermost lines: sum of investments $x + d$; vertical gray dashed/dotted lines: indifference points between disclosure regimes.

No breach information sharing

Solid lines within the interval $0 \leq \alpha < \phi_0$ in Figure 7.2 (a) depict a social planner's optimal investment decisions if he does not introduce information sharing. In this interval, the social optimum (x_1^*, d_1^*) does not depend on the authority's information sharing effectiveness α (cf. Equations (7.14) and (7.15)). Therefore, a social planner's investments in preventive and detective controls are constant. Specifically, investment in breach prevention is at $x_1^* = .013$, and investment in breach detection at $d_1^* = .151$. And the total security investments are constant at $x_1^* + d_1^* = .164$. In the discussed interval, a planner always invests less in preventive than detective controls, i. e., $x_1^* < d_1^*$.

Breach information sharing

Solid lines within the interval $\phi_0 \leq \alpha \leq 1$ in Figure 7.2 (a) depict a planner's optimal investment decisions if he introduces information sharing. Here, the social optimum (x_2^*, d_2^*) depends on the authority's effectiveness (cf. Equations (7.17) and (7.18)). At the reference point ϕ_0, investment in breach prevention and detection is at $x_2^* = .165$ and $d_2^* = .002$, respectively. And these two investments monotonically decrease in the effectiveness α. Thus, the maximum investments are at $x_2^* + d_2^* = .167$. In the discussed interval, a planner always invests more in preventive than detective controls, i. e., $x_2^* > d_2^*$.

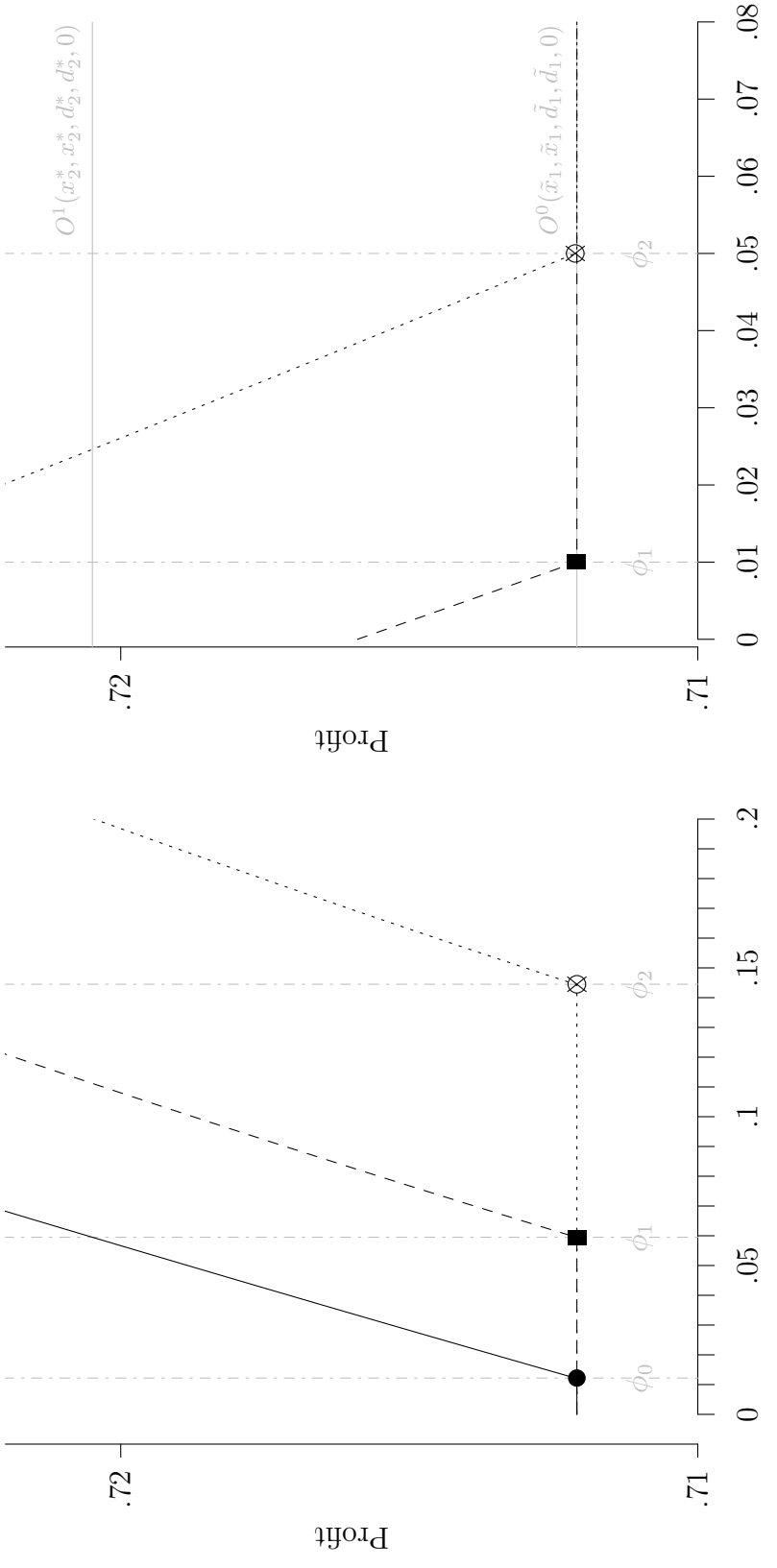
Welfare

The solid line in Figure 7.3 (a) depicts a social planner's generated profit at the social optimum as a function of the authority's information sharing effectiveness α. For an effectiveness in the interval $0 \leq \alpha < \phi_0$, this profit is constant at $O^0(x_1^*, x_1^*, d_1^*, d_1^*, 0) = .712$, as no breach information sharing is introduced. By contrast, once a social planner introduces breach information sharing, i. e., in the interval $\phi_0 \leq \alpha \leq 1$, the profit increases in the authority's effectiveness (at a decreasing rate, not visible in the figure).

7.3.5 Decisions of firms

The dashed and dotted lines in Figure 7.2 (a) show firms' investment decisions as a function of the authority's information sharing effectiveness, influencing the regulator's decision to effectively enforce a law with sanctions.[3] The lower dashed and dotted lines describe firms' investment in detective controls at the equilibrium $\tilde{d}$. And the upper dashed and dotted lines show the sum of investments at the Nash equilibrium $\tilde{x} + \tilde{d}$.

[3]Recall that effective enforcement of breach information sharing is always accompanied by a security audit probability of $a = 1$.



(a) $\sigma = .01$ (dashed); $\sigma = .05$ (dotted)

(b) $\alpha = .06$ (dashed); $\alpha = .14$ (dotted)

Fig. 7.3 Profit at the social optimum (solid lines) and the Nash equilibria (dashed and dotted lines); vertical gray dashed/dotted lines: indifference points between disclosure regimes.

Dashed lines in Figure 7.2 (a) describe firms' investment decisions in a scenario where the regulator requires low sanctions $\sigma = .01$ to effectively enforce the breach notification law. In this scenario, the regulator does not enforce the law if the authority's effectiveness α is below the reference point ϕ_1, i. e., the equation in Proposition 7.3.2 is not fulfilled for an effectiveness in $0 \leq \alpha < \phi_1$, resulting in the equilibrium $(\tilde{x}_1, \tilde{d}_1)$. By contrast, if the effectiveness is in $\phi_1 \leq \alpha \leq 1$, the regulator effectively enforces information sharing as this is socially beneficial, resulting in the equilibrium $(\tilde{x}_2, \tilde{d}_2)$.

Dotted lines in Figure 7.2 (a) describe firms' investment decisions in a scenario where the regulator requires high sanctions $\sigma = .05$ to effectively enforce the law. In this scenario, the regulator does not enforce information sharing if the authority's effectiveness α is below the reference point $\phi_2 > \phi_1$, i. e., the equation in Proposition 7.3.2 is not fulfilled for an effectiveness in $0 \leq \alpha < \phi_2$, such that the equilibrium $(\tilde{x}_1, \tilde{d}_1)$ emerges. By contrast, the regulator effectively enforces the breach notification law if the authority's effectiveness is in the interval $\phi_2 \leq \alpha \leq 1$, resulting in the Nash equilibrium $(\tilde{x}_2, \tilde{d}_2)$.

Observe from Figure 7.2 (a) that firms respond to higher expected sanctions with increased investment in detective controls. Subsequently, we analyze the low sanctions scenario and extend our discussion to a setup with varying sanctions where necessary.

No breach information sharing

Dashed lines within the interval $0 \leq \alpha < \phi_1$ in Figure 7.2 (a) depict firms' investment decisions if the regulator does not enforce the notification law. In this interval, the decisions of firms are the same as those of a social planner who does not introduce information sharing, i. e., $(x_1^*, d_1^*) = (\tilde{x}_1, \tilde{d}_1)$ (cf. the corresponding social optimum and Nash equilibrium in Sections 7.3.1 and 7.3.2). Thus, we may refer to Section 7.3.4 for a discussion of the firms' optimal decisions to invest in breach prevention and detection.

Breach information sharing

Dashed lines within the interval $\phi_1 \leq \alpha \leq 1$ in Figure 7.2 (a) depict firms' security investment decisions if the regulator effectively enforces the breach notification law with sanctions $\sigma = .01$. In this interval, firms invest at the Nash equilibrium $(\tilde{x}_2, \tilde{d}_2)$ which depends on the information sharing effectiveness of the authority (cf. Equations (7.23) and (7.24)). At the reference point ϕ_1, investment in breach prevention and detection is at $x_2^* = .155$ and at $\tilde{d}_2 = .009$, respectively. These investments in preventive and detective controls both monotonically decrease in the effectiveness α. Thus, the maximum total investments in security controls are at $\tilde{x}_2 + \tilde{d}_2 = .164$. In the interval $\phi_1 \leq \alpha \leq 1$, firms always invest more in preventive than detective controls $\tilde{x}_2 > \tilde{d}_2$.

Dashed lines in Figure 7.2 (b) depicts firms' security investment decisions as a function of the regulator's introduced sanction level that effectively enforces information sharing, given that the authority's effectiveness is at $\alpha = .06$. The lowermost dashed line captures investment in detective controls $\tilde{d}_2$, and the uppermost dashed line marks the resulting sum of firms' investments $\tilde{x}_2 + \tilde{d}_2$. Observe from the constant distance between both dashed lines that firms' investment in preventive controls does not depend on the sanction level (cf. Equation (7.23)). Furthermore, we find that investment in detective controls increases in the collected sanctions (cf. Equation (7.24)). Thereby, firms' investment in detective controls is always higher than the corresponding optimal investment introduced by a social planner (cf. the lowermost dashed line and the lowermost gray line). And high sanctions may cause firms to over-invest in controls altogether (cf. the uppermost dashed line and the uppermost gray line). As only the investment in detective controls increases in the sanction level, high sanctions may change firms' investment priority from preventive to detective controls, i. e., $\tilde{x}_2 < \tilde{d}_2$.

Dotted lines in Figure 7.2 (b) depicts firms' investment decisions as a function of the regulator's introduced sanction level that effectively enforces information sharing, but given that the authority's effectiveness is at $\alpha = .14$. By comparing this scenario to the one in the previous paragraph, we find that a higher effectiveness of the authority results in overall lower security investments at firms (cf. the uppermost dashed line and the uppermost dotted line in the figure). This reproduces a substitution effect of cyber risk information sharing and security investments, previously observed in Section 3.2.2.

Welfare

The dashed line in Figure 7.3 (a) shows firms' generated profit at the Nash equilibrium as a function of the authority's information sharing effectiveness, assuming that the regulator requires sanctions $\sigma = .01$ to effectively enforce the breach notification law. In the effectiveness interval $0 \leq \alpha < \phi_1$, the regulator does not enforce breach information sharing and firms' profit is constant at $O^0(\tilde{x}_1, \tilde{x}_1, \tilde{d}_1, \tilde{d}_1, 0) = .712$. However, if he effectively enforces the law, i. e., in the interval $\phi_1 \leq \alpha \leq 1$, the profit of firms increases in the authority's effectiveness α (but at a decreasing rate, not visible in the figure).

The dotted line in Figure 7.3 (a) also depicts firms' profit at the Nash equilibrium as a function of the authority's effectiveness, but assuming that the regulator requires sanctions $\sigma = .05$ to effectively enforce the breach notification law. By comparing this scenario to the one in the previous paragraph, we find that the regulator requiring higher sanctions only effectively enforces the breach notification law to increase firms' profit if the authority's information sharing effectiveness is above a greater threshold $\phi_2 > \phi_1$.

Table 7.2 Effect of exogenous actions by the regulator or authority on investments.

Regime		Exogenous parameters	
		effectiveness	sanctions
Endogenous parameters		$\alpha \uparrow$	$\sigma \uparrow$
Baseline (without information sharing)			
preventive controls	$x_1^* = \tilde{x}_1$	$\rightarrow$	$\rightarrow$
detective controls	$d_1^* = \tilde{d}_1$	$\rightarrow$	$\rightarrow$
Social optimum (with information sharing)			
preventive controls	x_2^*	$\downarrow$	$\rightarrow$
detective controls	d_2^*	$\downarrow$	$\rightarrow$
social welfare	$O^1(x_2^*, x_2^*, d_2^*, d_2^*, 0)$	$\uparrow$	$\rightarrow$
Nash equilibrium (with information sharing)			
preventive controls	$\tilde{x}_2$	$\downarrow$	$\rightarrow$
detective controls	$\tilde{d}_2$	$\downarrow$	$\uparrow$
social welfare	$O^1(\tilde{x}_2, \tilde{x}_2, \tilde{d}_2, \tilde{d}_2, 1)$	$\uparrow$	$\downarrow$

The dashed line in Figure 7.3 (b) depicts firms' profit at the Nash equilibrium as a function of the regulator's required sanction level to incentivize breach information sharing, assuming the authority's effectiveness is at $\alpha = .06$. If the regulator requires rather low sanctions $0 < \sigma \leq \phi_1$ to incentivize information sharing, he effectively enforces the law and the Nash equilibrium is $(\tilde{x}_2, \tilde{d}_2)$. Firms' profit in this interval reveals that a lower sanction level is socially beneficial (cf. the dashed line and the lowermost solid gray line, capturing firms' profit without a law). Therefore, the regulator maximizes profit by setting sanctions on the minimum level required to (just) incentivize firms' breach information sharing. However, regardless of the sanction level that effectively enforces the notification law, firms' profit is always below the profit generated by a social planner (cf. the dashed line and the uppermost solid gray line). On the other hand, if the regulator requires rather high sanctions $\phi_1 < \sigma$ to incentivize breach information sharing, he abstains from enforcing the law and the equilibrium is $(\tilde{x}_1, \tilde{d}_1)$. This is because a sanction level above the reference point is socially detrimental: firms generate higher profit without an enforced disclosure regime.

The dotted line in Figure 7.3 (b) depicts firms' profit at the Nash equilibrium as a function of the regulator's required sanction level to incentivize information sharing, but assuming the authority's effectiveness is at $\alpha = .14$. By comparing this setup to the one in the last paragraph, we find that if the regulator effectively enforces the law, a higher effectiveness of the authority raises firms' profit (cf. the dashed and dotted line).

Table 7.3 Comparison of firms' total spendings with those of a social planner.

Condition	**Notation**	**Question a)**
Interval of sharing effectiveness	Social optimum, Nash equilibrium	Total security spendings
$0 \leq \alpha < \phi_0$	(x_1^*, d_1^*),$(\tilde{x}_1, \tilde{d}_1)$	$x_1^* + d_1^* = \tilde{x}_1 + \tilde{d}_1$
$\phi_0 \leq \alpha < \phi_{1,2}$	(x_2^*, d_2^*),$(\tilde{x}_1, \tilde{d}_1)$	$x_2^* + d_2^* > \tilde{x}_1 + \tilde{d}_1$
$\phi_{1,2} \leq \alpha \leq 1$ (S1)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$x_2^* + d_2^* > \tilde{x}_2 + \tilde{d}_2$
$\phi_{1,2} \leq \alpha \leq 1$ (S2)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$x_2^* + d_2^* \leq \tilde{x}_2 + \tilde{d}_2$

7.3.6 Results

We next provide a response to the research questions formulated in Section 7.1. The effects of our model parameters on the social optima and Nash equilibria are summarized in Table 7.2. This table indicates that most results depend on the authority's effectiveness. As this effectiveness is unknown in practice, we discuss all relevant scenarios and give the intervals for the effectiveness scale where specific results apply. We extend our explanation of results on the effect of sanctions where appropriate.

Total security spendings

Table 7.3 indicates how an effective enforcement of breach information sharing changes affected firms' total security spendings as compared to a social planner's optimal spendings. In the interval without need for breach information sharing, i. e., $0 \leq \alpha < \phi_0$, the sum of firms' investments and the total spendings of a planner are equal. By contrast, in case of a high effectiveness, i. e., $\phi_0 \leq \alpha < \phi_{1,2}$, total investments of firms are lower than those of a planner, who introduces information sharing. If the authority's information sharing effectiveness is in the interval $\phi_{1,2} \leq \alpha \leq 1$, firms may under- *or* over-invest in security. We refer to the two possible scenarios as scenario 1 (S1) and scenario 2 (S2). In scenario 1, the regulator effectively enforces breach information sharing with low sanctions. In turn, firms' total security investments are below the total investments of a social planner. In scenario 2, the regulator effectively enforces information sharing with high sanctions. This can potentially lead to security over-investments of firms.

Investment priority

Table 7.4 summarizes how an effective enforcement of breach information sharing changes affected firms' investment priority, as compared to a social planner's optimal decisions. In the interval without need for breach information sharing, i. e., $0 \leq \alpha < \phi_0$,

Table 7.4 Comparison of firms' investment priority with decisions of a social planner.

Condition	**Notation**	**Question b)**	
Interval of sharing effectiveness	Social optimum, Nash equilibrium	Investment priority	Preventive & detective security spending
$0 \leq \alpha < \phi_0$	(x_1^*, d_1^*),$(\tilde{x}_1, \tilde{d}_1)$	$x_1^* < d_1^*$,$\tilde{x}_1 < \tilde{d}_1$	$x_1^* = \tilde{x}_1$,$d_1^* = \tilde{d}_1$
$\phi_0 \leq \alpha < \phi_{1,2}$	(x_2^*, d_2^*),$(\tilde{x}_1, \tilde{d}_1)$	$x_2^* > d_2^*$,$\tilde{x}_1 < \tilde{d}_1$	$x_2^* > \tilde{x}_1$,$d_2^* < \tilde{d}_1$
$\phi_{1,2} \leq \alpha \leq 1$ (S1)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$x_2^* > d_2^*$,$\tilde{x}_2 > \tilde{d}_2$	$x_2^* > \tilde{x}_2$,$d_2^* < \tilde{d}_2$
$\phi_{1,2} \leq \alpha \leq 1$ (S2)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$x_2^* > d_2^*$,$\tilde{x}_2 < \tilde{d}_2$	$x_2^* > \tilde{x}_2$,$d_2^* < \tilde{d}_2$

both a social planner and firms prioritize investment in detective controls. At the same time, there is no difference in the security spending on single controls. By contrast, in case that the authority has a high information sharing effectiveness, i. e., $\phi_0 \leq \alpha < \phi_{1,2}$, firms' priority differs from the preference of a social planner, who introduces breach information sharing. Specifically, firms favor to invest in detective controls while a planner prioritizes investment in breach prevention. If the authority's information sharing effectiveness is in the interval $\phi_{1,2} \leq \alpha \leq 1$, it depends on the scenario whether or not firms and a social planner set different security investment priorities. In scenario 1, where the regulator effectively enforces breach information sharing with low sanctions, firms and a planner prioritize investment in preventive controls. In scenario 2, where the regulator effectively enforces compliance with high sanctions, firms are incentivized to prioritize investment in detective controls, differing from a social planner's optimal preference. In the interval where the planner introduces information sharing because of the authority's high effectiveness, i. e., $\phi_0 \leq \alpha \leq 1$, firms appear to under-invest in preventive and over-invest in detective controls.

Social welfare

Table 7.5 summarizes how firms' profit under diverse disclosure regimes compares to a social planner's optimal profit. In the interval without need for breach information sharing, i. e., $0 \leq \alpha < \phi_0$, firms invest at the socially optimal level. Therefore, they gain the same profit as a social planner does. If the authority's information sharing effectiveness is high, i. e., in the interval $\phi_0 \leq \alpha \leq 1$, firms under-invest in preventive controls and over-invest in detective controls, such that they generate less profit than a social planner. However, if the authority's effectiveness is in the interval $\phi_{1,2} \leq \alpha \leq 1$, such that the regulator effectively enforces mandatory breach information sharing, firms generate more profit as compared to the same setup without a disclosure regime.

Table 7.5 Comparison of firms' and a social planner's profit.

Condition	**Notation**	**Question c)**
Interval of sharing effectiveness	Social optimum, Nash equilibrium	Social welfare (profit)
$0 \leq \alpha < \phi_0$	(x_1^*, d_1^*),$(\tilde{x}_1, \tilde{d}_1)$	$O^0(x_1^*, x_1^*, d_1^*, d_1^*, 0) = O^0(\tilde{x}_1, \tilde{x}_1, \tilde{d}_1, \tilde{d}_1, 0)$
$\phi_0 \leq \alpha < \phi_{1,2}$	(x_2^*, d_2^*),$(\tilde{x}_1, \tilde{d}_1)$	$O^1(x_2^*, x_2^*, d_2^*, d_2^*, 0) > O^0(\tilde{x}_1, \tilde{x}_1, \tilde{d}_1, \tilde{d}_1, 0)$
$\phi_{1,2} \leq \alpha \leq 1$ (S1)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$O^1(x_2^*, x_2^*, d_2^*, d_2^*, 0) > O^1(\tilde{x}_2, \tilde{x}_2, \tilde{d}_2, \tilde{d}_2, 1)$
$\phi_{1,2} \leq \alpha \leq 1$ (S2)	(x_2^*, d_2^*),$(\tilde{x}_2, \tilde{d}_2)$	$O^1(x_2^*, x_2^*, d_2^*, d_2^*, 0) > O^1(\tilde{x}_2, \tilde{x}_2, \tilde{d}_2, \tilde{d}_2, 1)$

7.4 Discussion

Our game-theoretic model covers important characteristics of firms' trade-offs between investments in productive activity and security controls. Even though the model cannot fully represent reality, we can use results derived from its analysis to infer effects of effectively enforcing breach notification laws such as the NIS Directive on affected firms' incentives to invest in breach prevention and detection, presented in Section 7.4.1. These inferences may have some possible implications, discussed in Section 7.4.2. With respect to the inferences and their possible implications, a critical reflection of model limitations that call for future research is required, conducted in Section 7.4.3. Note that this section can be understood as a complement to our discussion in Section 6.4.

7.4.1 Inferences from our analysis

Based on our model setup, regulators do not enforce mandatory breach information sharing of firms with authorities in cases where the latter cannot effectively deal with received information. The model analysis predicts that if authorities' effectiveness is low, firms conduct investments in preventive and detective controls at optimal levels, i. e., comparable to levels that a social planner would introduce. With our exogenous parameter choice, these investments account for about 16.4 % of firms' total budget. This security budget is primarily allocated on detective rather than preventive controls.

The situation in which firms make optimal investments changes if regulators do not enforce mandatory breach information sharing even though authorities' effectiveness is high. In this scenario, firms do not deviate from decisions introduced in the last paragraph, but a social planner establishes information sharing and adapts his investments. Our model analysis predicts that a planner spends more than 16.4 % of firms' total budget on controls, such that there is security under-investment at firms.

Furthermore, a planner prioritizes investment in preventive over detective controls, which differs from the preference of firms. As a consequence, it turns out that firms under-invest in preventive controls, and over-invest in measures to detect breaches. These sub-optimal budget allocations result in profit below the socially optimal level.

Regulators effectively enforce firms' mandatory breach information sharing by audits and sanctions if the effectiveness of authorities is very high. Our model analysis predicts that in turn, affected firms adapt their investment decisions depending on the introduced sanction level, while a social planner does not deviate from decisions discussed in the last paragraph. In case that regulators effectively enforce information sharing with a low sanction level, firms under-invest in security controls. In this scenario, firms prioritize budget allocations on breach prevention rather than detection. At the same time, they under-invest in breach prevention and over-invest in detection. By contrast to the low sanction level scenario, if regulators use a high sanction level to effectively enforce mandatory information sharing, it is likely that firms over-invest in security. Thereby, they primarily allocate budget on detective rather than preventive controls. And as with the low sanction level scenario, firms under-invest in breach prevention and over-invest in detection. Based on our model setup, once regulators effectively enforce information sharing by a low or high sanction level, affected firms generate more profit as compared to the scenario without a disclosure regime, yet less than a social planner. The following principle is applicable to regulators: an effective enforcement of laws by a lower sanction level leads to higher profit at affected firms.

7.4.2 Possible implications

Our analysis suggests that laws which mandate breach information sharing with authorities, enforced by audits and a rather low sanction level, may incentivize affected firms to under-invest in security controls. This implies that enforced laws bear the risk of firms making their cyber risk management efforts dependent on the information shared by authorities. However, in reality, it is conceivable that these information are noise rather than signals, e. g., because informed authorities have a very low information sharing effectiveness. If firms misjudge and make use of such information to manage cyber risk, they may get a false conception for the security of their ICT systems. In order to avoid firms' plain substitution of security investments by engagement in information exchange, authorities have to clarify how useful their shared information really are. For instance, it is conceivable that they indicate the relevance of disseminated information for different types of recipients. In turn, firms may better judge whether received information can be used to improve cyber risk management efforts or not.

Furthermore, enforced laws incentivize affected firms to under-invest in preventive and over-invest in detective controls. Our model predicts that firms' incentives to under-invest in breach prevention are not influenced by the introduced audit probability or sanction level – contradicting the results in Chapter 6. This implies that firms affected by laws always take preventive measures that lead to insufficient ICT system security levels. On the other hand, our analysis indicates that firms' incentives to over-invest in breach detection exacerbate by an increase in expected sanctions. The explanation is intuitive: as initiated audits cannot differentiate malicious concealment of breaches from benign nescience, firms fear the threat of sanctions that may apply for undetected and thus unreported breaches; now, if higher sanctions have to be expected, this provides stronger incentives for firms to detect and report breaches. Yet, it is conceivable that at the time when firms detect, report, and initiate corrective actions for breaches, attackers already leveraged ICT systems to propagate attacks.[4] Thus, firms' over-investment in breach detection and under-investment in breach prevention leads to negative externalities that the laws intend to counter in the first place. We can imagine that regulators correct such failures of laws by introducing minimum standards for firms' preventive controls that – if met – lead to less sanctions in cases of non-compliance.

7.4.3 Limitations of the model

Caution is needed when transferring previous conclusions to the real world, as they are derived from a game-theoretic model that simplifies reality. Specifically, our model comprises two important simplifications that help us to obtain closed-form solutions for investment decisions of the two representative firms. Most notably, we assume that the firms have ICT systems whose security is interdependent, and that these systems are badly protected such that they generate negative externalities justifying regulatory intervention in form of a breach notification law. By capturing this interdependence, future models may be able to make more reliable inferences regarding firms' incentives to allocate budget on productive activity, preventive controls, and detective controls. Furthermore, our model implies that profit of the two firms is driven by constant return on investment in productive activity. A more realistic model may consider that profit of the firms depends on their competition in an oligopolistic market, e. g., Betrand or Cournot duopoly – depending on the types of goods the firms sell –, and capture that this competition is influenced by the authority's information disclosures to the public.

[4] In particular, this may hold truth for worm attacks – such as the recent "WannaCry worm" attack that breached and propagated between ICT systems all over the world in a very short amount of time (https://www.symantec.com/security_response/writeup.jsp?docid=2017-051310-3522-99).

Part III

Summary

Chapter 8

Conclusion

Conventional wisdom suggests that cyber risk information sharing is an easy and cheap way to give defenders an edge over attackers. All that needs to be done is to build the right technology and agree on a data model and exchange format. This view neglects that information sharing is less a technical than an economic problem. In the first part of this work, we reviewed literature that sheds light into the economics driving defenders' information sharing decisions. In the second part, we identified conditions under which effectively enforcing laws that mandate firms' breach information sharing with authorities improves welfare, and evaluated effects of enforcements on firms' incentives to prevent and detect breaches. This part of the dissertation summarizes our most important findings in Section 8.1, and presents an outlook on future laws that mandate defenders' cyber risk information sharing with authorities in Section 8.2.

8.1 Summary

Our first goal was to provide a consolidated understanding of defenders' incentives to share cyber risk information. In this context, we conducted a literature survey to identify enablers and disablers for information exchange. It turns out that voluntary private cyber risk information sharing among different kinds of defenders is primarily enabled by mechanism that help to build mutual trust over time, monetary compensations from information recipients, and expected cyber risk reductions. The latter also is an enabler for defenders' deliberate information disclosures to the public, while firms' voluntary announcements of cyber risk information in particular seem to be driven by expected increases in profit. However, defenders often engage less in voluntary information sharing than is socially desirable, mainly because such actions may result

in an increase of cyber risk or forgone profit. In order to counter lacking incentives, regulators can make use of mechanism to enforce laws that mandate defenders' cyber risk information exchange. Our survey reveals a research gap regarding studies on the economics of defenders' mandatory cyber risk information sharing with authorities.

Our second goal was to identify conditions under which regulators' effective enforcement of laws that mandate firms' breach information sharing with authorities improves welfare. For this purpose, we devised and analyzed a game-theoretic model. The model analysis indicates that to effectively enforce laws, regulators have to verify affected firms' compliance with breach reporting obligations and sanction non-compliance. However, regulators should only effectively enforce laws if three conditions are fulfilled simultaneously. First, the interdependence of affected firms' ICT system security is high. Otherwise, these systems may not generate sufficient negative externalities to justify government intervention in form of laws mandating information exchange. Second, affected firms' costs associated with breach information sharing obligations are low. In fact, if these disclosure costs exceed other losses that come along with ICT system breaches, effectively enforcing breach information sharing imposes burdens on firms that are almost certainly socially detrimental. Third, authorities are very effective in drawing and sharing conclusions from received information with others. In case that authorities are ineffective, their actions cannot lead to societal benefits that outweigh the burdens imposed on firms by the disclosure regimes. If the three conditions are fulfilled, regulators' effective enforcement of breach information sharing with authorities may improve welfare: it incentivizes affected firms to internalize negative externalities from their ICT systems, and promotes cyber risk information symmetry in economies.

Our third goal was to evaluate effects of effectively enforcing laws that mandate breach information sharing with authorities on affected firms' decisions to invest in breach prevention and detection. Therefore, we devised and analyzed a variant of the model used to achieve our second goal. This analysis indicates that if laws are enforced, expected sanctions for non-compliance drive affected firms' security investment decisions. Low expected sanctions incentivize security under-investments. Thereby, firms prioritize investment in breach prevention over detection. By contrast, high expected sanctions can lead to security over-investments. At the same time, firms prefer investment in breach detection over preventive measures. The reason is that high expected sanctions create strong incentives for firms to comply with the laws, which can only be achieved if breaches get detected in the first place. Overall, an effective enforcement of laws by low or high expected sanctions at firms may lead to societal benefits, if informed authorities can provide sufficient cyber risk management support to other defenders.

8.2 Outlook

Laws that mandate defenders' cyber risk information sharing with authorities are high on the policy agenda around the globe. Yet, to the best of our knowledge, regulators to date only enacted firms' breach information sharing with authorities. We subsequently use the framework dimensions from Chapter 2 to present an outlook on future laws with objectives presented in Section 5.1, and highlight some associated implications.

We expect future laws that mandate cyber risk information sharing with authorities to oblige firms rather than other defenders (*Who*). The reason is that one of these laws' objectives is to let affected parties internalize negative externalities from their ICT systems, and firms' systems potentially are responsible for very large amounts of negative externalities in economies. For instance, consider the impact of ICT system breaches at firms that operate critical national infrastructure, such as nuclear power plants, financial institutions, or telecommunications facilities. However, firms may already be affected by other types of laws with similar objectives, e. g., ex ante regulation (such as security standards), and ex post liability policies [Romanosky and Acquisti, 2009]. This calls for examinations of the status quo and joint economic effects of laws that aim to increase firms' ICT system security levels. Corresponding inquiries may help regulators to better trade-off the costs and benefits of new disclosure regimes.

We also expect that in the future, regulators will enforce laws that mandate firms' breach rather than other cyber risk information (*What*) sharing with authorities. The argument is based on the operating principles of notification laws to achieve their two main objectives. First, enforced cyber risk information sharing is supposed to provide incentives for affected firms to protect their ICT systems and thus evade costs associated with the disclosure regime. This dissertation indicates that of all defined information types (cf. Section 2.2), enforced vulnerability or breach information sharing potentially meets this requirement: authorities' disclosures of either information can lead to reputation damages at affected firms (cf. Sections 4.3.1 and 4.3.3). Second, reported information are supposed to be used by authorities to inform other defenders such that they can better manage their cyber risk. Yet, if vulnerability information get reported, authorities may have strong incentives to stockpile them for the use in "cyber wars", rather than forward them to affected vendors that can initiate cyber risk mitigations (cf. Section 3.2.1). By contrast, if breach information get reported, authorities appear to engage in information forwarding [Bisogni et al., 2017] which in turn may help recipients. These rationales indicate that regulators' enforcement of laws mandating firms' breach information sharing with authorities will likely work best.

However, laws that mandate firms' breach information sharing with authorities only (*Whom*), such as EU Directive 2016/1148, have to be put into question. In light that breaches at firms likely generate negative externalities affecting other defenders, why not mandate information sharing directly with those affected? After all, corresponding US laws appear to be effective (cf. Section 5.3). Against this backdrop, an approach that may be considered in the future – and already commonly gets implemented in the EU – is to oblige firms' breach information sharing with authorities *and* affected defenders [Anderson et al., 2008]. Such notification laws are accompanied by positive effects resulting from informed authorities' actions. Also, the laws most definitely empower affected defenders, allowing them to better manage cyber risk and "name and shame" firms by disclosing received information to the public. In turn, the approach establishes very effective incentives at affected firms to prevent breaches, and ensures that scholars can come up with good statistics on how pervasive breaches really are.

As regulators will likely enforce laws that mandate firms' cyber risk information sharing with authorities in the future, further theoretical and first empirical studies are required to investigate properties that characterize effective (*What Effect*) approaches, also indicated by Table 5.2. New theoretical studies may evaluate opportunities to promote firms' self-regulation and voluntary compliance with the laws [Winn, 2009]. For instance, it is reasonable that these studies inquire regulators' strategic use of political instruments such as subsidies, liabilities, and taxes to incentivize firms' information sharing with authorities. On the other hand, first empirical studies are much needed that shed light into determinants for effective informed authorities. For example, scholars may collaborate with competent authorities in EU member states that already receive and handle cyber risk information, such as the "Bundesamt für Sicherheit in der Informationstechnik" (BSI) in Germany, to assess the performance of their procedures used to draw and share conclusions from received reports with others in the economy.

After all, regulators should only bring in authorities to the race for cyber risk information if there is evidence that their actions give defenders an edge.

Bibliography

Ablon, L., Heaton, P., Lavery, D., and Romanosky, S. (2016). Consumer attitudes toward data breach notifications and loss of personal information. In *Workshop on the Economics of Information Security (WEIS)*, Berkeley, CA, USA.

Acquisti, A., Friedman, A., and Telang, R. (2006). Is there a cost to privacy breaches? An event study. In *Workshop on Economics of Information Security (WEIS)*, University of Cambridge, UK.

Acquisti, A. and Varian, H. R. (2005). Conditioning prices on purchase history. *Marketing Science*, 24(3):367–381.

Akerlof, G. (1970). The market for "Lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Anderson, R. (2001). Why information security is hard – An economic perspective. In *Annual Computer Security Applications Conference (ACSAC)*, New Orleans, LA, USA.

Anderson, R. (2008). *Security engineering: A guide to building dependable distributed systems*. Wiley, 2 edition.

Anderson, R., Böhme, R., Clayton, R., and Moore, T. (2008). Security economics and the internal market. Technical report, European Network and Information Security Agency (ENISA).

Anderson, R. and Moore, T. (2006). The economics of information security. *Science*, 314(5799):610–613.

Arora, A., Caulkins, J. P., and Telang, R. (2006a). Research note—Sell first, fix later: Impact of patching on software quality. *Management Science*, 52(3):465–471.

Arora, A., Forman, C., Nandkumar, A., and Telang, R. (2010a). Competition and patching of security vulnerabilities: An empirical analysis. *Information Economics and Policy*, 22(2):164–177.

Arora, A., Krishnan, R., Telang, R., and Yang, Y. (2010b). An empirical analysis of software vendors' patch release behavior: Impact of vulnerability disclosure. *Information Systems Research*, 21(1):115–132.

Arora, A., Nandkumar, A., and Telang, R. (2006b). Does information security attack frequency increase with vulnerability disclosure? An empirical analysis. *Information Systems Frontiers*, 8(5):350–362.

Arora, A., Telang, R., and Xu, H. (2008). Optimal policy for software vulnerability disclosure. *Management Science*, 54(4):642–656.

August, T., Dao, D., Laube, S., and Niculescu, F. (2017). Economics of ransomware attacks. In *Workshop on Information Systems and Economics (WISE)*, Seoul, South Korea.

August, T. and Tunca, T. I. (2008). Let the pirates patch? An economic analysis of software security patch restrictions. *Information Systems Research*, 19(1):48–70.

Bandyopadhyay, T., Mookerjee, V. S., and Rao, R. C. (2009). Why IT managers don't go for cyber-insurance products. *Communications of the ACM*, 52(11):68–73.

Barnum, S. (2012). Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™). Technical report, MITRE Corporation.

Beattie, S., Arnold, S., Cowan, C., Wagle, P., and Wright, C. (2002). Timing the application of security patches for optimal uptime. In *Proceedings of the USENIX Systems Administration Conference (LISA)*, pages 233–242, Philadelphia, PA, USA.

Bertino, E., Choo, K.-K. R., Georgakopolous, D., and Nepal, S. (2016). Internet of Things (IoT): Smart and secure service delivery. *ACM Transactions on Internet Technology (TOIT)*, 16(4):22.

Bisogni, F. (2016). Proving limits of state data breach notification laws: Is a federal law the most adequate solution? *Journal of Information Policy*, 6:154–205.

Bisogni, F., Asghari, H., and van Eeten, M. (2017). Estimating the size of the iceberg from its tip. In *Workshop on the Economics of Information Security (WEIS)*, San Diego, CA, USA.

Böhme, R. (2006). A comparison of market approaches to software vulnerability disclosure. In Müller, G., editor, *Emerging Trends in Information and Communication Security (ETRICS)*, volume 3995 of *Lecture Notes in Computer Science*, pages 298–311, Berlin Heidelberg. Springer.

Böhme, R. (2012). Security audits revisited. In Keromytis, A., editor, *Financial Cryptography and Data Security*, volume 7397 of *Lecture Notes in Computer Science*, pages 129–147, Berlin Heidelberg. Springer.

Böhme, R. and Laube, S. (2014). Das IT-Sicherheitsgesetz. In Baetge, J. and Kirsch, H.-J., editors, *Mittelstand im Blick: Compliance und Risikomanagement*, pages 17–36, Düsseldorf, Germany. IDW.

Böhme, R., Laube, S., and Riek, M. (2018). A fundamental approach to cyber risk analysis. *Variance*, 11(2).

Campbell, K., Gordon, L. A., Loeb, M. P., and Zhou, L. (2003). The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. *Journal of Computer Security*, 11:431–448.

Caralli, R. A., Stevens, J. F., Young, L. R., and Wilson, W. R. (2007). Introducing OCTAVE Allegro: Improving the information security risk assessment process. Technical report, Software Engineering Institute, Carnegie Mellon University.

Carr, N. G. (2003). IT doesn't matter. *Harvard Business Review*, May:5–12.

Cavusoglu, H., Cavusoglu, H., and Raghunathan, S. (2007). Efficiency of vulnerability disclosure mechanisms to disseminate vulnerability knowledge. *IEEE Transactions on Software Engineering*, 33(3):171–185.

Cavusoglu, H., Cavusoglu, H., and Zhang, J. (2008). Security patch management: Share the burden or share the damage? *Management Science*, 54(4):657–670.

Cavusoglu, H., Mishra, B., and Raghunathan, S. (2004a). A model for evaluating IT security investments. *Communications of the ACM*, 47(7):87–92.

Cavusoglu, H., Mishra, B., and Raghunathan, S. (2004b). 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):69–104.

Cetin, O., Jhaveri, M. H., Ganán, C., van Eeten, M., and Moore, T. (2016). Understanding the role of sender reputation in abuse reporting and cleanup. *Journal of Cybersecurity*, 2(1):83–98.

Choi, J. P., Fershtman, C., and Gandal, N. (2010). Network security: Vulnerabilities and disclosure policy. *The Journal of Industrial Economics*, LVIII(4):868–894.

Choi, S. and Johnson, M. E. (2017). Do hospital data breaches reduce patient care quality? In *Workshop on the Economics of Information Security (WEIS)*, San Diego, CA, USA.

Collins, M., Gates, C., and Kataria, G. (2006). A model for opportunistic network exploits: The case of P2P worms. In *Workshop on the Economics of Information Security (WEIS)*, University of Cambridge, UK.

Dacey, R. F. (2003). Progress made, but challenges remain to protect federal systems and the nation's critical infrastructures. Testimony.

Dekker, D. M., Karsberg, C., and Daskala, B. (2012). Cyber incident reporting in the EU – An overview of security articles in EU legislation. Technical report, European Union Agency for Network and Information Security (ENISA).

Deutscher Bundestag (2015). Gesetz zur Erhöhung der Sicherheit informationstechnischer Systeme (IT-Sicherheitsgesetz). *Bundesgesetzblatt*, I(31):1324–1331.

Durumeric, Z., Kasten, J., Adrian, D., Halderman, J. A., Bailey, M., Li, F., Weaver, N., Amann, J., Beekman, J., Payer, M., et al. (2014). The matter of heartbleed. In *Proceedings of the Internet measurement conference (IMC)*, pages 475–488, Vancouver, BC, Canada.

Edelman, B. (2011). Adverse selection in online "trust" certifications and search results. *Electronic Commerce Research and Applications*, 10(1):17–25.

Egelman, S., Herley, C., and van Oorschot, P. C. (2013). Markets for zero-day exploits: Ethics and implications. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, pages 41–46, Banff, AB, Canada.

ENISA (2015). Cyber security information sharing: An overview of regulatory and non-regulatory approaches. Technical report, European Union Agency For Network And Information Security (ENISA).

Finifter, M., Akhawe, D., and Wagner, D. (2013). An empirical study of vulnerability rewards programs. In *Proceedings of the USENIX Security Symposium*, pages 273–288, Washington, DC, USA.

Fischer-Hübner, S. (2001). IT-Security. In Goos, G., Hartmanis, J., and van Leeuwen, J., editors, *IT-Security and Privacy: Design and Use of Privacy-enhancing Security Mechanisms*, volume 1958 of *Lecture Notes in Computer Science*, pages 35–105, Berlin, Heidelberg. Springer.

Frei, S., Schatzmann, D., Plattner, B., and Trammell, B. (2010). Modeling the security ecosystem – The dynamics of (in)security. In Moore, T., Pym, D., and Ioannidis, C., editors, *Economics of Information Security and Privacy*, pages 79–106. Springer US.

Freiling, F. and Schwittay, B. (2007). A common process model for incident response and digital forensics. In Frings, S., Göbel, O., Günther, D., Hase, H. G., Nedon, J., Schadt, D., and Brömme, A., editors, *IMF 2007: IT-Incident Management & IT-Forensics*, volume 114 of *Lecture Notes in Informatics*, pages 13–40, Stuttgart, Germany. Gesellschaft für Informatik.

Gal-Or, E. and Ghose, A. (2005). The economic incentives for sharing security information. *Information Systems Research*, 16(2):186–208.

Gatzlaff, K. M. and McCullough, K. A. (2010). The effect of data breaches on shareholder wealth. *Risk Management and Insurance Review*, 13(1):61–83.

Gay, S. (2016). Strategic news bundling and privacy breach disclosures. In *Workshop on the Economics of Information Security (WEIS)*, Berkeley, CA, USA.

Gordon, L. A. and Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, 5(4):438–457.

Gordon, L. A., Loeb, M. P., and Lucyshyn, W. (2003). Sharing information on computer systems security: An economic analysis. *Journal of Accounting and Public Policy*, 22(6):461–485.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., and Zhou, L. (2015). The impact of information sharing on cybersecurity underinvestment: A real options perspective. *Journal of Accounting and Public Policy*, 34(5):509–519.

Gordon, L. A., Loeb, M. P., and Sohail, T. (2010). Market value of voluntary disclosures concerning information security. *MIS Quarterly*, 34(3):567–594.

Gordon, L. A., Loeb, M. P., and Zhou, L. (2011). The impact of information security breaches: Has there been a downward shift in costs? *Journal of Computer Security*, 19(1):33–56.

Hausken, K. (2007). Information sharing among firms and cyber attacks. *Journal of Accounting and Public Policy*, 26(6):639–688.

Hausken, K. (2015). A strategic analysis of information sharing among cyber attackers. *Journal of Information Systems and Technology Management*, 12(2):245–270.

Hausken, K. (2017). Information sharing among cyber hackers in successive attacks. *International Game Theory Review*, 19(02).

Hiller, J. S. and Russell, R. S. (2013). The challenge and imperative of private sector cybersecurity: An international comparison. *Computer Law & Security Review*, 29(3):236–245.

Hovav, A. and D'Arcy, J. (2004). The impact of virus attack announcements on the market value of firms. *Information Systems Security*, 13(3):32–40.

Ishiguro, M., Tanaka, H., Matsuura, K., and Murase, I. (2006). The effect of information security incidents on corporate values in the Japanese stock market. In *International Workshop on the Economics of Securing the Information Infrastructure (WESII)*, Washington, DC, USA.

Jhaveri, M. H., Cetin, O., Gañán, C., Moore, T., and van Eeten, M. (2017). Abuse reporting and the fight against cybercrime. *ACM Computing Surveys (CSUR)*, 49(4).

Kamien, M. I., Muller, E., and Zang, I. (1992). Research joint ventures and R&D cartels. *The American Economic Review*, 82(5):1293–1306.

Kampanakis, P. (2014). Security automation and threat information-sharing options. *IEEE Security & Privacy*, 12(5):42–51.

Kannan, K., Rees, J., and Sridhar, S. (2007). Market reactions to information security breach announcements: An empirical analysis. *International Journal of Electronic Commerce*, 12(1):69–91.

Kannan, K. and Telang, R. (2005). Market for software vulnerabilities? Think again. *Management Science*, 51(5):726–740.

Katti, S., Krishnamurthy, B., and Katabi, D. (2005). Collaborating against common enemies. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference (IMC)*, pages 365–378, Berkeley, CA, USA.

Khouzani, A. M. H. R., Pham, V., and Cid, C. (2014a). Incentive engineering for outsourced computation in the face of collusion. In *Workshop on the Economics of Information Security (WEIS)*, The Pennsylvania State University, PA, USA.

Khouzani, A. M. H. R., Pham, V., and Cid, C. (2014b). Strategic discovery and sharing of vulnerabilities in competitive environments. In Poovendran, R. and Saad, W., editors, *Decision and Game Theory for Security*, volume 8840 of *Lecture Notes in Computer Science*, pages 59–78. Springer International Publishing.

Kim, S., Zimmermann, T., and Nagappan, N. (2011). Crash graphs: An aggregated view of multiple crashes to improve crash triage. In *International Conference on Dependable Systems & Networks (DSN)*, Hong Kong, China.

Kirby, A. J. (1988). Trade associations as information exchange mechanisms. *The RAND Journal of Economics*, 29(1):138–146.

Ko, M. and Dorantes, C. (2006). The impact of information security breaches on financial performance of the breached firms: An empirical investigation. *Journal of Information Technology Management*, 17(2):13–22.

Koivunen, E. (2012). "Why wasn't I notified?": Information security incident reporting demystified. In Aura, T., Järvinen, K., and Nyberg, K., editors, *Information Security Technology for Applications*, volume 7127 of *Lecture Notes in Computer Science*, pages 55–70, Berlin, Heidelberg. Springer.

Kruidhof, O. (2014). Evolution of national and corporate CERTs – Trust, the key factor. In Hathaway, M. E., editor, *Best Practices in Computer Network Defense: Incident Detection and Response*, pages 81–96.

Kunreuther, H. and Heal, G. (2003). Interdependent security. *The Journal of Risk and Uncertainty*, 26(2/3):231–249.

Kwon, J. and Johnson, M. E. (2015). The market effect of healthcare security: Do patients care about data breaches? In *Workshop on the Economics of Information Security (WEIS)*, Delft University of Technology, The Netherlands.

Laffont, J.-J. and Martimort, D. (2002). *The theory of incentives: The principal-agent model*. Princeton University Press, Princeton, NJ, USA.

Landwehr, C. E., Bull, A. R., McDermott, J. P., and Choi, W. S. (1994). A taxonomy of computer program security flaws. *ACM Computing Surveys (CSUR)*, 26(3):211–254.

Lasswell, H. D. (1948). The structure and function of communication in society. In Bryson, L., editor, *The communication of ideas: A series of addresses*, pages 37–51, New York. Harper and Brothers.

Laszka, A., Felegyhazi, M., and Buttyan, L. (2014). A survey of interdependent information security games. *ACM Computing Surveys (CSUR)*, 47(2).

Laube, S. and Böhme, R. (2015a). Mandatory security information sharing with authorities: Implications on investments in internal controls. In *ACM Conference on Computer and Communication Security (ACM CCS), Workshop on Information Sharing and Collaborative Security (WISCS)*, Denver, CO, USA.

Laube, S. and Böhme, R. (2015b). Meldepflichten für IT-Sicherheitsvorfälle: Ein Prinzipal-Agent-Ansatz. In Thomas, O. and Teuteberg, F., editors, *Tagungsband Wirtschaftsinformatik*, pages 1146–1162, Osnabrück, Germany.

Laube, S. and Böhme, R. (2015c). The economics of mandatory security breach reporting to authorities. In *Workshop on the Economics of Information Security (WEIS)*, Delft University of Technology, The Netherlands.

Laube, S. and Böhme, R. (2016). The economics of mandatory security breach reporting to authorities. *Journal of Cybersecurity*, 2(1):29–41.

Laube, S. and Böhme, R. (2017). Strategic aspects of cyber risk information sharing. *ACM Computing Surveys*, 50(5).

Li, P. and Rao, H. R. (2007). An examination of private intermediaries' roles in software vulnerabilities disclosure. *Information Systems Frontiers*, 9(5):531–539.

Lippmann, R., Haines, J. W., Fried, D. J., Korba, J., and Das, K. (2000). The 1999 DARPA off-line intrusion detection evaluation. *Computer networks*, 34(4):579–595.

Liu, D., Ji, Y., and Mookerjee, V. (2011). Knowledge sharing and investment decisions in information security. *Decision Support Systems*, 52(1):95–107.

Macho-Stadler, I. and Pérez-Castrillo, D. (2009). *Principal-agent models*, volume 1, pages 6977–6990. Springer, New York.

Machuletz, D., Sendt, H., Laube, S., and Böhme, R. (2016). Users protect their privacy if they can: Determinants of webcam covering behavior. In *Proceedings of the 1st European Workshop on Usable Security (EuroUSEC)*, Darmstadt, Germany.

MacKinlay, A. C. (1997). Event studies in economics and finance. *Journal of Economic Literature*, 35(1):13–39.

Maillart, T. and Sornette, D. (2010). Heavy-tailed distribution of cyber-risks. *The European Physical Journal B*, 75(3):357–364.

Maillart, T., Zhao, M., Grossklags, J., and Chuang, J. (2016). Given enough eyeballs, all bugs are shallow? Revisiting Eric Raymond with bug bounty programs. In *Workshop on the Economics of Information Security (WEIS)*, Berkeley, CA, USA.

Meng, G., Liu, Y., Zhang, J., Pokluda, A., and Boutaba, R. (2015). Collaborative security: A survey and taxonomy. *ACM Computing Surveys (CSUR)*, 48(1).

Metcalf, L. and Spring, J. M. (2014). Blacklist ecosystem analysis update: 2014. Technical report, Carnegie Mellon University.

Mitra, S. and Ransbotham, S. (2015). Information disclosure and the diffusion of information security attacks. *Information Systems Research*, 26(3):565–584.

Moody, D. L. and Walsh, P. (1999). Measuring the value of information – An asset valuation approach. In *European Conference on Information Systems*, Copenhagen, Denmark.

Moore, T. and Clayton, R. (2008). The consequence of non-cooperation in the fight against phishing. In *APWG eCrime Researchers Summit*, Atlanta, GA, USA.

Moore, T. and Clayton, R. (2011). The impact of public information on phishing attack and defense. *Communications and Strategies*, 1(81):45–68.

Moore, T., Friedman, A., and Procaccia, A. D. (2010). Would a 'cyber warrior' protect us: Exploring trade-offs between attack and defense of information systems. In *Proceedings of the New Security Paradigms Workshop (NSPW)*, pages 85–94, Concord, MA, USA.

Moores, T. (2005). Do consumers understand the role of privacy seals in e-commerce? *Communications of the ACM*, 48(3):86–91.

Naghizadeh, P. and Liu, M. (2016). Inter-temporal incentives in security information sharing agreements. In *AAAI-16 Workshop on Artificial Intelligence for Cyber Security (AICS)*, Phoenix, AZ, USA.

Nash, J. F. (1950). *Non-cooperative games*. PhD thesis, Princeton University, NJ, USA.

National Concil of ISACs (2017). The reach of Information Sharing and Analysis Centers (ISAC). Technical report, National Concil of Information Sharing and Analysis Centers (ISACs).

NIST (2011). Specification for asset identification 1.1. Technical Report 7693, National Institute of Standards and Technology (NIST).

NIST (2012). Guide for conducting risk assessments. Technical Report 800-30 Rev 1, National Institute of Standards and Technology (NIST).

Nizovtsev, D. and Thursby, M. (2007). To disclose or not? An analysis of software user behavior. *Information Economics and Policy*, 19(1):43–64.

Öğüt, H., Cavusoglu, H., and Raghunathan, S. (2008). Intrusion-detection policies for IT security breaches. *INFORMS Journal on Computing*, 20(1):112–123.

Öğüt, H., Memon, N., and Raghunathan, S. (2005). Cyber insurance and IT security investment: Impact of interdependent risk. In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, MA, USA.

Osborne, M. J. (2003). *An introduction to game theory*. Oxford University Press, Oxford.

Ozment, A. (2007). Improving vulnerability discovery models. In *ACM Conference on Computer and Communication Security (ACM CCS), Workshop on Quality of protection*, pages 6–11, Alexandria, VA, USA.

Ozment, A. and Schechter, S. E. (2006). Milk or wine: Does software security improve with age? In *Proceedings of the USENIX Security Symposium*, pages 93–104, Vancouver, BC, Canada.

Polinsky, A. M. and Shavell, S. (2007). The theory of public enforcement of law. *Handbook of law and economics*, 1:403–454.

Ransbotham, S. and Mitra, S. (2013). The impact of immediate disclosure on attack diffusion and volume. In Schneier, B., editor, *Economics of information security and privacy III*, pages 1–12, New York. Springer.

Ransbotham, S., Mitra, S., and Ramsey, J. (2012). Are markets for vulnerabilities effective? *MIS Quarterly*, 36(1):43–64.

Rescorla, E. (2003). Security holes . . . who cares. In *USENIX Security Symposium*, Washington, DC, USA.

Romanosky, S. (2016). Examining the costs and causes of cyber incidents. *Journal of Cybersecurity*.

Romanosky, S. and Acquisti, A. (2009). Privacy costs and personal data protection: Economic and legal perspectives. *Berkeley Technology Law Journal*, 24(3):1061–1101.

Romanosky, S., Hoffman, D., and Acquisti, A. (2014). Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, 11(1):74–104.

Romanosky, S., Sharp, R., and Acquisti, A. (2010). Data breaches and identity theft: When is mandatory disclosure optimal? In *Workshop on the Economics of Information Security (WEIS)*, Harvard University, MA, USA.

Romanosky, S., Telang, R., and Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30(2):256–286.

Samuelson Law, Technology & Public Policy Clinic (2007). Security breach notification laws: Views from chief security officers. Technical report, University of California, Berkeley School of Law.

Schechter, S. E. (2004). *Computer security strength & risk: A quantitative approach*. PhD thesis, Harvard University, MA, USA.

Schneier, B. (2000). *Secret & lies: Digital security in a networked world*, chapter 21, pages 318–333. John Wilesy & Sons.

Shahzad, M., Shafiq, M. Z., and Liu, A. X. (2012). A large scale exploratory analysis of software vulnerability life cycles. In *Proceedings of the International Conference on Software Engineering (ICSE)*, pages 771–781, Zürich, Switzerland.

Shapiro, C. and Varian, H. R. (1998). *Information rules: A strategic guide to the network economy*. Harvard Business Review Press.

Skopik, F., Settanni, G., and Fiedler, R. (2016). A problem shared is a problem halved: A survey on the dimensions of collective cyber defense through security information sharing. *Computers & Security*, 60:154–176.

Statistisches Bundesamt (2013). *Statistisches Jahrbuch Deutschland und Internationales*. Statistisches Bundesamt, Wiesbaden.

Swire, P. P. (2004). A model for when disclosure helps security: What is different about computer and network security. *Journal on Telecommunications and High Technology Law*, 3:163–208.

Tang, Q., Linden, L., Quarterman, J. S., and Whinston, A. B. (2013). Improving Internet security through social information and social comparison: A field quasi-experiment. In *Workshop on Economics of Information Security (WEIS)*, Washington, DC, USA.

Telang, R. and Wattal, S. (2007). An empirical analysis of the impact of software vulnerability announcements on firm stock price. *IEEE Transactions on Software Engineering*, 33(8):544–557.

Thomas, K., Amira, R., Ben-yoash, A., Folger, O., Hardon, A., Berger, A., Bursztein, E., and Bailey, M. (2016). The abuse sharing economy: Understanding the limits of threat exchanges. In Monrose, F., Dacier, M., Blanc, G., and Garcia-Alfaro, J., editors, *Research in Attacks, Intrusions, and Defenses*, volume 9854 of *Lecture Notes in Computer Science*, pages 143–164. Springer International Publishing.

Varian, H. R. (2002). System reliability and free riding. In Camp, L. J. and Lewis, S., editors, *Economics of Information Security*, volume 12 of *Advances in Information Security*, pages 1–15. Springer US.

Vasek, M. and Moore, T. (2012). Do malware reports expedite cleanup? An experimental study. In *Workshop on Cyber Security Experimentation and Test*, Bellevue, WA, USA.

Vasek, M., Weeden, M., and Moore, T. (2016). Measuring the impact of sharing abuse data with web hosting providers. In *ACM Conference on Computer and Communication Security (ACM CCS), Workshop on Information Sharing and Collaborative Security (WISCS)*, Vienna, Austria.

von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wagner, C., Dulaunoy, A., Wagener, G., and Iklody, A. (2016). MISP: The design and implementation of a collaborative threat intelligence sharing platform. In *ACM Conference on Computer and Communication Security (ACM CCS), Workshop on Information Sharing and Collaborative Security (WISCS)*, Vienna, Austria.

Wang, T., Kannan, K. N., and Ulmer, J. R. (2013). The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, 24(2):201–218.

West-Brown, M., Stikvoort, D., Kossakowski, K.-P., Killcrece, G., Ruefle, R., and Zajicek, M. (2003). Handbook for Computer Security Incident Response Teams (CSIRTs). Technical report, Carnegie Mellon Software Engineering Institute.

Winn, J. K. (2009). Are 'better' security breach notification laws possible? *Berkeley Technology Law Journal*, 24(3):1133–1165.

Zhao, M., Grossklags, J., and Liu, P. (2015). An empirical study of web vulnerability discovery ecosystems. In *ACM Conference on Computer and Communication Security (ACM CCS)*, Denver, CO, USA.

Zou, C. C., Gong, W., and Towsley, D. (2002). Code red worm propagation modeling and analysis. In *ACM Conference on Computer and Communication Security (ACM CCS)*, Washington, DC, USA.

Glossary

All subsequently listed acronyms are defined when they first appear in this dissertation. Some of the acronyms appear only once in this work. Nevertheless, we include them in this glossary as they are assumed to be common knowledge by many security experts. The provided definitions are not intended to be comprehensive.

Acronym	
APWG	Anti-Phishing Working Group; an industry association with the aim to unify the global response to cybercrime.
BSI	Bundesamt für Sicherheit in der Informationstechnik; the federal office for information security in Germany.
CAR	Cumulative Abnormal Returns; the sum of abnormal returns of a stock: differences between expected returns and actual returns. This is a common metric of the event study method.
CC	The Common Criteria; a standard for certifying ICT systems.
CERT	Computer Emergency Response Team; a type of non-profit organization that, among other things, provides an expert group to handle cyber risk information sharing.
CVE	Common Vulnerabilities and Exposure; a standard to reference publicly known vulnerabilities.
CVSS	Common Vulnerability Scoring System; a standard to score the cyber risk associated with vulnerabilities.
CybOX	Cyber Observable eXpression; a standard that can be used to encode events that (potentially) occur in ICT systems.
DHS	The Department of Homeland Security.

EU	The European Union.
FBI	The Federal Bureau of Investigation.
GDPR	General Data Protection Regulation; an EU regulation that aims to harmonize and unify existing EU privacy breach reporting obligations.
GLBA	Gramm–Leach–Bliley Act; an act that mandates firms in the US financial sector to inform their primary federal regulator and sometimes individuals on privacy breaches.
H.R.	The House of Representatives; the lower chamber of the United States Congress.
HIPAA	Health Insurance Portability and Accountability Act; an act that mandates firms in the US health care sector to report breaches of health information to affected individuals.
HITECH	Health Information Technology for Economic and Clinical Health Act; an amendment to HIPAA.
ICT	Information and Communication Technology; an umbrella term for all kinds of communication devices or applications, including: computer, hardware, and software. An ICT system consists of hardware, software, data, and the individuals that use it.
IDMEF	Intrusion Detection Message Exchange Format; a standard to share cyber risk information between all kinds of security controls.
IDS	Intrusion Detection System; a device or software application that is able to detect successful attacks.
IOC	Indicator of Compromise; artifacts observed on ICT systems that – with a high probability – indicate a successful attack.
IODEF	Incident Object Description Exchange Format; a standard that provides a framework for the communication of operational and statistical cyber risk information.

IoT	Internet of Things; the connection of physical devices, embedded with software, to the Internet.
ISAC	Information Sharing and Analysis Center; a type of non-profit organization that provides, among other things, resources for cyber risk information sharing.
ISAO	Information Sharing and Analysis Organization; a broader term for ISAC.
ITRC	The Identity Theft Resource Center; a non-profit organization that supports individuals in managing their cyber risk.
MAEC	Malware Attribute Enumeration and Characterization; a standard that enables the description of malware based on its attributes.
NCSL	The National Conference of State Legislatures; a non-governmental organization in the US that serves members and staff of state legislatures.
NIS	Network and Information Security.
NVD	National Vulnerability Database; a repository for information about vulnerabilities.
PRC	The Privacy Rights Clearinghouse; a non-profit organization informing defenders on best practices to strengthen privacy.
PSD2	Revised Directive on Payment Services; an EU directive aiming to protect defenders when they pay online.
S.	United States Senate; the upper chamber of the United States Congress.
SEC	The Securities and Exchange Commission.
STIX	Structured Threat Information eXpression; a standard that defines a structure for machine-processable storage, analysis, and sharing of cyber risk information.

TAXII	Trusted Automated eXchange of Indicator Information; a standard that provides an automated transport mechanism for cyber risk information expressed with STIX.
URL	Uniform Resource Locator; the global address of a website on the Internet.
US	The United States of America.

Appendix A

Proof sketches for Chapter 6

A.1 Social optima

A.1.1 Investment in preventive controls

The first derivation of Equation (6.10) w. r. t. x is

$$\frac{\partial O}{\partial x} = [\gamma \cdot H(s^*) \cdot (1 - P(x)) + (1 - \gamma \cdot H(s^*) \cdot P(x))] \cdot I(s^*, 0) \cdot \frac{\partial P(x)}{\partial x} + 1 \ . \quad \text{(A.1)}$$

The root of the above first-order condition $\partial O / \partial x = 0$ is

$$x^*(s^*) = -\frac{\log\left(\frac{\gamma \cdot H(s^*)+1}{4 \cdot \gamma \cdot H(s^*)} - \sqrt{\frac{(\gamma \cdot H(s^*)+1)^2}{16 \cdot \gamma^2 \cdot H(s^*)^2} - \frac{1}{2 \cdot \gamma \cdot \log(\theta) \cdot H(s^*) \cdot I(s^*,0)}}\right)}{\log(\theta)} \ . \quad \text{(A.2)}$$

This expression corresponds to Equation (6.11).

A.1.2 Breach reporting and proof of Lemma 6.3.1

Lemma. *For any optimal investment x^*, if $\alpha > 0$, $\gamma > 0$, $\lambda_2 > 0$, and $\epsilon > 0$, a reporting strategy $0 < s < 1$ is not socially optimal. Under these conditions, the socially optimal reporting strategy is a boundary value, i. e., $s^*(x^*) \in \{0, 1\}$.*

Proof. The first derivation of Equation (6.10) w. r. t. s is

$$\frac{\partial O}{\partial s} = (1-\epsilon) \cdot P(x^*) \cdot ((1-P(x^*)) \cdot (\gamma \cdot \lambda_2 \cdot H(s) - \alpha \cdot \gamma \cdot I(s)) + \lambda_2) \ . \tag{A.3}$$

A further derivation of the above equation $\partial^2 O / \partial s^2$ leads to

$$\frac{\partial^2 O}{\partial s^2} = -2 \cdot \alpha \cdot \gamma \cdot \lambda_2 \cdot (1-\epsilon)^2 \cdot (1-P(x^*)) \cdot P(x^*) \ . \tag{A.4}$$

Based on Equation (A.4), we observe that, for $\alpha > 0$, $\gamma > 0$, $\lambda_2 > 0$, $\epsilon > 0$ and any $x^*(\cdot)$, $\partial^2 c / \partial s^2 < 0$. Therefore, the cost function in Equation (6.10) is concave in s, and $s^*(x^*) \in \{0, 1\}$ are boundary values. □

A.2 Nash equilibria

A.2.1 Investments in preventive controls

The first derivation of Equation (6.13) w. r. t. x_i is

$$\frac{\partial O_i}{\partial x_i} = (1 - \gamma \cdot H(s_{1-i}) \cdot P(x_{1-i})) \cdot I_i(s_i, a) \cdot \frac{\delta P(x_i)}{\delta x_i} + 1 \ . \tag{A.5}$$

The root of this first-order condition $\partial O_i / \partial x_i = 0$, i. e., the best response of firm i, is

$$x_i^+(x_{1-i}, s_i, s_{1-i}, a) = \sup \left\{ -\frac{\log\left(\frac{1}{\log(\theta) \cdot I(s_i,a) \cdot \left(1 - \gamma \cdot H(s_{1-i}) \cdot \theta^{-x_{1-i}}\right)}\right)}{\log(\theta)}, 0 \right\} . \tag{A.6}$$

The mutual best responses $\tilde{x}(\tilde{s}, a) = x_i^+(\tilde{x}, \tilde{s}, \tilde{s}, a)$ lead to investments at Nash equilibria. Note that because of the constraint in Equation (6.13), there also exists a corner case if $0 = x_i^+(0, \tilde{s}, \tilde{s}, a)$. Thus, equilibria can imply the investments in preventive controls

$$\tilde{x}_{1,2}(\tilde{s}, a) = -\frac{\log\left(\frac{1}{2 \cdot \gamma \cdot H(\tilde{s})} \pm \sqrt{\frac{1}{4 \cdot \gamma^2 \cdot H(\tilde{s})^2} - \frac{1}{\gamma \cdot \log(\theta) \cdot H(\tilde{s}) \cdot I(\tilde{s},a)}}\right)}{\log(\theta)} \ , \tag{A.7}$$

$$\tilde{x}_3(\tilde{s}, a) = 0 \ . \tag{A.8}$$

This corresponds to Equations (6.14) and (6.15).

A.2.2 Proof of Lemma 6.3.3

Lemma. *If the Nash equilibrium implying $\tilde{x}_1(\tilde{s}, a)$ exists, then two other equilibria that contain the investments in preventive controls $\tilde{x}_{2,3}(\tilde{s}, a)$ exist simultaneously. Thereby, it holds that $\tilde{x}_3(\tilde{s}, a) = 0 \leq \tilde{x}_1(\tilde{s}, a) \leq \tilde{x}_2(\tilde{s}, a)$. Moreover, there are settings where only the equilibrium with investment $\tilde{x}_2(\tilde{s}, a)$ or $\tilde{x}_3(\tilde{s}, a) = 0$ persist.*

Proof. The amount of present equilibria depends on the discriminant in Equation (6.14). If the discriminant is negative, e. g., because:

$$\gamma > \frac{\log(\beta) \cdot I(\tilde{s}, a)}{4 \cdot H(s)} \ , \tag{A.9}$$

then only one equilibrium implying $\tilde{x}_3(\tilde{s}, a) = 0$ exists. Otherwise, the two equilibria with investments $\tilde{x}_{1,2}(\tilde{s}, a)$ may exist additionally, where $\tilde{x}_3(\tilde{s}, a) \leq \tilde{x}_1(\tilde{s}, a) \leq \tilde{x}_2(\tilde{s}, a)$. Based on Equations (A.6) and (A.7), the three equilibria implying investments in breach prevention $\tilde{x}_{1,2,3}(\tilde{s}, a)$ can exist simultaneously under the conditions

$$-\frac{\log\left(\frac{1}{2\cdot\gamma\cdot H(\tilde{s})} + \sqrt{\frac{1}{4\cdot\gamma^2\cdot H(\tilde{s})^2} - \frac{1}{\gamma\cdot\log(\theta)\cdot H(\tilde{s})\cdot I(\tilde{s},a)}}\right)}{\log(\theta)} \geq 0 \quad \text{, and} \tag{A.10}$$

$$-\frac{\log\left(\frac{1}{\log(\theta)\cdot I(s_i,a)\cdot(1-\gamma\cdot H(s_{1-i})\cdot\theta^{-0})}\right)}{\log(\theta)} \leq 0 \ . \tag{A.11}$$

Both conditions are fulfilled iff

$$\gamma \geq \frac{\log(\theta) \cdot I(\tilde{s}, a) - 1}{\log(\theta) \cdot H(\tilde{s}) \cdot I(\tilde{s}, a)} \ . \tag{A.12}$$

Equations (A.9) and (A.12) can also be solved for all other parameter and the audit probability: this probability and all parameter influence the existence of all equilibria.

Based on the previous constraints, we can differentiate between three categories of model parameter settings that lead to different Nash equilibria: (1) an equilibrium implying $\tilde{x}_3(\tilde{s}, a)$ exist alone iff Equation (A.9) holds; (2) the equilibria with investments in breach prevention $\tilde{x}_{1,2,3}(\tilde{s}, a)$ all exist simultaneously if Equation (A.12) holds, but not Equation (A.9); and (3) the equilibrium implying $\tilde{x}_2(\tilde{s}, a)$ exists alone iff Equation (A.12) does not hold. □

A.2.3 Breach reporting and proof of Lemma 6.3.4

Lemma. *If breach reporting does not associate disclosure costs $\lambda_2 = 0$, only Nash equilibria where firms voluntarily report breaches exist $\tilde{s} = 1$. Otherwise, in case that $\lambda_2 > 0$, only equilibria implying that firms do not report breaches exist $\tilde{s} = 0$, unless a security audit probability $a \geq a_{\min} = \lambda_2/(\lambda_2 + \sigma)$ is introduced.*

Proof. The first derivation of Equation (6.13) w. r. t. s_i is

$$\frac{\partial O_i}{\partial s_i} = \underbrace{P_i(x_i, x_{1-i}, s_{1-i})}_{>0} \cdot \underbrace{(1-\epsilon) \cdot (\lambda_2 - a \cdot (\lambda_2 + \sigma))}_{\text{sign depends on } a,\ \sigma, \text{ and } \lambda_2} . \tag{A.13}$$

If $a = 0 \wedge \lambda_2 > 0$, a firm does not have incentives to report breaches, i. e., $\partial O_i/\partial s_i > 0$. Consequently, Nash equilibria implying $\tilde{s}(\tilde{x}, 0) = 0$ evolve. Otherwise, if $a = 0 \wedge \lambda_2 = 0$, firms are indifferent to reporting, i. e., $\partial O_i/\partial s_i = 0$. Based on our assumptions on firms' reporting behavior in Section 6.2.3, this leads to compliance. If $\lambda_2 > 0$, a regulator can incentivize firms to report breaches with the initiation of audits $a > 0$, which may lead to $\partial O_i/\partial s_i \leq 0$. In order to determine the regulator's minimum audit probability $a = a_{\min}$ that incentivizes breach reporting of firms, we use the second part of Equation (A.13) that depends on a, σ and λ_2, i. e.,

$$0 = (1-\epsilon) \cdot (\lambda_2 - a_{\min} \cdot (\lambda_2 + \sigma)) \text{ , such that} \tag{A.14}$$

$$\Leftrightarrow a_{\min} = \frac{\lambda_2}{\lambda_2 + \sigma} . \tag{A.15}$$

A regulator can incentivize firms to report breaches with the introduction of an audit probability $a \geq a_{\min} = \lambda_2/(\lambda_2 + \sigma)$. Consequently, Nash equilibria implying the reporting strategy $\tilde{s}(\tilde{x}, a) = 1$ evolve if the audit probability is high enough. □

A.2.4 Proof of Lemma 6.3.5

Lemma. *If the sanction level is positive $\sigma > 0$, social costs always increase in the audit probability except for the case where this probability incites $\tilde{s} = 1$.*

Proof. The first derivation of Equation (6.17), w. r. t. a, is

$$\frac{\partial O}{\partial a} = 2 \cdot P_i(x_i, x_{1-i}, s_{1-i}) \cdot ((1 - s_i) \cdot (1-\epsilon) \cdot (\lambda_2 + \sigma) + \epsilon \cdot (\lambda_2 + \sigma)) . \tag{A.16}$$

Based on Equation (A.16), if the sanction level is positive $\sigma > 0$, and except for the case where the audit probability incites $\tilde{s} = 1$, we find that

$$\frac{\partial O}{\partial a} > 0 \ . \tag{A.17}$$

□

A.2.5 Proof of Lemma 6.3.6

Lemma. *If $\lambda_2 > 0$, $\sigma > 0$, and $\tilde{a} = a_{\min}$, the audit probability at all Nash equilibria $\tilde{a}$ decreases in the sanction level σ and increases in firms' disclosure costs λ_2.*

Proof. In Appendix A.2.3, we derived the audit probability to incentivize reporting of firms, i. e., $a_{\min} = \lambda_2/(\lambda_2 + \sigma)$. The first derivations of $a_{\min}$, w. r. t. σ and λ_2, are

$$\frac{\partial a_{\min}}{\partial \sigma} = -\frac{\lambda_2}{(\lambda_2 + \sigma)^2} \quad \text{, and} \tag{A.18}$$

$$\frac{\partial a_{\min}}{\partial \lambda_2} = \frac{\sigma}{(\lambda_2 + \sigma)^2} \ . \tag{A.19}$$

Thus, if $\lambda_2 > 0$ and $\sigma > 0$, we find that $\partial a_{\min}/\partial \sigma < 0$ and $\partial a_{\min}/\partial \lambda_2 > 0$. □

Appendix B

Proof sketches for Chapter 7

B.1 Social optima

B.1.1 Without breach information sharing $s = 0$

The first derivations of Equation (7.13) w. r. t. x and d are

$$\frac{\partial O^0}{\partial x} = \theta^{-x} \cdot \log(\theta) \cdot ((\lambda_1 - \lambda_3) \cdot F(d,d) + \lambda_3) - \rho \text{ , and} \tag{B.1}$$

$$\frac{\partial O^0}{\partial d} = \vartheta^{-d} \cdot \log(\vartheta) \cdot (\lambda_3 - \lambda_1) \cdot P(x,x) - \rho \text{ .} \tag{B.2}$$

The roots of the above conditions, $\partial O^0/\partial x = 0$ and $\partial O^0/\partial d = 0$, are

$$x = \frac{\log\left(\frac{\log(\theta)\cdot((\lambda_1-\lambda_3)\cdot F(d,d)+\lambda_3)}{\rho}\right)}{\log(\theta)} \text{ , and} \tag{B.3}$$

$$d = \frac{\log\left(\frac{\log(\vartheta)\cdot(\lambda_3-\lambda_1)\cdot P(x,x)}{\rho}\right)}{\log(\vartheta)} \text{ .} \tag{B.4}$$

Solving these two equations simultaneously results in the social optimum with investment in preventive and detective controls, respectively,

$$x_1^* = \frac{\log\left(-\frac{\lambda_1\cdot\log(\theta)\cdot\log(\vartheta)}{\rho\cdot\log(\theta)-\rho\cdot\log(\vartheta)}\right)}{\log(\theta)} \text{ , and} \tag{B.5}$$

$$d_1^* = \frac{\log\left(\frac{(\lambda_1-\lambda_3)\cdot(\log(\theta)-\log(\vartheta))}{\lambda_1\cdot\log(\theta)}\right)}{\log(\vartheta)} \text{ .} \tag{B.6}$$

These equations correspond to Equations (7.14) and (7.15).

B.1.2 With breach information sharing $s = 1$

The first derivations of Equation (7.16) w. r. t. x and d are

$$\frac{\partial O^1}{\partial x} = \frac{(1+\alpha)\cdot(\lambda_3\cdot\log(\theta)+(\lambda_1+\lambda_2-\lambda_3)\cdot\log(\theta)\cdot F(d,d))-\rho\cdot\theta^{\alpha\cdot x+x}}{\theta^{(1+\alpha)\cdot x}} \text{ , and} \tag{B.7}$$

$$\frac{\partial O^1}{\partial d} = \frac{-(1+\alpha)\cdot(\lambda_1+\lambda_2-\lambda_3)\cdot\log(\vartheta)\cdot P(x,x)-\rho\cdot\vartheta^{\alpha\cdot d+d}}{\vartheta^{(1+\alpha)\cdot d}} \ . \tag{B.8}$$

The roots of the above conditions, $\partial O^1/\partial x = 0$ and $\partial O^1/\partial d = 0$, are

$$x = \frac{\log\left(-\frac{(\alpha+1)\cdot\log(\theta)\cdot(-\lambda_3-(\lambda_1+\lambda_2-\lambda_3)\cdot F(d,d))}{\rho}\right)}{(\alpha+1)\cdot\log(\theta)} \text{ , and} \tag{B.9}$$

$$d = \frac{\log\left(-\frac{(\alpha+1)\cdot(\lambda_1+\lambda_2-\lambda_3)\cdot\log(\vartheta)\cdot P(x,x)}{\rho}\right)}{(\alpha+1)\cdot\log(\vartheta)} \ . \tag{B.10}$$

Solving these two equations simultaneously results in the social optimum with investment in preventive and detective controls, respectively,

$$x_2^* = \frac{\log\left(-\frac{(\alpha+1)\cdot\log(\theta)\cdot\log(\vartheta)\cdot(\lambda_1+\lambda_2)}{\rho\cdot(\log(\theta)-\log(\vartheta))}\right)}{(\alpha+1)\cdot\log(\theta)} \text{ , and} \tag{B.11}$$

$$d_2^* = \frac{\log\left(\frac{(\log(\theta)-\log(\vartheta))\cdot(\lambda_1+\lambda_2-\lambda_3)}{\log(\theta)\cdot(\lambda_1+\lambda_2)}\right)}{(\alpha+1)\cdot\log(\vartheta)} \ . \tag{B.12}$$

These equations correspond to Equations (7.17) and (7.18).

B.2 Nash equilibria

B.2.1 Without mandatory breach information sharing $s = 0$

The first derivations of Equation (7.19) w. r. t. x_i and d_i are

$$\frac{\partial O_i^0}{\partial x_i} = \theta^{-x_i}\cdot\log(\theta)\cdot((\lambda_1-\lambda_3)\cdot F_i(d_i,d_{1-i})+\lambda_3)-\rho \text{ , and} \tag{B.13}$$

$$\frac{\partial O_i^0}{\partial d_i} = \vartheta^{-d_i}\cdot\log(\vartheta)\cdot(\lambda_3-\lambda_1)\cdot P_i(x_i,x_{1-i})-\rho \ . \tag{B.14}$$

The roots of the first-order conditions of Equation (7.19), i. e., the best response of firm i $\partial O_i^0/\partial x_i = 0$ and $\partial O_i^0/\partial d_i = 0$, are

$$x_i^+(d_i, d_{1-i}) = \frac{\log\left(\frac{\log(\theta)\cdot((\lambda_1-\lambda_3)\cdot F_i(d_i,d_{1-i})+\lambda_3)}{\rho}\right)}{\log(\theta)} \text{ , and} \tag{B.15}$$

$$d_i^+(x_i, x_{1-i}) = \frac{\log\left(\frac{\log(\vartheta)\cdot(\lambda_3-\lambda_1)\cdot P_i(x_i,x_{1-i})}{\rho}\right)}{\log(\vartheta)} \text{ .} \tag{B.16}$$

Solving the last two equations simultaneously results in the Nash equilibrium with investment in preventive and detective controls, respectively,

$$\tilde{x}_1 = \frac{\log\left(-\frac{\lambda_1\cdot\log(\theta)\cdot\log(\vartheta)}{\rho\cdot\log(\theta)-\rho\cdot\log(\vartheta)}\right)}{\log(\theta)} \text{ , and} \tag{B.17}$$

$$\tilde{d}_1 = \frac{\log\left(\frac{(\lambda_1-\lambda_3)\cdot(\log(\theta)-\log(\vartheta))}{\lambda_1\cdot\log(\theta)}\right)}{\log(\vartheta)} \text{ .} \tag{B.18}$$

These equations correspond to Equations (7.20) and (7.21).

B.2.2 With mandatory breach information sharing $s = 1$

The first derivations of Equation (7.22) w. r. t. x_i and d_i are

$$\frac{\partial O_i^1}{\partial x_i} = P_i(x_i, x_{1-i}) \cdot \log(\theta) \cdot (\lambda_2 + \lambda_3 + \sigma - (\lambda_3 - \lambda_1 + \sigma) \cdot F_i(d_i, d_{1-i})) - \rho \text{ , and} \tag{B.19}$$

$$\frac{\partial O_i^1}{\partial d_i} = (1 - F_i(d_i, d_{1-i})) \cdot \log(\vartheta) \cdot (\lambda_3 - \lambda_1 + \sigma) \cdot P_i(x_i, x_{1-i}) - \rho \text{ .} \tag{B.20}$$

The roots of the first-order conditions of Equation (7.22), i. e., the best response of firm i $\partial O_i^1/\partial x_i = 0$ and $\partial O_i^1/\partial d_i = 0$, are

$$x_i^+(x_{1-i}, d_i, d_{1-i}) = \frac{\log\left(\frac{\log(\theta)\cdot\left((\lambda_1+\lambda_2)\cdot\vartheta^{\alpha\cdot d_{1-i}+d_i}-\lambda_1+\lambda_3+\sigma\right)}{\rho\cdot\vartheta^{\alpha\cdot d_{1-i}+d_i}\cdot\theta^{\alpha\cdot x_{1-i}}}\right)}{\log(\theta)} \text{ , and} \tag{B.21}$$

$$d_i^+(x_i, x_{1-i}, d_{1-i}) = \frac{\log\left(\frac{\left(P_i(x_i,x_{1-i})\cdot\vartheta^{-d_{1-i}\cdot\alpha}\cdot\log(\vartheta)\cdot(\sigma-\lambda_1+\lambda_3\right)}{\rho}\right)}{\log(\vartheta)} \text{ .} \tag{B.22}$$

Based on the mutual best responses $\tilde{x}(\tilde{d}) = x_i^+(\tilde{x}, \tilde{d}, \tilde{d})$ and $\tilde{d}(\tilde{x}) = d_i^+(\tilde{x}, \tilde{x}, \tilde{d})$, the Nash equilibrium has to satisfy

$$\tilde{x}(\tilde{d}) = \frac{\log(\lambda_2 + \lambda_3 + \sigma - F(\tilde{d}, \tilde{d}) \cdot (\lambda_3 - \lambda_1 + \sigma)) + \log(\log(\theta)) - \log(\rho)}{(\alpha + 1) \cdot \log(\theta)} \text{ , and} \tag{B.23}$$

$$\tilde{d}(\tilde{x}) = \frac{\log(P(\tilde{x}, \tilde{x})) + \log(\lambda_3 - \lambda_1 + \sigma) + \log(\log(\vartheta)) - \log(\rho)}{(\alpha + 1) \cdot \log(\vartheta)} \text{ .} \tag{B.24}$$

Solving these two equations simultaneously results in the Nash equilibrium with investment in preventive and detective controls, respectively,

$$\tilde{x}_2 = \frac{\log\left(-\frac{\log(\theta) \cdot \log(\vartheta) \cdot (\lambda_1 + \lambda_2)}{\rho \cdot (\log(\theta) - \log(\vartheta))}\right)}{(\alpha + 1) \cdot \log(\theta)} \text{ , and} \tag{B.25}$$

$$\tilde{d}_2 = \frac{\log\left(\frac{(\log(\theta) - \log(\vartheta)) \cdot (\lambda_1 - \lambda_3 - \sigma)}{\log(\theta) \cdot (\lambda_1 + \lambda_2)}\right)}{(\alpha + 1) \cdot \log(\vartheta)} \text{ .} \tag{B.26}$$

These equations correspond to Equations (7.23) and (7.24).

Declaration

This dissertation "Cyber Risk Information Sharing with Authorities" is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text.

Ich versichere an Eides statt, dass ich die eingereichte Dissertation "Cyber Risk Information Sharing with Authorities" selbständig verfasst habe. Anderer als der von mir angegebenen Quellen und Hilfsmittel habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht. Auch wurde die Dissertation nicht bereits anderweitig als Prüfungsarbeit vorgelegt.

Stefan Laube
November 2017