Tyler Moore
David J. Pym
Christos Ioannidis
Editors

Economics of Information Security and Privacy

Springer

Economics of Information Security and Privacy

Tyler Moore • David J. Pym • Christos Ioannidis
Editors

Economics of Information Security and Privacy

Springer

Editors
Dr. Tyler Moore
Harvard University
Center for Research on
Computation and Society
33 Oxford St.
Cambridge, MA 02138
USA
tmoore@seas.harvard.edu

Prof. Christos Ioannidis
University of Bath
Department of Economics
Claverton Down
Bath BA2 7AY
United Kingdom
ci200@bath.ac.uk

Prof. David J. Pym
School of Natural and Computing Sciences
University of Aberdeen
King's College
Aberdeen AB24 3UE
United Kingdom
d.j.pym@abdn.ac.uk

ISBN 978-1-4419-6966-8 e-ISBN 978-1-4419-6967-5
DOI 10.1007/978-1-4419-6967-5
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010931088

Preface

The Workshop on the Economics of Information Security was established in 2002 to bring together computer scientists and economists to understand and improve the poor state of information security practice. WEIS was borne out of a realization that security often fails for non-technical reasons. Rather, the incentives of both defender and attacker must be considered. Earlier workshops have answered questions ranging from finding optimal levels of security investement to understanding why privacy has been eroded. In the process, WEIS has attracted participation from the diverse fields such as law, management and psychology. WEIS has now established itself as the leading forum for interdisciplinary scholarship on information security.

The eigth installment of the conference returned to the United Kingdom, hosted by University College London on June 24-25, 2009. Approximately 100 researchers, practitioners and government officials from across the globe convened in London to hear presentations from authors of 21 peer-reviewed papers, in addition to a panel and keynote lectures from Hal Varian (Google), Bruce Schneier (BT Counterpane), Martin Sadler (HP Labs), and Robert Coles (Merrill Lynch). Angela Sasse and David Pym chaired the conference, while Christos Ioannidis and Tyler Moore chaired the program committee.

We are very grateful for the service of the WEIS program committee: Alessandro Acquisti(Carnegie Mellon University, USA), Ross Anderson (University of Cambridge, UK), Rainer Böhme (Technische Universität Dresden, Germany), Jean Camp (Indiana University, USA), Huseyin Cavusoglu (University of Texas at Dallas, USA), Nicolas Courtois (University College London, UK), Neil Gandal (Tel Aviv University, Israel), Larry Gordon (University of Maryland, USA), Eric Johnson (Dartmouth College, USA), Marty Loeb (University of Maryland, USA), Tyler Moore (Harvard University, USA), Andrew Odlyzko (University of Minnesota, USA), Andy Ozment (Office of the Secretary of Defense, USA), David Pym (HP Labs Bristol & University of Bath, UK), M. Angela Sasse (University College London, UK), Stuart Schechter (Microsoft Research, USA), Bruce Schneier(BT Counterpane, USA), Rahul Telang (Carnegie-Mellon University, USA), and Catherine Tucker (Massachusetts Institute of Technology, USA).

We are also grateful to Adam Beautement and JJ Giwa at UCL for their assistance in organizing local arrangements. Finally, we thank HP Labs, the UK Economic and Social Research Council and Unisys for their kind sponsorship.

Cambridge, Massachusetts, USA — *Tyler Moore*
Bath, United Kingdom — *David Pym*
January 2010 — *Christos Ioannidis*

List of Contributors

Ross Anderson
Computer Laboratory, University of Cambridge, UK, e-mail: Ross.Anderson@cl.cam.ac.uk

Stefan Berthold
Karlstads Universitet, Fakulteten för Ekonomi, Kommunikation och IT, Karlstad, Sweden, e-mail: stefan.berthold@kau.se

Rainer Böhme
International Computer Science Institute, Berkeley, CA, USA, e-mail: rainer.boehme@icsi.berkeley.edu

Joseph Bonneau
Computer Laboratory, University of Cambridge, UK, e-mail: jcb82@cl.cam.ac.uk

Ruth Breu
Institute of Computer Science, University of Innsbruck, Austria, e-mail: ruth.breu@uibk.ac.at

Huseyin Cavusoglu
School of Management, The University of Texas at Dallas, Richardson, TX, USA, e-mail: huseyin@utdallas.edu

Asunur Cezar
Middle East Technical University, Ankara, Turkey, e-mail: asunur@metu.edu.tr

Nicolas Christin
CyLab, Carnegie Mellon University, Pittsburgh, PA, USA, e-mail: nicolasc@andrew.cmu.edu

Richard Clayton
Computer Laboratory, University of Cambridge, UK, e-mail: richard.

clayton@cl.cam.ac.uk

Ramón Compañó
European Commission - Directorate General Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), e-mail: Ramon.compano@ec.europa.eu

Wainer Lusoli
European Commission - Directorate General Joint Research Centre (JRC), Visiting Research Fellow, University of Chester, UK, e-mail: Wainer.lusoli@ec.europa.eu

Mark Felegyhazi
International Computer Science Institute, Berkeley, CA, USA, e-mail: mark@icsi.berkeley.edu

Stefan Frei
Communication Systems Group, ETH Zurich, Switzerland, e-mail: frei@techzoom.net

Shailendra Fuloria
Computer Laboratory, University of Cambridge, UK, e-mail: Shailendra.Fuloria@cl.cam.ac.uk

Sören Preibusch
Computer Laboratory, University of Cambridge, UK, e-mail: sdp36@cl.cam.ac.uk

Dinei Florêncio
Microsoft Research, Redmond, WA, USA, e-mail: dinei@microsoft.com

Makoto Goto
Graduate School of Economics and Business Administration, Hokkaido University, Sapporo, Japan, e-mail: goto@econ.hokudai.ac.jp

Jens Grossklags
Center for Information Technology Policy, Princeton University, NJ, USA e-mail: jensg@princeton.edu

Cormac Herley
Microsoft Research, Redmond, WA, USA, e-mail: cormac@microsoft.com

Jeffrey Hunker
Jeffrey Hunker Associates, e-mail: hunker@jeffreyhunker.com

Frank Innerhofer–Oberperfler
Institute of Computer Science, University of Innsbruck, Austria, e-mail: frank.innerhofer-oberperfler@uibk.ac.at

Christos Ioannidis
University of Bath, UK, e-mail: c.ioannidis@bath.ac.uk

Benjamin Johnson

CyLab, Carnegie Mellon University, Pittsburgh, PA, USA, e-mail: johnsonb@andrew.cmu.edu

Tyler Moore
Center for Research on Computation & Society, Harvard University, Cambridge, MA, USA, e-mail: tmoore@seas.harvard.edu

Bernhard Plattner
Communication Systems Group, ETH Zurich, Switzerland, e-mail: plattner@tik.ee.ethz.ch

Christian W. Probst
Technical University of Denmark, e-mail: probst@imm.dtu.dk

David Pym
HP Labs, Bristol, UK & University of Bath, UK, e-mail: david.pym@cantab.net

Srinivasan Raghunathan
School of Management, The University of Texas at Dallas, Richardson, TX, USA, e-mail: sraghu@utdallas.edu

Dominik Schatzmann
Communication Systems Group, ETH Zurich, Switzerland, e-mail: schatzmann@tik.ee.ethz.ch

Galina Schwartz
University of California at Berkeley, CA, USA, e-mail: schwartz@eecs.berkeley.edu

Vicente Segura
Department of Network and Services Security, Telefonica I+D, Spain, e-mail: vsg@tid.es

Nikhil Shetty
University of California at Berkeley, CA, USA, e-mail: nikhils@eecs.berkeley.edu

Ken-ichi Tatsumi
Faculty of Economics, Gakushuin University, Tokyo, Japan, e-mail: Kenichi.Tatsumi@gakushuin.ac.jp

Brian Trammell
Hitachi Europe, ICTL Secure Systems Team, Zurich, Switzerland, e-mail: trammell@tik.ee.ethz.ch

Jean Walrand
University of California at Berkeley, CA, USA, e-mail: wlr@eecs.berkeley.edu

Contents

Chapter 1
Introduction and Overview

Tyler Moore, David Pym, and Christos Ioannidis

1.1 Introduction

The Workshop on the Economics of Information Security (WEIS) is the leading forum for interdisciplinary research and scholarship on information security and privacy, combining ideas, techniques, and expertise from the fields of economics, social science, business, law, policy, and computer science.

In 2009, WEIS was held in London, at UCL, a constituent college of the University of London. The papers included in this volume were all presented at WEIS 2009, having been carefully reviewed by a program committee composed of leading researchers. The presented papers were grouped into nine sessions — identity theft, modeling uncertainty's effects, future directions in the economics of information security, economics of privacy, options, misaligned incentives in systems, cyber-insurance, modeling security dynamics — and we follow the same organization in this volume.

The program of WEIS 2009 included four Keynote Addresses, described below in order of presentation.

- Hal Varian (Google and UC Berkeley): 'Computer Mediated Transactions'. Starting from a discussion of computed-mediated transactions placed in an historical context, Hal Varian gave an insight into his current thinking on topics such as accountability, trust, and enforceability in computer-mediated contracts and transactions. He emphasized the importance of customization and personalization, trade-offs enhanced service and privacy, and the expected impact of cloud computing.

Tyler Moore
Harvard University, USA, e-mail: tmoore@seas.harvard.edu

David Pym
HP Labs, Bristol and University of Bath, UK, e-mail: david.pym@hp.com

Christos Ioannidis
University of Bath, UK, e-mail: c.ioannidis@bath.ac.uk

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_1,

- Bruce Schneier (BT Counterpane): 'Security and Human Behavior'. Bruce Schneier gave a lively presentation of his view of the rôle of psychology in information security. He emphasized that psychological issues and perspectives impact upon security design, risk perception, attack strategies, and usability. He described how security is both a 'feeling' and a 'reality', and discussed the significance and detection of differences between feeling and reality,
- Martin Sadler (HP Labs, Bristol): 'From Mathematical Modeling to Automation: Turning Research on the Economics of Security into Economic Value'. Martin Sadler gave an illuminating discussion of some key research with HP's Systems Security Lab. He explained why modeling, from a range of economic and mathematical perspectives, is necessary if the management of information security investments is to evolve into a more rigorous engineering science. He emphasized the need for innovative thinking to support effective interactions between researchers and practitioners, be they in academia, industry and commerce, or government.
- Robert Coles (Bank of America): 'Information Security — Art or Science?'. Robert Coles gave an informative and reflective account of practical concerns of a CISO in large organization. He described the strategic issues and budgeting processes, in the contexts of constraints such as Basel II, and standards such as ISO27001 and COBIT. He explained how a range of example incidents might be handled, and, from the perspective of developing a science of information security, the difficulties involved in obtaining useful data.

The program also included a Panel Session, entitled 'A Broader View of Cyber Security Economics'. The panel members were Lance Hoffman (George Washington University, USA), Shari Lawrence Pfleeger (RAND Corporation), David Good (University of Cambridge, UK), Ann Cavoukian (Information and Privacy Commissioner, Province of Ontario, Canada), and Alessandro Acquisti (Carnegie Mellon University, USA).

The discussion in the Panel Session was primarily concerned with privacy, with a focus on the concerns of the citizen in dealing with the state, particularly in the context of the provisions of services using cloud computing.

1.2 The Economics of Information Security and Privacy

Since its inception in 2002, the research appearing at WEIS has steadily evolved. In its formative years, papers at WEIS used economics to explain longstanding challenges to information security: concepts such as incentives, game theory, externalities and information asymmetries were applied to solve information security problems. Models of security investment, patch management and privacy-enhancing technology adoption were presented. The topics discussed at WEIS have continued to evolve. The scientific organizers of WEIS 2009 (Ioannidis, Moore, Pym, and Sasse) encouraged submissions emphasizing two particular aspects:

- First, a whole-systems view of information systems security problems, including human behavior, technological aspects of systems, and the integration of these with the economic environments within which systems exist and operate. Understanding the incentive structures that operate across the spectrum of information security actors is essential, ranging from attackers to those with system-level defensive-responsibilities to the remaining majority of users.
- Second, the use of empirical methods (including data collection, analysis and simulation methods) to support economic and social studies of information security policies and technologies.

One theme common throughout the papers comprising the conference program has been use of methods of risk assessment and control, of the kind found in economics and finance, to model threats to information security as well as the timing and size of investments in information security operations.

Several included papers investigate privacy. The association between the concerns and management of privacy and information security is quite strong. Privacy is a valuable commodity, so individuals and organizations are are increasingly prepared to invest to protect it. Moreover, privacy is a matter of increasing concern for both national- and international-level regulators, as mandatory disclosure of privacy breaches has been adopted in many US states and is under consideration in Europe and beyond.

There is an inherent contradiction between the development of the web, which encourages the sharing of information, and stated preferences for privacy. Social networking sites sites solicit and distribute private information, but must (or, at least, should) take account the implicit requirements for privacy protection assumed by the providers of information.

The relationship of privacy to security is also important. For example, the users of social networking sites choose to promote the availability of the information that they own, but they also expect that the operators of the site will maintain its integrity and protect the confidentiality of their information by restricting access to legitimate members of the site. When citizens disclose private information to governments, by contrast, a different set of incentives, preferences, and trade-offs apply.

Informed by the perspective sketched above, we next introduce the papers included in this volume.

1.3 Overview of the Book's Contributions

The subsequent chapters in the book are broken down in seven sections, as indicated in Table 1.1.

The first two chapters consider the key role uncertainty plays when making security decisions. Uncertainty can take several forms — whether a firm will be targeted by attack, whether an attack will be successful or even detected, and so on. Firms are presented with a choice of spending money on security defenses and taking a small but certain loss, or skipping investment into protection but face an uncertain

Table 1.1 Chapters within each major section.

Modeling Uncertainty's Effects
The Price of Uncertainty in Security Games
Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty & the Underground Economy
Misaligned Incentives in Computing Systems
Security Economics and Critical National Infrastructure
Internet Multi-Homing Problems: Explanations from Economics
Threat Assessment
Modeling the Security Ecosystem — The Dynamics of (In)Security
Modeling the Economic Incentives of DDoS attacks: Femtocell Case Study
Privacy
The Privacy Jungle: On the Market for Data Protection in Social Networks
The Policy Maker's Anguish: Regulating Personal Data Behavior
Applications of Options Theory
Valuating Privacy with Option Pricing Theory
Optimal Timing of Information Security Investment: A Real Options Approach
Cyber Insurance
Competitive Cyber-Insurance and Internet Security
Potential Rating Indicators for Cyberinsurance: An Exploratory Qualitative Study
Risk Analysis
The Risk of Risk Analysis-And its relation to the Economics of Insider Threats
Competition, Speculative Risks and IT Security Outsourcing

but potentially higher loss. In Chapter 2, Grossklags, Johnson and Christin use game theory to compute the 'price of uncertainty' under different scenarios: the difference between the worst expected payoff under incomplete information and the payoff with complete information. In Chapter 3, Herley and Florêncio consider uncertainty from the perspective of miscreants participating on the underground economy. They note that stolen banking credentials are sold on the open market for a pennies on the dollar, and argue that such a low market-clearing price reflects a severe lemons market. Would-be fraudsters cannot tell whether the stolen credentials they are buying are real or fake. Chapter 3 also points to a more fundamental uncertainty, namely, the size and scope of the underground economy itself.

The next two chapters point out the existence of misaligned incentives for two computing applications. In Chapter 4, Anderson and Fuloria examine economic factors affecting critical national infrastructures, such as the IT systems that control power plants and chemical refineries. They argue that the systemic weaknesses present in these infrastructures can be explained by economic factors rather than technical ones. In Chapter 5, Richard Clayton examines the networking protocol SHIM6. He argues convincingly that SHIM6, while perfectly suitable from a technical perspective, is likely to be a commercial failure because the designers have failed to consider that would-be adopters have no short-term incentives to do so.

In Chapter 6, Frei, Schatzmann, Plattner, and Trammell continue the trend of arguing that poor incentives, flawed design and neglected implementation explain why security fails. They use empirical analysis to inform a model of organizational

response to vulnerability disclosure. The authors develop a methodology of how the 'security ecosystem' can be analyzed quantitatively using statistical analysis.Segura and Lahuerta develop and estimate an economic model of extortion by distributed denial-of-service (DDoS) attacks in Chapter 7. They put forward a model of the behavior of attackers as agents responding to economic incentives. Collecting relevant data on advertised prices in underground fora, the authors compute econometric estimates of the model parameters to understand the risk posed by DDoS attacks and develop appropriate responses. On the basis of their econometric evidence, the authors undertake a number of scenario simulations to explore the solution space for mitigating extortion attacks in the context of a wireless telecommunications service.Chapters 8 and 9 address the issue of privacy preservation from the points of view of policy makers and individual agents. A great deal of private information is lodged under the stewardship of social networks. In Chapter 8, Bonneau and Preibusch, undertake a comprehensive study of practices and policies with respect to the protection of privacy. They evaluate 45 social networking sites operating in a vigorously competitive environment as they seek additional users for their services. They find evidence that some of these sites are making a sustained effort to protect the privacy of their users, albeit with substantial diversity in the policies adopted. Privacy protection does not constitute a strong selling point and the relevant policies are often not conveyed clearly to the users. They conclude by suggesting a model of a privacy communication game, where the sites reveal in detail their privacy protection protocols to avoid criticism while hiding the privacy control interface to attract users.

In Chapter 9, Compañó and Lusoli conduct a cross-country on-line survey to identify differences in attitudes towards electronic identity across Europe. They identify a number of behavioral paradoxes which present governments with policy dilemmas. Most important is what the authors coin *the privacy paradox*, whereby young people happily provide a range of personal information despite awareness to the privacy risks. In protecting their privacy the participants reveal that they are asking for 'technology' that provides them with the means to protect their own identity data. At the same time, however, they are not confident in their expertise and organizational skills to keep their personal data safe and safeguard their privacy. Policy action faces systemic constraints as there is an apparent internal contradiction between the promotion and protection of privacy as public good and private sector interests. Cultural differences across EU Member States add to the difficulties of sound policy formation.

Chapter 10 proposes an innovative way to assess the long-term value of maintaining the privacy of personal data in the face of uncertainty. Berthold and Böhme apply a methodological tool from finance, the binomial option pricing model, to the valuation of personal information. By varying uncertainty surrounding future information disclosures as well as the agents' behavioral characteristics, the authors develop scenarios for the efficient pricing of 'privacy options'. In contrast to standard financial theory, the authors postulate stationary stochastic processes regarding the change in the value of information. The key idea of this work is that disclosing a single attribute value can be interpreted as writing an option for exploiting the at-

tribute in the future. Here, 'to exploit' refers to the act of using the attribute to draw inference on the data subject's identity or preference, and to base decisions on this information that may affect the data subject.

'Project valuation using real options' is a modern capital-budgeting technique that accounts for the explicit uncertainty inherent in all investment projects. Unlike traditional methods, where project managers are 'passive' once the investment decisions have been made, real options assume that managers maintain an active role in decisions regarding both the timing and size of the investment. In Chapter 11, Tatsumi and Goto apply the real options approach to firms' investment activity in information security technology and then provide a dynamic analysis of information security investment. They conclude that the statistical characteristics of threats to information security are important determinants of both the size and the timing of investment expenditure. Improvements in efficiency of vulnerability reduction technology encourages firms to undertake investment in information security earlier rather than later.

The question of how competitive cyber-insurers affect network security and welfare of the networked society is addressed by Shetty, Schwartz, Felegyhazi, and Walrand in Chapter 12. The existence of an efficient equilibrium insurance contract depends upon two different assumptions regarding the ability of the cyber-insurer to monitor the user's security. If insurers are unable to monitor the security of their clients, then under the existence of asymmetric information, it is unlikely that efficient cover contracts can be written. When user security is enforceable and costless, full-cover contracts are available. From numerical simulations, they conclude that a competitive cyber-insurance market improves both the welfare of users and network security.

Continuing the cyber-insurance theme in Chapter 13, Innerhofer-Oberperfler and Breu explore in a qualitative study the suitability of different rating indicators. The paper summarizes 36 semi-structured interviews with experts from Germany, Austria and Switzerland. The study's goal is to establish rating indicators for cyber-insurance, rank their importance, and explore the relationships between them. An initial list of 198 indicators is reduced to 94, which have subsequently been ranked in terms of importance by 29 experts. The results of this undertaking can aid cyber-insurers seeking to refine their premium-rating models with additional variables.

In Chapter 14, Probst and Hunker answer the question why organizations, given the significance of insider threats, choose policies that allow insider threats to proliferate. They examine how an organization's risk analysis and its assessment of trust in an insider develop over time. Insiders acquire more knowledge and vital information about the organization over time and thereby pose increasing risk. Faced with such a dynamic process, organizations have two choices for mitigation. The organization could choose to enhance the trust placed in insiders as they pose a bigger risk; alternatively, the organization must enforce an ever-increasing number of policies to regulate and monitor the insider's actions. The authors argue that since insiders with malicious intentions can alter their behavior to circumvent control systems, prevention becomes prohibitively expensive.

In the final chapter, Cezar, Cavusoglu and Raghunathan establish conditions for the efficient outsourcing of IT security. They show that the suitability of IT security outsourcing depends crucially upon the firm's perception of competitive externalities. When competitive externalities are ignored, firms choose to outsource provided the Managed Security Service Provider (MSSP) offers a quality (or a cost) advantage over in-house operations. However, when firms take externalities into account the quality improvement is no longer required for outsourcing to make sense. Even if the likelihood of a breach is higher under the MSSP, the expected benefit from the competitive demand externality may offset the loss from the higher likelihood of breaches.

Chapter 2
The Price of Uncertainty in Security Games

Jens Grossklags, Benjamin Johnson, and Nicolas Christin

Abstract In the realm of information security, lack of information about other users' incentives in a network can lead to inefficient security choices and reductions in individuals' payoffs. We propose, contrast and compare three metrics for measuring the *price of uncertainty* due to the departure from the payoff-optimal security outcomes under complete information. Per the analogy with other efficiency metrics, such as the price of anarchy, we define the price of uncertainty as the maximum discrepancy in expected payoff in a complete information environment versus the payoff in an incomplete information environment. We consider *difference*, *payoff-ratio*, and *cost-ratio* metrics as canonical nontrivial measurements of the price of uncertainty. We conduct an algebraic, numerical, and graphical analysis of these metrics applied to different well-studied security scenarios proposed in prior work (i.e., best shot, weakest-link, and total effort). In these scenarios, we study how a fully rational expert agent could utilize the metrics to decide whether to gather information about the economic incentives of multiple nearsighted and naïve agents. We find substantial differences between the various metrics and evaluate the appropriateness for security choices in networked systems.

Jens Grossklags
Princeton University, Center for Information Technology Policy, Sherrerd Hall, Princeton, NJ 08544, e-mail: jensg@princeton.edu

Benjamin Johnson
Carnegie Mellon University, CyLab, 4720 Forbes Ave, Pittsburgh, PA 15213, e-mail: johnsonb@andrew.cmu.edu

Nicolas Christin
Carnegie Mellon University, CyLab, 4720 Forbes Ave, Pittsburgh, PA 15213, e-mail: nicolasc@andrew.cmu.edu

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_2,

2.1 Introduction

The importance of (the lack of) information about security threats, response mechanisms, and associated expected losses and cost has long been identified in the computer science, risk management and economics communities. Granick, for example, argues that weaknesses in our understanding of the measurability of losses serve as an impediment in sentencing cybercrime offenders [14]. Swire adds that deterring fraudsters and criminals online is hampered if we cannot correctly aggregate their offenses across different jurisdictions [37].

The question arises how much defenders can gain by investing in techniques or other efforts to improve information availability for decision-making? Swire's analysis foreshadows significant costs to create an information exchange for law enforcement that could support evidence gathering. Similarly, private organizations struggle with how to accumulate data about security risks and incidents in their respective industries. Past work has, for example, considered the role of intermediaries such as Information Sharing & Analysis Centers to create incentives for exchanging and disclosing data between companies. Researchers investigated under which conditions organizations are willing to contribute to an information pool about security breaches and investments when (negative) competitive effects may result from this cooperation [10, 13]. In different contexts disclosure is not always voluntary and companies may question how much profit they squander when undesirable information is released. For example, other economics research explores the impact of (mandated) breach disclosures [5] or publication of software vulnerabilities [38] on the financial market value of corporations. While other work shows that the information gathering or disclosure effect is not always unambiguously positive or negative [7].

This trade-off between cost and benefits of information gathering, sharing or disclosure reappears in many contexts. From a viewpoint of individual rationality it is decided based on the difference of how much the individual can learn in comparison to the advantage gained by attackers or competitors [36].

Our contribution is to propose and evaluate a set of generic metrics that are applicable to different security decision-making situations to help with this trade-off calculation. In particular, we are interested in quantifying the payoff differential that results from the changes in security choices given different information available. In economic terms we thereby refer to the differences in payoff that results from changes in the underlying *information structure* of the scenario that makes explicit the nature of the utility of information to agents [27].

Specifically, we introduce the *price of uncertainty* metric that quantifies the maximum discrepancy in the total expected payoff between exactly two information conditions.[1] Our terminology is made per analogy with Koutsoupias and Papadim-

[1] After our initial proposal of the price of uncertainty [19], Balcan *et al.* published a research study in which they defined the price of uncertainty as the degree that small fluctuations in costs impact the result of natural best-response and improved-response dynamics [3].

itriou's "price of anarchy" [24]. We consider *difference*, *payoff-ratio*, and *cost-ratio* metrics as canonical nontrivial measurements of the price of uncertainty.

Since the possibilities for the economic formalization of information are vast, we illustrate our approach on an example model for security choices. Specifically, we introduce uncertainty by assuming that each agent faces a randomly drawn probability of being subject to a direct attack. We study how the decisions and payoffs of an individual agent differ if all draws are common knowledge, compared to a scenario where this information is only privately known [18].

We conduct this analysis within the framework of security games [15, 16] to understand the behavior of the price of uncertainty across different canonical interdependency cases: best shot, weakest-link and total effort [39]. We further consider a recent extension of our work in which we distinguish between the roles of a fully rational expert agent and naïve end users [17]. The inexperienced users conduct a simple self-centered cost-benefit analysis, and neglect interdependencies. We analyze the price of uncertainty from the perspective of the expert agent that fully comprehends the benefits of information in the context of the interrelationship with the naïve users [18]. This allows us to make a general observation. The value of information for the expert agent is always weakly positive [27] since naïve users do not strategize based on additional information.

In this model, the price of uncertainty can depend on several different parameters: the cost of security measures, the magnitude of potential losses, the initial security budget or endowment, and the number of other naïve agents. We study the impact of these parameters algebraically, numerically and graphically.

We show that the difference metric of the price of uncertainty increases linearly in losses, L, and decreases super-linearly in the number of agents, N. That is, only in the presence of extremely large losses would a decision-maker strictly prefer to explore the threat probabilities of other agents at a reasonable cost. The payoff-ratio metric is strictly decreasing in N and independent of the magnitude of potential losses, L. Finally, our cost-ratio metric serves as an example for misleading advice because it overemphasizes the need for action in the presence of relatively small costs.

By evaluating the price of uncertainty for a range of parameters in different security scenarios, we can determine which configurations can accommodate limited information environments (i.e., when being less informed does not significantly jeopardize an expert user's payoff). We also provide a framework for future work in the area of analysis of the value of security-relevant information. For example, we believe that the game-theoretic analysis in specialized scenarios, e.g., intrusion detection games [28], and security patrol versus robber avoidance scenarios [32] can benefit from a substantiation of the significance of informational assumptions by studying the price of uncertainty.

In Section 2.2, we summarize the security games framework we developed in prior work, and detail our assumptions about agent behaviors and information conditions. We present the different metrics for the price of uncertainty and describe our analysis methodology in Section 2.3. We conduct our analysis and discuss the

results in Section 2.4. Finally, we close with a discussion and concluding remarks in Section 8.8.

2.2 Decision Theoretic Model

Our current analysis of the *price of uncertainty* is based on the security games framework [15, 16] and our consecutive work that extends this model to an economy consisting of an expert user and several unsophisticated users that follow a simple but reasonable rule-of-thumb strategy [17, 18]. The latter investigation is a decision-theoretic approach [6, 12]. In the following, we present the key aspects of our model.

2.2.1 Basic Model

Self-protection and self-insurance. In practice, the action portfolio of a defender may include different options to prevent successful compromises and to limit losses that result from a breach. In Grossklags *et al.* [15] we provide a model that allows a decoupling of investments in the context of computer security. On the one hand, the perimeter can be strengthened with a higher self-protection investment (e.g., implementing or updating a firewall). On the other hand, the amount of losses can be reduced by introducing self-insurance technologies and practices (e.g., backup provisions). Formally, player i decides whether to invest in protection ($e_i = 1$) or not ($e_i = 0$). Similarly, each player can adopt a self-insurance technology ($s_i = 1$) or not ($s_i = 0$). In other words, e_i and s_i are two discrete decision variables.

Discrete choice decision-making captures many practical security problems. Examples include purchase and adoption investments as well as updating and patching of protection and self-insurance technologies [2, 25, 29, 30]. We have further conducted a sensitivity analysis with respect to the discrete choice assumption and find that, for the study in the present paper, the only differences between the discrete and continuous cases (where e_i and s_i are continuous variables over the interval $[0,1]$ as opposed to be mere binary variables) arise when there is strict equality between some of the terms in our case-specifying inequality conditions (see derivations in [18]). We believe that focusing on these boundary cases is of limited practical applicability, and could even be misleading. For comparison, we refer to our prior work where we considered the continuous case in a full information environment [15].

We further denote by $b \geq 0$ and $c \geq 0$ the cost of protection and self-insurance, respectively, which are homogeneous for the agent population. So, player i pays be_i for protection and cs_i for self-insurance.

Interdependency. Decisions by one defender frequently influence the incentives for security investments by her peers [39]. For example, the lack of protection efforts by

a subset of agents will often allow an attacker to also compromise resources of other agents if a common perimeter is breached. We denote H as a "contribution" function that characterizes the effect of e_i on agent's utility U_i, subject to the protection levels chosen (contributed) by *all* other players. We require that H be defined for all values over $[0,1]^N$. We distinguish three canonical cases that we discussed in-depth in prior work [15]:

- Best shot: $H = \max(e_i, e_{-i})$.
- Weakest-link: $H = \min(e_i, e_{-i})$.
- Total effort: $H = \frac{1}{N}\sum_k e_k$.

where, following common notation, e_{-i} denotes the set of protection levels chosen by players other than i.

Attack probabilities, network size and endowment. Each of $N \in \mathbb{N}$ agents receives an endowment M. If she is attacked and compromised successfully she faces a maximum loss of L. Her expected loss $p_i L$ is mitigated by a scaling factor p_i randomly drawn from a uniform distribution on $[0,1]$.[2] Instead of interpreting the parameter p_i as the probability of a successful attack; we consider the *expected loss*, $p_i L$, as the primary heterogeneous parameter under consideration. The same familiar notation with p_i considered as a heterogeneous mitigating factor as opposed to an attack probability facilitates this perspective.

The choice to consider a heterogeneous expected loss is motivated by practical considerations, as different targets may have highly variable liabilities, due to their economic, political, or reputational agenda. The choice of a uniform distribution on mitigating factors ensures the analysis remains tractable, while already providing numerous insights. We conjecture that different distributions (e.g., power law) may also be appropriate in practice.

2.2.2 Player Behavior

At the core of our analysis is the observation that expert and non-expert users differ in their understanding of the complexity of networked systems. Indeed, consumers' knowledge about risks and means of protection with respect to privacy and security can be quite varied [1], and field surveys separate between high and low expertise users [34].

Sophisticated (expert) user. Advanced users can rely on their superior technical and structural understanding of computer security threats and defense mechanisms, to analyze and respond to changes in the environment [8]. In the present context, expert users, for example, have less difficulty to conclude that the goal to avoid

[2] Technically, our analysis does not require complete knowledge of the distribution on the various p_i. The distribution informs the probability that a given number of p_j are above the rule-of-thumb threshold; but to conduct our analysis, it suffices to know only these threshold probabilities, and not the full distribution.

censorship points is a best shot scenario, whereas the protection of a corporate network frequently suggests a weakest-link optimization problem [15]. Accordingly, a sophisticated user correctly understands her utility to be dependent on the interdependencies that exist in the network:

$$U_i = M - p_i L(1-s_i)(1-H(e_i, e_{-i})) - be_i - cs_i \ .$$

Naïve (non-expert) user. Average users underappreciate the interdependency of network security goals and threats [1, 34]. We model the *perceived* utility of each naïve agent to only depend on the direct security threat and the individual investment in self-protection and self-insurance. The investment levels of other players are *not* considered in the naïve user's decision making, despite the existence of interdependencies. We define the perceived utility for a specific naïve agent j as:

$$PU_j = M - p_j L(1-s_j)(1-e_j) - be_j - cs_j \ .$$

Clearly, perceived and realized utility actually differ: by failing to incorporate the interdependencies of all agents' investment levels in their analysis, naïve users may achieve sub-optimal payoffs far below their anticipated expected payoffs. This paper does not aim to resolve this conflict, and, in fact, there is little evidence that users will learn the complexity of network security over time [34]. We argue that non-expert users would repeatedly act in an inconsistent fashion. This hypothesis is supported by findings in behavioral economics that consumers repeatedly deviate from rationality, however, in the same predictable ways [23].

2.2.3 Information Conditions

Our analysis is focused on the decision making of the expert user subject to the bounded rational behaviors of the naïve network participants. That is, more precisely, the expert agent maximizes their expected utility subject to the available information about other agents' drawn threat probabilities and their resulting actions. Two different information conditions may be available to the expert agent:

Complete information: Actual draws of attack probabilities p_j for all $j \neq i$, and her own drawn probability of being attacked p_i.

Incomplete information: Known probability distribution of the unsophisticated users' attack threat, and her own drawn probability of being attacked p_i.

Therefore, the expert agent can accurately infer what each agent's investment levels are in the complete information scenario. Under incomplete information the sophisticated user has to develop an expectation about the actions of the naïve users.

2.2.4 Remarks on Basic Results

We have conducted the basic analysis of this scenario in [18]. Below we are making several general observations to guide the reader through the results in this paper.

Every security scenario (i.e., best-shot, weakest-link and total effort) involves simple cost-benefit analyses for both sophisticated and naïve agents [11]. Agents remain passive when the cost of self-protection and self-insurance exceeds the expected loss. Further, they differentiate between the two types of security actions based on their relative cost. This behavior describes what we would usually consider as basic risk-taking that is part of everyday life: It is not always worth protecting against known risks.

One important feature of our model is the availability of self-insurance. If the cost of self-insurance c is less than the cost of protection b, the decision scenario significantly simplifies for all games and both information conditions. This is because once self-insurance is applied, the risk and interdependency among the players is removed. The interesting cases for all three games arise when $b \leq c$ and protection is a potentially cost-effective option. Within this realm insurance has a more subtle effect on the payoffs.

Tables 2.1, 2.2 and 2.3 contain the total expected payoff for decisions made by the sophisticated agent, but also for the naïve agents. We have already highlighted that for $c < b$ all agents follow the same simple decision rule to decide between passivity and self-insurance. Therefore, payoffs in this region are identical for all agent types in the case of homogeneous security costs. But, there are payoff differences among all three information conditions for some parts of the parameter range when $b \leq c$.

It is intuitive that the naïve agents suffer in the weakest-link game since they do not appreciate the difficulty to achieve system-wide protection. Similarly, in the best shot game too many unsophisticated agents will invest in protection lowering the average payoff. In the total effort game, sophisticated agents realize that their contribution is only valued in relation to the network size. In comparison, naïve agents invest more often. Further, the payoff profile of the unsophisticated agents remains flat for $b < c$. This reflects the fact that the naïve agent ignores the insurance option whenever protection is cheaper.

We can observe that the sophisticated agents will suffer from their misallocation of resources in the weakest-link game when information is incomplete. In the best shot game this impact is limited, but there is a residual risk that no naïve agent willingly protects due to an unlikely set of draws. In such cases the fully informed expert could have chosen to take it upon herself to secure the network. In the total effort game we observe a limited payoff discrepancy for expert users as a result of limited information.

2.2.5 Outlook on Further Analyses

Above we have provided a short summary of the key results that help to distinguish the three canonical scenarios and the decision-making of the expert and naïve agents (as detailed in [18]). From this point on we venture into new territory.

We start with the total payoff results in Tables 2.1, 2.2, and 2.3 and derive metrics to compare the impact of the important decision making parameters on the payoffs achievable in the two different information conditions. Thereby, we focus on the choices and payoffs garnered by the expert agent.

2.3 Price of Uncertainty Metrics

2.3.1 The Price of Uncertainty

In previous work we discussed two information conditions (complete information and incomplete information) for an expert player in three canonical security games. In this context, the price of uncertainty measures the disadvantage of the expert player when she has incomplete information, compared to when she has complete information. Depending on the form this measure takes, the price of uncertainty potentially depends on five different parameters:

1. Cost of protection b,
2. Cost of insurance c,
3. Magnitude of potential losses L,
4. Initial endowment M, and
5. Number of other players N.

Because the analysis of five-variable functions is somewhat cumbersome, a central objective in our metric-creation exercise is to reduce the number of parameters in a manner such that something both relevant and interesting can be said. Therefore, we focus on how the price of uncertainty depends on the magnitude of potential losses L and the number of other players N. To eliminate M we choose a canonical value of either 0 or L, and to eliminate b and c we chose the values that cause the price of uncertainty to have the greatest significance. This choice depends on the metric.

2.3.2 Three Metrics for the Price of Uncertainty

For each of the security games (i.e., best shot, weakest link, and total effort), we define and analyze three metrics for the price of uncertainty:

1. The difference metric $PoU_1(L,N)$, defined by

$$\max_{b,c\in[0,L]} [\text{Expected Payoff Complete}(b,c,L,L,N) - \text{Expected Payoff Incomplete}(b,c,L,L,N)]$$

2. The payoff-ratio metric $PoU_2(L,N)$ defined by

$$\max_{b,c\in[0,L]} \left[\frac{\text{Expected Payoff Complete}(b,c,L,L,N)}{\text{Expected Payoff Incomplete}(b,c,L,L,N)}\right]$$

3. The cost-ratio metric $PoU_3(L,N)$ defined by

$$\min_{b,c\in[0,L]} \left[\frac{\text{Expected Payoff Complete}(b,c,L,0,N)}{\text{Expected Payoff Incomplete}(b,c,L,0,N)}\right]$$

2.3.3 Discussion of the Definitions

2.3.3.1 The Difference Metric

The difference metric is our most straightforward metric. It says the price of uncertainty is the worst case difference in payoff between complete and incomplete information, where the maximum is taken over all possible prices for protection and self-insurance. In this metric, a completely insignificant price of uncertainty yields an output of zero, and the metric's output increases directly as the price of uncertainty becomes more significant.

2.3.3.2 The Payoff-Ratio Metric

The payoff-ratio metric is motivated by the game-theoretic notion of the "price of anarchy", which is defined as a payoff-ratio of a game's socially optimal equilibrium to its worst case Nash equilibrium [24]. By analogy, we define the price of uncertainty as the worst case ratio between the payoffs for the expert with complete information to the expert with incomplete information, with the worst case taken over all possible prices of protection and self-insurance. One advantage of using a ratio-style metric of this type is that its output is currency-independent. In other words, while our difference metric might depend on say dollars or euros, this ratio metric is just a pure number. In the payoff-ratio metric, a completely insignificant price of uncertainty yields an output of 1, and the metric's output *increases* as the price of uncertainty becomes more significant.

2.3.3.3 The Cost-Ratio Metric

The cost-ratio metric is similar to the payoff-ratio metric, but with a different canonical choice of 0 for the initial endowment M. This metric directly measures the ratio of costs induced by the expert's choices. These costs are reflected in formulas involving b, c, L, and N. Mathematically, the cost-ratio allows for a simpler algebraic analysis due to an abundance of term cancellations. A minor disadvantage of this metric's formulation is that it has a somewhat nonstandard orientation, in the sense that it decreases as the price of uncertainty becomes more significant. There are two justifications for this choice. First, we wanted to cast this metric as being a simpler analogue to the payoff-ratio metric. Second, we wanted to avoid values at infinity, which would have resulted had we used this metric's multiplicative inverse. In our cost-ratio metric, a completely insignificant price of uncertainty yields an output of 1, and the metric's output *decreases* toward zero as the price of uncertainty becomes more significant.

2.4 Analysis

In this section, we analyze the price of uncertainty as defined by each of the three metrics in each of the canonical security games. In each case the analysis proceeds as follows. First, considering the magnitude of potential loss L and the number of other players N as fixed parameters, we determine the protection cost b and self-insurance cost c which cause the metric under consideration to yield its most significant value. This process defines a function of two parameters L and N, which we then analyze as a measure of the price of uncertainty. In some scenarios we are able to produce clean algebraic results with tight asymptotic bounds. For others we must rely almost completely on computer-aided numerical analysis and graphs. Each subsection contains graphs of all relevant metrics and maximizing parameters, and concludes with some important observations.

2.4.1 Best Shot Game

In the best shot game (introduced in [15]), the security of the network is determined by the protection level of the individual with the highest investment. The relevant expected payoffs for an expert player in the best shot game are shown in Table 2.1. These results will be used throughout this section. Complete derivations for these payoffs can be found in [18].

Table 2.1 Best shot security game: Total expected game payoffs.

Case Name	Case Condition	Information Type	Total Expected Payoff
BC1	$c < b$	Complete	$M - c + \frac{c^2}{2L}$
BC2	$b \leq c$	Complete	$M - b\left(1 - \frac{b}{2L}\right)\left(\frac{b}{L}\right)^{N-1}$
BI1	$c < b$	Incomplete	$M - c + \frac{c^2}{2L}$
BI2	$b \leq c$	Incomplete	$M - \frac{L}{2}\left(\frac{b}{L}\right)^{N-1}$
BN1	$c < b$	Naive	$M - c + \frac{c^2}{2}$
BN2	$b \leq c$	Naive	$M - b + \frac{b^2}{2L}$

2.4.1.1 The Best Shot Difference Metric: $BPoU_1(L,N)$

In this section, we analyze the price of uncertainty metric $BPoU_1(L,N)$ defined as:

$$\max_{b,c\in[0,L]} [\text{Best Shot Expected Payoff Complete}(b,c,L,M,N) - \text{Best Shot Expected Payoff Incomplete}(b,c,L,M,N)] \tag{2.1}$$

In the best shot game, the complete and incomplete payoffs are the same when $c < b$; hence to compute the maximum payoff difference we may assume that $b \leq c$. Observe that in this case, the payoffs do not depend on c at all. This will help to simplify our analysis, and in fact allows us to compute $BPoU_1(L,N)$ in a purely algebraic manner.

We find that any b maximizing this equation satisfies

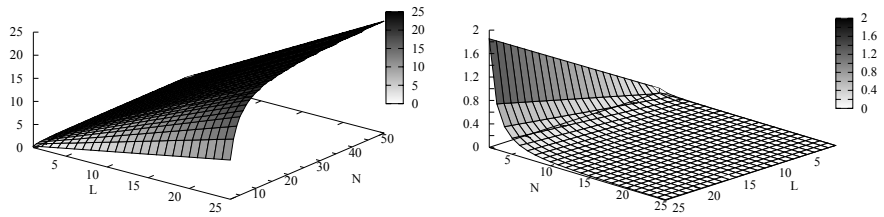
$$b = L \cdot \left(\frac{N-1}{N+1}\right),$$

and that consequently,

$$BPoU_1(L,N) = 2L \cdot \frac{(N-1)^{N-1}}{(N+1)^{N+1}}. \tag{2.2}$$

To give an asymptotic analysis, we begin by noting that $\lim_{n\to\infty}\left(\frac{N-1}{N+1}\right)^{N-1} = \frac{1}{e^2}$. Rewriting the expression above as $2L\left(\frac{N-1}{N+1}\right)^{N-1} \cdot \frac{1}{(N+1)^2}$, we see that the first part approaches $\frac{2L}{e^2}$ as N gets large, and that the second part decreases to zero quadratically in $\frac{1}{N}$. Hence this metric for the price of uncertainty increases linearly in L for fixed N and decreases quadratically to zero in $\frac{1}{N}$ for fixed L. Figure 2.1(a) shows a graph of the maximizing b for $BPoU_1$ as a function of N and L; while Figure 2.1(b) shows a graph of the metric $BPoU_1$ as a function of N and L. A complete algebraic derivation is also available in the workshop version of this paper [19].

Observations. The interpretation of our numerical results for this metric is that the price of uncertainty increases with the potential losses, but as the number of players



(a) The maximizing b for $BPoU_1(L,N)$ (b) $BPoU_1(L,N)$

Fig. 2.1 Best shot – Difference metric: $BPoU_1(L,N)$. The metric grows linearly in the potential loss L for a fixed network size N, and decreases inverse-quadratically in the network size N for a fixed loss L.

increases, the price of uncertainty diminishes (unless the losses are quite high) and approaches the square of the number of players.

2.4.1.2 The Best Shot Payoff-Ratio Metric $BPoU_2(L,N)$

In this section, we analyze the price of uncertainty metric $BPoU_2(L,N)$, defined as

$$\max_{b,c\in[0,L]} \left[\frac{\text{Best Shot Expected Payoff Complete}(b,c,L,L,N)}{\text{Best Shot Expected Payoff Incomplete}(b,c,L,L,N)} \right]. \tag{2.3}$$

After substituting $B = \frac{b}{L}$ we may derive

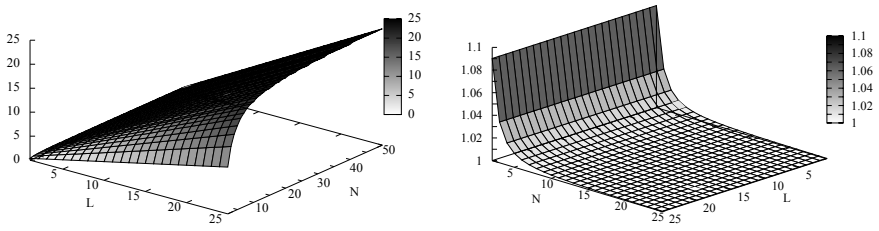
$$BPoU_2(L,N) = \max_{B\in[0,1]} 1 + \frac{\frac{1}{2}B^{N-1}(1-B)^2}{1-\frac{1}{2}B^{N-1}},$$

and the maximizing B for this equation occurs when

$$0 = \frac{1-B}{2}B^{N-2}\left(B^N - B(N+1) + N - 1\right).$$

Observing that $B^N - B(N+1) + N - 1$ is positive at $B = 0$ and negative at $B = 1$, and making additional arguments, it can be shown that this equation has exactly one solution in $(0,1)$. Due to well-known algebraic results, this solution cannot be expressed algebraically for $N \geq 5$, but we can plot the solution graphically. Figure 2.2 Grossklags/plots a graph of the maximizing $b = LB$ as a function of N and L. Figure 2.2(b) Grossklags/plots $BPoU_2$ as a function of N. As can be seen from the graph (or from our derivation), this metric does not depend on L, and it approaches 1 as N increases.

Observations. Since 1 represents the smallest price possible in this metric, the interpretation would be that the price of uncertainty is independent of the magnitude of potential losses and becomes insignificant as the number of players increases.



(a) The maximizing b for $BPoU_2(L,N)$ (b) $BPoU_2(L,N)$

Fig. 2.2 Best shot – Payoff-ratio metric: $BPoU_2(L,N)$. The metric is independent of L.

2.4.1.3 The Best Shot Cost-Ratio Metric $PoU_3(B,L,N)$

In this section, we analyze the price of uncertainty metric $BPoU_3(L,N)$, defined as

$$\min_{b,c\in[0,L]} \left[\frac{\text{Best Shot Expected Payoff Complete}(b,c,L,0,N)}{\text{Best Shot Expected Payoff Incomplete}(b,c,L,0,N)} \right]. \tag{2.4}$$

This metric is expressed in terms of our payoff functions, but by starting with an initial endowment of zero, it becomes a ratio of costs. If the cost of limited information is great compared to the cost of complete information, this ratio will tend toward zero. On the other hand, if the costs are similar, then the ratio will tend toward one. We select the minimizing b and c for this ratio so as to obtain the most significant price of uncertainty under the metric. Using this strategy, we obtain

$$BPoU_3(L,N) = \min_{b\in[0,L]} \frac{2b}{L}\left(1-\frac{b}{2L}\right).$$

Clearly the minimum value (of zero) for this expression (assuming $0 \le b \le L$) is achieved by taking $b = 0$. This cost-ratio metric always measures the price of uncertainty at its greatest possible value, independent of N or L.

Observations. The most direct interpretation for this result would be that the price of uncertainty is very significant, regardless of the number of players or the potential losses. An alternative, and arguably better explanation is that this particular metric is not a very useful provider of information for the best shot game.

2.4.2 Weakest Link Game

In the weakest link game (introduced in [15]), the security of the network is determined by the protection level of the individual with the lowest protection investment. The relevant expected payoffs for the weakest link game are shown in Table 2.2. These results will be used throughout this section. Complete derivations for these payoffs can be found in [18].

Table 2.2 Weakest link security game: Total expected game payoffs.

Case Name	Case Condition	Information Type	Total Expected Payoff
WC1	$c < b$	Complete	$M - c + \frac{c^2}{2L}$
WC2	$b \leq c$	Complete	$M - c + \frac{c^2}{2L} + (c-b)\left(1 - \frac{c+b}{2L}\right)\left(1 - \frac{b}{L}\right)^{N-1}$
WI1	$c < b$	Incomplete	$M - c + \frac{c^2}{2L}$
WI2	$b \leq c \leq \frac{b}{\left(1-\frac{b}{L}\right)^{N-1}}$	Incomplete	$M - c + \frac{c^2}{2L}$
WI3	$\frac{b}{\left(1-\frac{b}{L}\right)^{N-1}} < c < b + L\left(1 - \left(1 - \frac{b}{L}\right)^{N-1}\right)$	Incomplete	$M - c + \frac{b^2}{2L\left(1-\frac{b}{L}\right)^{N-1}} + \frac{(c-b)^2}{2L\left(1-\left(1-\frac{b}{L}\right)^{N-1}\right)}$
WI4	$\frac{b}{\left(1-\frac{b}{L}\right)^{N-1}} < b + L\left(1 - \left(1 - \frac{b}{L}\right)^{N-1}\right) \leq c$	Incomplete	$M - b - \frac{L}{2}\left(1 - \left(1 - \frac{b}{L}\right)^{N-1}\right) + \frac{b^2}{2L\left(1-\frac{b}{L}\right)^{N-1}}$
WN1	$c < b$	Naive	$M - c + \frac{c^2}{2}$
WN2	$b \leq c$	Naive	$M - b + \frac{b^2}{2L} - \frac{L}{2}\left(1 - \frac{b^2}{L^2}\right)\left(1 - \left(1 - \frac{b}{L}\right)^{N-1}\right)$

In the weakest link game, the complete and incomplete payoffs are the same when $c < b$, but for $b \leq c$ there is a wide variety of cases to consider, and without some direction it is not obvious which direction we should take in our analysis. Unlike the best shot game in which most of our equational analysis involved a single variable b in a relatively simple expression, a soft algebraic analysis of the weakest link game is much more difficult to conduct. Our strategy is to use numerical approximations and graphs to determine which cases to consider, and consequently which equations to work with. Thus, most of our algebraic work for this game takes the form of supporting, verifying, and clarifying the numerical analysis.

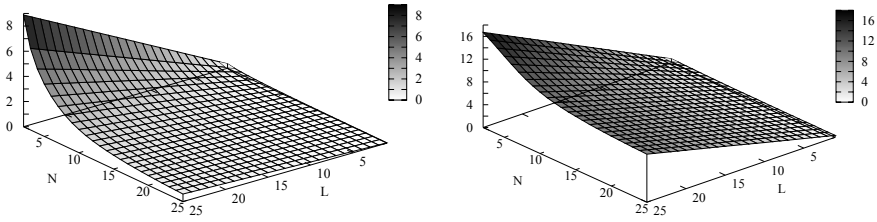
2.4.2.1 The Weakest Link Difference Metric: $WPoU_1(L,N)$

In this section, we analyze the price of uncertainty metric $WPoU_1(L,N)$ defined as:

$$\max_{b,c\in[0,L]} [\text{Weakest Link Expected Payoff Complete}(b,c,L,L,N) - \text{Weakest Link Expected Payoff Incomplete}(b,c,L,L,N)]. \tag{2.5}$$

Our numerical analysis of this difference metric indicates that all the highest values lie in the weakest link game's case WI3, in which we have $\frac{b}{\left(1-\frac{b}{L}\right)^{N-1}} < c < b + L\left(1 - \left(1 - \frac{b}{L}\right)^{N-1}\right)$. Using this, we may derive an expression for this metric involving equations; however the minimizing values of b and c that yield the final

solution are roots of polynomial equations whose degree depends on N. Here we will dispense with the partial derivations and refer the reader to the graphs. Figure 2.3 gives the maximizing b and c (respectively) as functions of L and N. Then, Figure 2.4 gives the weakest link difference metric $WPoU_1$ as a function of L and N.



(a) Maximizing b (b) Maximizing c

Fig. 2.3 Weakest Link – Difference metric: The maximizing b and c (respectively) for $WPoU_1(L,N)$.

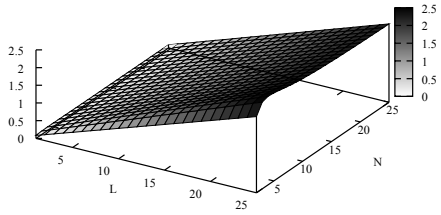


Fig. 2.4 Weakest Link – Difference metric: $WPoU_1(L,N)$. The metric grows linearly in the losses L and remains relatively constant for fixed L regardless of the network size N.

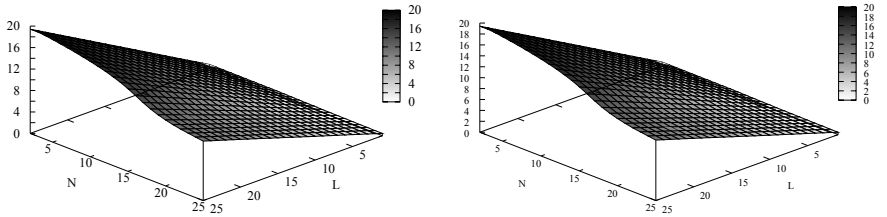
Observe that the maximizing b decreases to 0 as a function of N but increases linearly in L. The maximizing c also decreases in N and increases linearly in L. The difference metric itself increases linearly in L, but remains relatively-constant as N grows. This phenomenon can be explained by the following observation. The maximizing b for this metric satisfies the relation $\frac{b}{L} \in O\left(\frac{1}{N}\right)$, whence the expression $\left(1-\frac{b}{L}\right)^{N-1}$ approaches a constant as N increases. All terms in $WPoU_1(L,N)$ involving N have this form; thus as N grows the function value does not change. The graph shows additionally that the convergence to a constant value is quite fast in N.

Observations. The interpretation for these numerical results is that the price of uncertainty in the weakest link game is highest when protection is cheap and self-insurance is competitively priced. The price of uncertainty increases directly with the potential loss, and is unaffected by the number of other players.

2.4.2.2 The Weakest Link Payoff-Ratio Metric $WPoU_2(L,N)$

In this section, we analyze the price of uncertainty metric $WPoU_2(L,N)$, defined as

$$\max_{b,c\in[0,L]} \left[\frac{\text{Weakest Link Expected Payoff Complete}(b,c,L,L,N)}{\text{Weakest Link Expected Payoff Incomplete}(b,c,L,L,N)} \right]. \tag{2.6}$$



(a) Maximizing b (b) Maximizing c

Fig. 2.5 Weakest Link – Payoff-ratio metric: The maximizing b and c (respectively) for $WPoU_2(L,N)$.

We begin by considering the graphs in Figure 2.5, which give as functions of L and N the b and c (respectively) which maximize the price of uncertainty under this metric. We see that the maximizing b increases linearly with L, but decreases to zero super-linearly in $\frac{1}{N}$. The maximizing c also increases linearly with L, and decreases with N. For the weakest link payoff-ratio metric, we observe that the metric has no dependence on L, and that there is a local maximum very close to $N = 4$, and that after $N = 4$ the ratio decreases toward zero as N increases.

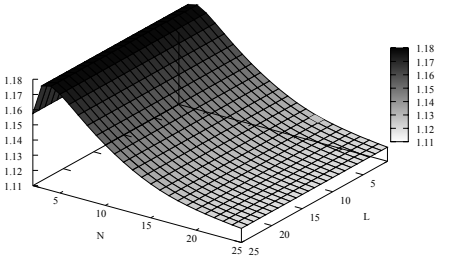


Fig. 2.6 Weakest Link – Payoff-ratio metric: $WPoU_2(L,N)$. Numeric simulations confirm the metric is independent of L.

The graph for the payoff-ratio metric is given in Figure 2.6. We see from the figure that the metric does not depend on L. We can also derive this observation by considering the equations as we did in the best shot case, specifically noting that it is without loss of generality to consider a maximum over $\frac{b}{L}$ and $\frac{c}{L}$ in place of b and

c, respectively. Because the metric only depends on $\frac{b}{L}$ and $\frac{c}{L}$ with the conditions $0 \leq b, c \leq L$, it follows that $L = 1$ without loss of generality, and hence the metric does not depend on L.

Observations. We observe that in the weakest link payoff-ratio metric, the price of uncertainty is highest when there are exactly 4 players, and it decreases toward its minimum possible value as the number of players increases.

2.4.2.3 The Weakest Link Cost-Ratio Metric $WPoU_3(L,N)$

In this section, we analyze the price of uncertainty metric $WPoU_3(L,N)$, defined as

$$\min_{b,c\in[0,L]} \left[\frac{\text{Weakest Link Expected Payoff Complete}(b,c,L,0,N)}{\text{Weakest Link Expected Payoff Incomplete}(b,c,L,0,N)} \right]. \tag{2.7}$$

Plotting as functions of L and N the b and c (respectively) which maximize the price of uncertainty under this metric (not shown for space purposes) shows that the maximum value for b is always achieved when b (and consequently $\frac{b}{L}$) is close to zero. The maximizing c is attained when $\frac{c}{L}$ is scaled with $\frac{b}{L}$ appropriately.

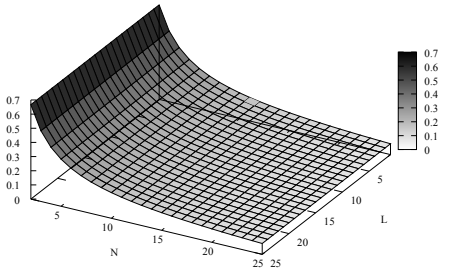


Fig. 2.7 Weakest Link – Cost-ratio metric: $WPoU_3(L,N)$.

The graph for the payoff ratio metric is given in Figure 2.7. As with the payoff-ratio metric considered above, this ratio-based metric does not depend on L. The plot gives nonzero values for all N but decreases to zero as N increases. Recall that zero in this metric represents the most significant price of uncertainty.

Observations. The results for this metric can be interpreted as saying that the price of uncertainty becomes more significant as the number of players increases. This interpretation contradicts our observations in the difference and payoff-ratio metrics for this game, and serves as a prime example to illustrate that the choice of metric makes a significant difference in the interpretation. Our explanation of the discrepancy is that this cost-ratio metric focuses on comparing costs which are insignificantly small in both the complete and incomplete information environments, but whose limiting ratio indicates a significant discrepancy. Based on this observa-

tion, a blunt assessment is that the cost-ratio metric for the weakest link game does not measure what we most generally think of as important.

2.4.3 Total Effort Game

In the total effort game (introduced in [15]), the security of the network is determined by the average protection level of all individual players in the network. The relevant expected payoffs for the total effort game are shown in Table 2.3. These results will be used throughout this section. Complete derivations for these payoffs can be found in [18].

Table 2.3 Total effort security game: Total expected payoffs.

Case Name	Case Condition	Information Type	Total Expected Payoff
TC1	$c < b$	Complete	$M - c + \frac{c^2}{2L}$
TC2	$bN \leq L$ and $b \leq c$	Complete	$\sum_{k=0}^{\lfloor N-\frac{c}{b} \rfloor} Pr[k] \cdot \left(M - c + \frac{c^2}{2L\left(1-\frac{k}{N}\right)} \right)$ $+ \sum_{k=\lfloor N-1-\frac{N}{L}(c-b) \rfloor}^{\lfloor N-\frac{c}{b}+1 \rfloor} Pr[k] \cdot \left(M - c + \frac{b^2 N}{2L} + \frac{(c-b)^2}{2L\left(1-\frac{k+1}{N}\right)} \right)$ $+ \sum_{k=\lfloor N-\frac{N}{L}(c-b) \rfloor}^{N-1} Pr[k] \cdot \left(M - b - \frac{L}{2}\left(1 - \frac{k+1}{N}\right) + \frac{b^2 N}{2L} \right)$
TC2	$L < bN$ and $b \leq c$	Complete	$\sum_{k=0}^{\lfloor N-\frac{cN}{L} \rfloor} Pr[k] \cdot \left(M - c + \frac{c^2}{2L\left(1-\frac{k}{N}\right)} \right)$ $+ \sum_{k=\lfloor N-\frac{cN}{L}+1 \rfloor}^{N-1} Pr[k] \cdot \left(M - \frac{L}{2N}(N-k) \right)$
TI1	$c < b$	Incomplete	$M - c + \frac{c^2}{L}$
TI2	$bN \leq L$ and $b \leq c \leq b + \frac{b^2}{L}(N-1)$	Incomplete	$M - c + \frac{c^2}{2\left(b + \frac{L-b}{N}\right)}$
TI3	$bN \leq L$ and $b + \frac{b^2}{L}(N-1) < c < 2b - \frac{b}{N}$	Incomplete	$M - c + \frac{b^2 N}{2L} + \frac{(c-b)^2}{2\left(b - \frac{b}{N}\right)}$
TI4	$bN \leq L$ and $2b - \frac{b}{N} \leq c$	Incomplete	$M - b - \frac{1}{2}\left(b - \frac{b}{N}\right) + \frac{b^2 N}{2L}$
TI5	$L < bN$ and $b \leq c < b + \frac{L-b}{N}$	Incomplete	$M - c + \frac{c^2}{2\left(b + \frac{L-b}{N}\right)}$
TI6	$L < bN$ and $b + \frac{L-b}{N} \leq c$	Incomplete	$M - \frac{1}{2}\left(b + \frac{L-b}{N}\right)$
TN1	$c < b$	Naive	$M - c + \frac{c^2}{2}$
TN2	$b \leq c$	Naive	$M - b - \frac{1}{2}\left(b - \frac{b}{N}\right) + \frac{b^2}{L}\left(1 - \frac{1}{2N}\right)$

2.4.3.1 The Total Effort Difference Metric: $TPoU_1(L,N)$

In this section, we analyze the price of uncertainty metric $TPoU_1(L,N)$ defined as:

$$\max_{b,c\in[0,L]}[\text{Total Effort Expected Payoff Complete}(b,c,L,M,N)- \text{Total Effort Expected Payoff Incomplete}(b,c,L,M,N)]. \tag{2.8}$$

As with the weakest link game, there are a number of cases to consider when beginning to analyze the price of uncertainty metrics. Numerical evidence suggests that the maximizing b and c for this game are in the total effort game's case TI3, in which we have $bN \le L$ and $b+\frac{b^2}{L}(N-1) < c < 2b-\frac{b}{N}$. Using the payoff equations from this case, we can make some progress toward an algebraic solution, deriving the following condition for (b,c) to maximize the payoff difference:

$$c = \frac{\sum_{k=\lfloor N-\frac{c}{b}+1\rfloor}^{\lfloor N-1-\frac{N}{L}(c-b)\rfloor}\left(\frac{Pr[k]}{L\left(1-\frac{k+1}{N}\right)}\right)-\sum_{k=\lfloor N-\frac{N}{L}(c-b)\rfloor}^{N-1}Pr[k]-\frac{b}{\left(b-\frac{b}{N}\right)}}{\sum_{k=0}^{\lfloor N-\frac{c}{b}\rfloor}\left(\frac{Pr[k]}{L\left(1-\frac{k}{N}\right)}\right)+\sum_{k=\lfloor N-\frac{c}{b}+1\rfloor}^{\lfloor N-1-\frac{N}{L}(c-b)\rfloor}\left(\frac{Pr[k]}{L\left(1-\frac{k+1}{N}\right)}\right)-\frac{1}{b-\frac{b}{N}}}.$$

This equation meets the frontiers of our algebraic simplification skills and motivates our haste in proceeding to the numerical analysis. Figure 2.8 Grossklags/plots the price of uncertainty as a function of N and L. We observe that the price of uncertainty in this metric increases linearly in L and decreases to zero with N significantly more quickly than $\frac{1}{N}$.

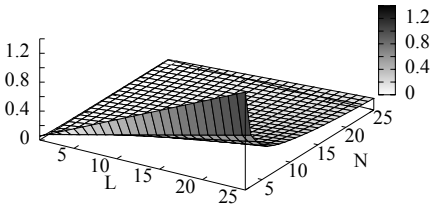


Fig. 2.8 Total Effort – Difference metric: $TPoU_1(L,N)$.

Observations. The interpretation of our numerical results for this metric is that the price of uncertainty increases with the potential losses, but as the number of players increases, the price of uncertainty diminishes quickly.

2.4.3.2 The Total Effort Payoff-Ratio Metric: $TPoU_2(L,N)$

In this section, we analyze the price of uncertainty metric $TPoU_2(L,N)$ defined as:

$$\max_{b,c\in[0,L]}\left[\frac{\text{Total Effort Expected Payoff Complete}(b,c,L,L,N)}{\text{Total Effort Expected Payoff Incomplete}(b,c,L,L,N)}\right]. \tag{2.9}$$

For the remaining total effort metrics, our analysis relies exclusively on numerical approximations. Figure 2.9(b) Grossklags/plots the total effort game's payoff-

ratio price of uncertainty as a function of N. The figure shows that the price of uncertainty does not depend on L and that it decreases toward 1 as N increases.

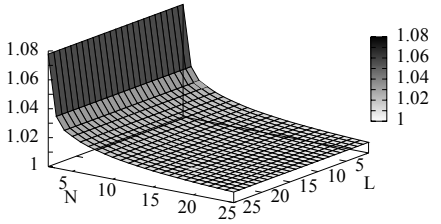


Fig. 2.9 Total Effort – Payoff-ratio metric: $TPoU_2(L,N)$.

Observations. In the total effort game, the payoff-ratio metric depends only on the number of players, and it diminishes to its least significant possible value as the number of players increases.

2.4.3.3 The Total Effort Cost-Ratio Metric: $TPoU_3(L,N)$

In this section, we analyze the price of uncertainty metric $TPoU_3(L,N)$ defined as:

$$\max_{b,c\in[0,L]} \left[\frac{\text{Total Effort Expected Payoff Complete}(b,c,L,0,N)}{\text{Total Effort Expected Payoff Incomplete}(b,c,L,0,N)} \right]. \tag{2.10}$$

Figure 2.10(c) Grossklags/plots the total effort game's cost-ratio price of uncertainty as a function of N. As can be seen from the graph, the price of uncertainty does not depend on L, and decreases as N increases.

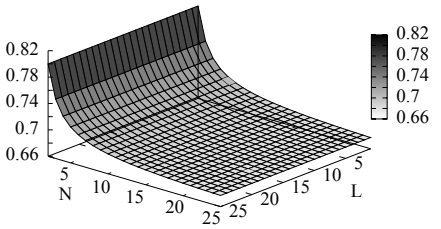


Fig. 2.10 Total Effort – Cost-ratio metric: $TPoU_3(L,N)$.

Observations. Using the cost-ratio metric for the total effort game, the price of uncertainty becomes more significant with an increase in the number of players. Once again this goes against the analogous conclusions drawn with the other two metrics. We surmise that this happens because the cost-ratio metric focuses on the

cases where the costs for the complete and incomplete information scenarios are quite small, while the ratio indicates a significant distinction.

2.5 Conclusions

Users frequently fail to deploy, or upgrade security technologies, or to carefully preserve and backup their valuable data [22, 31], which leads to considerable monetary losses to both individuals and corporations every year. This state of affairs can be partly attributed to economic considerations.

Significant challenges for average users arise when they have to determine optimal security strategies in the presence of interdependencies between security choices of other agents [15, 25]. Struggling with this task we anticipate the vast majority of users to be naïve, and to apply approximate decision-rules that fail to accurately appreciate the impact of their decisions on others [1].

In this paper we continue our investigation into the incentives of an individual expert user that rationally responds to the security choices of unsophisticated end-users under different informational assumptions [18]. In particular, we study how the expert evaluates the importance of improving the information available for her decision-making. We propose three variations of the *price of uncertainty* metric that may serve as a decision help for the expert user. We distinguish between a difference, a payoff-ratio, and a cost-ratio metric.

Our work complements the rich area of security metrics that are commonly technical, financial [21] or market-based [4]. However, the price of uncertainty is motivated by game-theory and, more specifically, by Koutsoupias and Papadimitriou's metric to evaluate worst-case equilibria [24], and adds to the rich literature on information sharing, (mandatory) disclosure, and notice and consent that we reviewed in the introductory section.

Our research yields a number of somewhat counter-intuitive results:

- Using cost-ratio metrics can be misleading, as two negligible costs in front of a large endowment may still produce a large ratio when divided by each other. While mathematically trivial, such a pitfall is relatively easy to get into. We showed that, unfortunately, for *all* games we studied, cost-ratios are *never* an appropriate metric. The cynic in ourselves could actually point out that their main use would be for marketing purposes. Beware of snake oil!
- Aside from the cost-ratio metric, the other metrics show a relatively low price of uncertainty across all the scenarios we considered, and this is especially true with a large number of players. The difference metric shows some signs of a penalty for lack of information, but if we consider the absolute payoff values (reported in Tables 2.1, 2.2, and 2.3) we find the price of uncertainty in the difference metric is at most 20% of the magnitude of the potential loss. Accordingly, we can summarize that in scenarios with many players the lack of information does not penalize an expert too much. On the other hand, the lack of knowledge (about interdependencies) that makes a user naïve, as opposed

to expert, results in significant payoff degradation regardless of the number of players [18].

- Assuming fixed possible losses, the more players are in a network, the less information matters. This is actually good news, as full information typically gets increasingly difficult to gather as the number of players grows large.
- In contrast to our arguments in favor of difference-based metrics behavioral research has shown that individuals are frequently influenced by ratio-difference evaluations [33]. However, this makes consumers more vulnerable to (numerical) framing differences that change perceptions about the benefits of additional information. For example, experimental research has reported robust evidence for consumers' preferences for benefits that are presented as large ratios in comparison to small ratios [26]. In the security context, marketers could easily switch the framing from a security to a reliability measure and thereby vary the size of the benefit ratio (e.g., from 3% vs. 5% failure to 97% vs. 95% reliability). As a result, individuals may exaggerage the importance of changes when risks or benefits are small [20, 35].
- We have also shown that the payoff-ratio and the cost-ratio metrics are independent of the size of the losses, L. Human-subject experiments suggest, however, that decision-makers may falsely utilize ratio considerations in the presence of (apparently) irrelevant information. For example, psychologists have found that investments in measures leading to savings of a fixed number of lives were preferred if the total number of individuals at risk was decreased [9]. Unfortunately, such a bias would lead to even less optimal decisions when considering the difference metric since the loss, L, is shown to be positively and linearly related to the price of uncertainty.

Of course, we should not forget that we consider a rather specialized environment, where only one single expert is alone in a population of naïve users. However stringent this assumption may sound, one should note that in reality, the number of expert users is dwarfed by the number of "lambda" users, that may not have the expertise, or inclination, to act very strategically.

Regardless of these limitations, we hope that our work will be a useful starting point for a serious discussion of information metrics applied to interdependent security scenarios. As we have shown here, picking the right metric is not a straightforward choice, and several pitfalls exist.

Acknowledgements We thank John Chuang, the anonymous reviewers and the participants of the Workshop on the Economics of Information Security for their helpful comments to an earlier version of this paper. All remaining errors are our own. This work is supported in part by the National Science Foundation under ITR award ANI-0331659 (100x100), with a University of California MICRO project grant in collaboration with DoCoMo USA Labs, and by CyLab at Carnegie Mellon under grant DAAD19-02-1-0389 from the Army Research Office. The views and conclusions contained here are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either express or implied, of ARO, CMU, or the U.S. Government or any of its agencies. We further want to thank Christos Ioannidis, Tyler Moore and David Pym for assembling this edited volume.

References

1. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security & Privacy **3**(1), 26–33 (2005)
2. August, T., Tunca, T.: Network software security and user incentives. Management Science **52**(11), 1703–1720 (2006)
3. Balcan, M., Blum, A., Mansour, Y.: The price of uncertainty. In: Proceedings of the ACM Conference on Electronic Commerce (EC), pp. 285–294. ACM Press, New York (2009)
4. Böhme, R., Nowey, T.: Economic security metrics. In: I. Eusgeld, F. Freiling, R. Reussner (eds.) Dependability Metrics, *LNCS*, vol. 4909, pp. 176–187. Springer, Berlin Heidelberg (2008)
5. Campbell, K., Gordon, L., Loeb, M., L. Zhou, L.: The economic cost of publicly announced information security breaches: Empirical evidence from the stock market. Journal of Computer Security **11**(3), 431–448 (2003)
6. Cavusoglu, H., Raghunathan, S., Yue, W.: Decision-theoretic and game-theoretic approaches to IT security investment. Journal of Management Information Systems **25**(2), 281–304 (2008)
7. Choi, J., Fershtman, C., Gandal, N.: Network security: Vulnerabilities and disclosure policy. Journal of Industrial Economics (forthcoming)
8. Dörner, D.: The Logic Of Failure: Recognizing And Avoiding Error In Complex Situations. Metropolitan Books (1996)
9. Fetherstonhaugh, D., Slovic, P., Johnson, S., Friedrich, J.: Insensitivity to the value of human life: A study of psychophysical numbing. Journal of Risk & Uncertainty **14**(3), 283–300 (1997)
10. Gal-Or, E., A. Ghose, A.: The economic incentives for sharing security information. Information Systems Research, **16**(2), 186–208 (2005)
11. Gordon, L., Loeb, M.: Managing Cyber-Security Resources: A Cost-Benefit Analysis. McGraw-Hill (2006)
12. Gordon, L.A., Loeb, M.: The economics of information security investment. ACM Transactions on Information and System Security **5**(4), 438–457 (2002)
13. Gordon, L.A., Loeb, M., Lucyshyn, W.: Sharing information on computer systems security: An economic analysis. Journal of Accounting and Public Policy, **22**(6), 461–485 (2003)
14. Granick, J.: Faking it: Calculating loss in computer crime sentencing. I/S: A Journal of Law and Policy for the Information Society **2**(2), 207–228 (2006)
15. Grossklags, J., Christin, N., Chuang, J.: Secure or insure? A game-theoretic analysis of information security games. In: Proceedings of the 17th International World Wide Web Conference (WWW), pp. 209–218. (2008)
16. Grossklags, J., Christin, N., Chuang, J.: Security and insurance management in networks with heterogeneous agents. In: Proceedings of the ACM Conference on Electronic Commerce (EC), pp. 160–169. ACM Press, New York (2008)
17. Grossklags, J., Johnson, B.: Uncertainty in the Weakest-link security game. In: Proceedings of GameNets, pp. 673-682. (2009)
18. Grossklags, J., Johnson, B., Christin, N.: When information improves information security. Tech. Rep. CMU-CyLab-09-004 (2009)
19. Grossklags, J., Johnson, B., Christin, N.: The price of uncertainty in security games. In: Proceedings of the 8th Workshop on the Economics of Information Security (WEIS). London, UK (2009)
20. Hershey, J., Baron, J.: Clinical reasoning and cognitive processes. Medical Decision Making **7**(4), 203–211 (1987)
21. Jaquith, A.: Security Metrics: Replacing Fear, Uncertainty, and Doubt. Pearson Education (2007)
22. Kabooza: Global backup survey: About backup habits, risk factors, worries and data loss of home PCs (2009). `http://www.kabooza.com/globalsurvey.html`
23. Kahneman, D., Tversky, A.: Choices, Values and Frames. Cambridge University Press (2000)

24. Koutsoupias, E., Papadimitriou, C.: Worst-case equilibria. In: Proceedings of the 16th Annual Symposium on Theoretical Aspects of Computer Science (STOC), pp. 404–413. ACM Press, New York (1999)
25. Kunreuther, H., Heal, G.: Interdependent security. Journal of Risk & Uncertainty **26**(2–3), 231–249 (2003)
26. Kwong, J., Wong, K.: The role of ratio differences in the framing of numerical information. International Journal of Research in Marketing **23**(4), 385–394 (2006)
27. Laffont, J.: The Economics of Uncertainty and Information. MIT Press (1989)
28. Liu, Y., Comaniciu, C., Man, H.: A Bayesian game approach for intrusion detection in wireless ad hoc networks. In: Proceedings of the Workshop on Game Theory for Communications and Networks, article no. 4. ACM Press, New York (2006)
29. Meier, D., Oswald, Y., Schmid, S., Wattenhofer, R.: On the windfall of friendship: Inoculation strategies on social networks. In: Proceedings of the ACM Conference on Electronic Commerce (EC), pp. 294–301. ACM Press, New York (2008)
30. Moscibroda, T., Schmid, S., Wattenhofer, R.: When selfish meets evil: Byzantine players in a virus inoculation game. In: Proceedings of the ACM Symposium on Principles of Distributed Computing (PODC), pp. 35–44. ACM Press, New York (2006)
31. NCSA/Symantec: Home user study (2008). `http://staysafeonline.org/`
32. Paruchuri, P., Pearce, J., Marecki, J., Tambe, M., Ordonez, F., Kraus, S.: Playing games for security: An efficient exact algorithm for solving Bayesian Stackelberg games. In: Proceedings of AAMAS, pp. 895–902. IFAAMAS, Richland, South Carolina (2008)
33. Quattrone, G., Tversky, A.: Contrasting rational and psychological analyses of political choice. The American Political Science Review **82**(3), 719–736 (1988)
34. Stanton, J., Stam, K., Mastrangelo, P., Jolton, J.: Analysis of end user security behaviors. Computers & Security **2**(24), 124–133 (2005)
35. Stone, E., Yates, F., Parker, A.: Risk communication: Absolute versus relative expressions of low-probability risks. Organizational Behavior & Human Decision Processes **3**(60), 387–408 (1994)
36. Swire, P.: A model for when disclosure helps security: What is different about computer and network security? Journal on Telecommunications and High Technology Law **3**(1), 163–208 (2004)
37. Swire, P.: No cop on the beat: Underenforcement in e-commerce and cybercrime. Journal on Telecommunications and High Technology Law **7**(1), 107–126 (2009)
38. Telang, R., Wattal, S.: An empirical analysis of the impact of software vulnerability announcements on firm stock price. IEEE Transactions on Software Engineering **33**(8), 544–557 (2007)
39. Varian, H.R.: System reliability and free riding. In: L.J. Camp and S. Lewis (eds.) Economics of Information Security, pp. 1–15. Kluwer Academic Publishers (2004)

Chapter 3
Nobody Sells Gold for the Price of Silver: Dishonesty, Uncertainty and the Underground Economy

Cormac Herley and Dinei Florêncio

Abstract The underground economy has attracted a lot of attention recently as a key component of cybercrime. In particular the IRC markets for stolen identities, phishing kits, botnets, and cybercrime related services have been extensively studied. It is suggested that sophisticated underground markets show great specialization and maturity. There are complex divisions of labor and service offerings for every need. Stolen credentials are traded in bulk for pennies on the dollar. It is suggested that large sums move on these markets.

We argue that this makes very little sense. Using basic arguments from economics we show that the IRC markets studied represent classic examples of lemon markets. The ever-present rippers who cheat other participants ensure that the market cannot operate effectively. Their presence represents a tax on every transaction conducted in the market. Those who form gangs and alliances avoid this tax, enjoy a lower cost basis and higher profit. This suggests a two tier underground economy where organization is the route to profit. The IRC markets appear to be the lower tier, and are occupied by those without skills or alliances, newcomers, and those who seek to cheat them. The goods offered for sale there are those that are easy to acquire, but hard to monetize. We find that estimates of the size of the IRC markets are greatly exaggerated. Finally, we find that defenders recruit their own opponents by publicizing exaggerated estimates of the rewards of cybercrime. Those so recruited inhabit the lower tier; they produce very little profit, but contribute greatly to the externalities of cybercrime.

Cormac Herley
Microsoft Research, Redmond, WA, e-mail: cormac@microsoft.com

Dinei Florêncio
Microsoft Research, Redmond, WA, e-mail: dinei@microsoft.com

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_3,

3.1 Introduction

There has been a recent surge of interest in the underground economy, both in the popular and academic presses. A common theme is the observation that there is a thriving market in the goods and services associated with online crime [12, 29–31]. Hackers who used to seek exploits for recreation or reputation have given way to those who are in it for the money. For example, an NY Times story "Black Market In Credit Cards Thrives on Web" (June 21, 2005) relates that "The online trade in credit card and bank account numbers, as well as other raw consumer information, is highly structured. There are buyers and sellers, intermediaries and even service industries." The underground economy described, in the NY Times and elsewhere, appears to mirror the real economy in many respects. There are well-defined specializations and complex divisions of labor. For example, some have stolen credentials for sale, while others act as cashiers to drain the accounts. Some develop phishing kits for sale while others maintain compromised hosts on which they can be deployed. Specialization is a usually a sign of developed economies, and generally increases the productivity of labor [28]. Thus the specialization observed is often taken as an indication of the size, maturity and value of the underground economy. There is a considerable and growing body of work documenting the activity on these markets.

While at first glance plausible, the accounts that reach us of this underground economy present a number of facts that do not make sense. First, common to most of the underground economy studies is the observation that stolen credit card numbers (CCNs) and credentials sell for pennies on the dollar. For example, Symantec [29] finds the asking price for a CCN varies between $0.5 and $12, even when the available balance is several thousand dollars. Thomas and Martin [31] quote an IRC exchange where 40k financial accounts with face value of $10 million are valued at less than $500. Published accounts of the underground economy are mostly silent on this gap. Occasionally it is presented this as evidence of how profitable the trade is; *i.e.*, sellers make so much that care little about maximizing return on any given CCN. This makes very little sense. Why would anyone sell for 50 cents an asset that is worth $2000? If the seller can't turn the CCN into cash himself surely someone will do it for less than the 99% and higher premiums that these numbers imply.

Second, several of the underground economy studies refer to large numbers of CCNs posted openly on the wire [12, 29–31]. That is the information is posted for all on the IRC channel to see. Symantec reports finding 44752 individual pieces of personal data, such as SSNs or CCNs, during a year [29]. Franklin *et al.* report a daily average of 465 CCNs posted on the single IRC channel they monitored. Some suggest that posting stolen CCNs is a way for participants to have their user nics verified [31], while others claim that posting free samples is a mechanism for sellers to attract business [12].

Third, the openness of the underground economy markets would appear to invite counter-measures. That is, if it has been easy for security researchers to find and function on these networks the same should be true of bank employees or law en-

forcement officials. According to Symantec [29]: "joining is usually open to anyone, often entailing registration with only a username." A bank can easily post honey pot CCNs on the channel to investigate the cash out strategy of hackers. Law enforcement can easily identify cashiers and drops by offering transactions and following the money. Franklin *et al.* [12] suggest a number of simple elegant counter-measures that could even be fatal to the market (see Section 3.4.1). It simply defies common sense to have a large underground economy that is so easily accessible to all.

Fourth, there is huge variance in the estimates of the amounts of money at stake in the underground economy. Popular press accounts tend to the sensational and talk about billions that trade on the underground economy markets. Symantec tallies the total asking price of all the CCNs they observed offered for sale at $163 million, but estimates the potential worth of those CCNs at $5.3 billion. Gartner puts 2007 phishing losses in the US at $3.2 billion. Yet, in previous work [19] we find that the losses are more likely in the vicinity of $60 million. Kanich *et al.* [22] found that a major spam campaign that involved 350 million emails sent, garnered revenue of only $2731.88. They point out that anecdotal reports of $80 per million for spam delivery would be too expensive by a factor of twenty for this campaign to make sense. Further, the revenue gained involved 75k botnet machines: that's a return of 50 cents per botnet machine per year. Even allowing that a botnet machine could be rented in parallel to do many things we have two orders of magnitude difference between the frequently quoted $1 per botnet machine per day and this measurement. Thus, there are enormous differences between the various estimates of the values of exploits: 32× between the ask and potential value of CCNs, 50× between two phishing estimates, 20× between return on a spam campaign and quoted spam rates, and 100× between the revenue from a botnet and quoted rental rates.

Finally, every account of the underground economy makes reference to rippers [9,12,29–31]. Rippers are participants in IRC markets who do not provide the goods or service for which they've been paid. They are energetic, inventive, and appear everywhere. How is it possible for a market to function when dishonesty is so easy and so profitable?

This paper is an attempt to resolve some of these apparent paradoxes using arguments from Economics. We find that the underground economy IRC markets are a classic example of a market for lemons [2]. That is, there is information asymmetry between buyer and seller: the uncertainty for a buyer in knowing whether he is dealing with a ripper or not. Every account we have of the underground economy makes clear that rippers are a real and ever-present menace. This uncertainty causes adverse selection, where rippers are attracted to the market (since they get money for nothing), while legitimate sellers tend to stay away (since the probability of getting ripped off is factored into everyone's buying decisions). Unchecked this causes the market to fail [2].

Essentially the risk of dealing with a ripper represents a tax on every transaction conducted in the market. Those who can avoid this tax have lower costs and higher profits; and the simplest way of avoiding the tax is to form deals repeatedly with partners who perform. This mirrors the Theory of the Firm by Coase [7] in which it is suggested that when market transactions are taxed, or expensive or risky it makes

sense to form relationships and ultimately firms rather than purchase resources in a market. It makes no sense to transact with anonymous market participants when there is considerable quality uncertainty *unless there is no alternative.*

This has a number of implications for the underground economy. While there is a great deal of activity in the underground economy market place, it does not imply a lot of dollars change hands. While it's tempting to regard this as representative, we believe the important deals happen where the ripper tax cannot reach them. This means that the IRC markets are a very low-value channel. Those who advertise goods for sale are either selling things that have no value to them, or hoping to exploit newcomers. Those who buy are either newcomers, or in need of connections. We believe that anyone who shows up on a non protected IRC channel hoping to trade profitably with anonymous partners runs the very real risk of being cheated. Thus, estimating the dollar size of the underground economy based on the asking price of good and services advertised on IRC networks appears unsound. Finally, the presence of a ripper tax on IRC channels points to a two tier system in the underground economy: those who are members of alliances and gangs remove a major cost from their business. Those who trade on IRC channels do so because they have no choice, or they are seeking to cheat. Thus, far from being a sophisticated clearing house where professional criminals trade specialties the IRC channels appear to be the bottom rung, where those with few skills, connections or experience mix with rippers.

3.2 Related Work

3.2.1 Studies of the Underground Economy

Thomas and Martin [31] were among the first to draw attention to and document the growing activity in the underground economy. They found enormous activity on IRC channels advertising stolen goods and services such as phishing kits, credentials *etc*. They colorfully describe the underground economy as an extremely busy market-place where individuals who specialize in particular activities trade goods and services to others. Some will produce phishing kits, some will offer hosting services, some sell the credentials harvested and still others offer to cash-out the actual dollars from compromised accounts. Franklin *et al.* [12] followed with a very detailed measurement study of an IRC market in the underground economy, and document the activity in a principled way. For example they found over 100k active user accounts on a single IRC trading channel, and measured an enormous quantity of credentials and services offered for sale.

Symantec has produced a series of reports on Internet Security and the underground economy. They appear to corroborate the view that the market for goods and services related to stolen credentials has become big business [30]: "The emergence of underground economy servers as the de facto trading place for illicit information

is indicative of the increased professionalization and commercialization of malicious activities over the past several years." In 2008 Symantec [29] finds that stolen credit cards are selling for as little as $0.10, but that the potential worth of all the credit cards was $5.3 billion: "Cybercrooks have developed sophisticated business models such that recognized job roles and specialisms have evolved in the 'recession proof' digital underground."

Geer and Conway [14, 15] informally suggest an Øwned Price Index of underground economy asking prices to track changes in the markets.

Dhanjani and Rios [9] also investigate an IRC marketplace and also observe impressive activity. Interestingly they find that phishers prey on each other: phishing kits offered for sale in the market turn out to have obfuscated backdoor code that reports the details of any credentials harvested to the author as well as to any user of the kit. Among their interesting observations are that many participants are unsophisticated and inexperienced, and a great many phishers struggle to monetize their exploits. Cova *et al.* [8] arrive at very similar conclusions.

Zhuge *et al.* [32] carry out a study on the Chinese underground economy, again focusing on activity and advertisements. The IRC channels popular elsewhere are less used in China.

Kanich [22] managed to invade a spamming botnet and track the amount of spam sent, the transactions conducted and the dollars that appear to have changed hands. They observe that in a 26 day study of a spamming botnet 350 million emails resulted in only 28 sales and total revenue of $2731.88. Interestingly, they suggest that the spam services they studied are produced by the controllers of the botnet itself. This suggests that the service is entirely integrated rather than sold as a commodity service to others (see Section 3.4.3).

John *et al.* [21] also provide a very detailed study of a spamming botnet. They find, for example, that a small number of botnets account for a majority of spam. This corroborates the view that a few well organized gangs dominate the space.

Holz *et al.* [20] carry out a study of dropzones (*i.e.* servers that are used to park stolen credentials). They observe the number of stolen credentials that get stored, but have no direct means of estimating the value of each. They take the Symantec [30] estimates of the value of credentials to arrive at a figure for the size of the underground economy they study.

3.2.2 Economics of Security and of the Underground Economy

Anderson [3] first proposed the comprehensive examination of security from an Economics perspective, and developed on the theme with others [4]. They observe, for example, that economists have long studied how misaligned incentives produce undesired outcomes, and many of these results carry lessons for security. The study of negative externalities, where economic actors do not bear the full cost of their actions also has great applicability in network and internet security.

Since 2001, the Workshop on the Economics of Information Security has explored these and other areas of overlap between economics and security. For example, there has been much interesting work on the establishment of a market for security vulnerabilities [27] and the Economics of Privacy [1].

Many works have studied the economics and mechanisms that govern the behavior of the "good guys" and study how things can be made better. There has also been examination of the economics and mechanisms that govern the behavior of the "bad guys" and study how things can be made worse. Fultz and Grosslags [13] examine the case where, like the defenders, attackers are in a resource constrained environment. When there are too many of them, all seeking easy returns, yield falls. Two related papers propose to insert uncertainty in the botnet infrastructure, with the objective of increasing uncertainty in the service provided by the botnets, thus reducing its value. Ford and Gordon [11] propose that once a machine is recovered from a botnet, instead of letting the botnet master know, we maintain association with the botnet, and even increase the click/display rate. This would increase the uncertainty, reducing the value of the service provided by adware. Li *et al.* [24] propose increasing uncertainty in the botnet economy by setting up honeypots and allowing them to infiltrate the botnet environment, increasing uncertainty in how many machines a botnet really has available for a denial of service attack.

In an earlier work we examined the Economics of phishing [19]. We found that phishing is a classic example of Tragedy of the Commons where open access to a shared resource drives the total returns down. One of our surprising findings was that total *direct* dollar losses from phishing appear to be far lower than generally thought. However, this fits neatly with the notion of security as an externality [4]: the direct dollar losses are far from being a complete accounting of the problem. We explore that question further in Section 3.4.6.

3.2.3 Economics Background

3.2.3.1 Asymmetric Information: The Market for Lemons

In a classic paper, Akerlof [2] examined the effect of uncertainty on markets. In a situation where sellers have better information than buyers about the quality of their wares there is adverse selection and the "bad drive out the good." Choosing the specific example of used cars where the seller knows whether the car is a Lemon or not the buyer will logically factor the average likelihood of getting a lemon into the price. Thus sellers of good cars get less than their cars are worth, while sellers of lemons get more. This leads to adverse selection where sellers of lemons are attracted to the market, while sellers of good cars tend to stay away. This increases the the percent of lemons in the market driving the average quality further down.

Where there is a continuum of quality the problem can lead to a market failure. Suppose a product has quality q uniformly distributed in the range $[0,1]$. Suppose that for every q there are sellers who are willing to sell their product for any price

above q, and buyers who are willing to buy for any price below $3q/2$. The price would then achieve equilibrium at some point between q and $3q/2$ if quality were observable. However, since the seller knows the quality while the buyer does not, the buyer can only decide his price based on the average or expected quality. At any possible equilibrium price p, only products in the quality range $[0, p]$ will be offered for sale, so that the average quality is $p/2$. However buyers will pay only $3p/4$, for a product of expected quality $p/2$ and thus no trades happen.

What is interesting is that even though willing buyers and willing sellers exist for products at every quality in the range $[0, 1]$ no trades happen. There are buyers who would happily pay $3q/2$ for a products of quality q and sellers who will take this price. But the buyer has no way of verifying that the product is really of the claimed quality, and the seller has no way of credibly disclosing q. A lemon market will be produced by the following:

- Asymmetry of information, in which no buyers can accurately assess the value of a product through examination before sale is made and all sellers can more accurately assess the value of a product prior to sale
- An incentive exists for the seller to pass off a low quality product as a higher quality one
- Sellers have no credible quality disclosure technology
- Either there exist a continuum of seller qualities or the average seller type is sufficiently low
- Deficiency of effective public quality assurances (by reputation or regulation and/or of effective guarantees / warranties)

Akerlof suggested that lemon Markets existed in the market for used cars, the insurance and job markets, and in the market for debt in underdeveloped economies.

The claim that security products are a market for lemons has been noted [3]. That is the buyer often has a poor understanding of the risk mitigated and the protection gained, and is poorly equipped to make an informed distinction between a good security product and a bad one. Grigg argues that security products are actually a market for silver bullets [17] since neither buyer nor seller actually understands the risks. While it is interesting that the market for lemon theory has been applied to security goods before, the argument we advance is quite different. We argue that several of the goods traded in the underground economy satisfy the criteria for a market for lemons.

3.2.3.2 The Theory of the Firm

A subject of great interest in Economics is the theory of the firm. That is, why do firms exist instead of letting the market decide all prices. For example, why does it make sense for a company to have long term employees rather than purchase labor as needed in the market.

Coase [7] advanced the transaction cost theory of the firm in 1937. When the transaction costs are high or uncertain it is advantageous to form firms.

3.3 The Underground Economy is a Market for Lemons

As discussed in Section 3.2.3.1, there are a number of factors that lead to a market for lemons: asymmetry of information; incentive to inflate quality claims; lack of credible disclosure; continuum of seller quality or poor seller quality; and lack of public quality assurance or government regulation. That there is an incentive to inflate quality claims requires no demonstration. We now go through each of the other criteria in turn and demonstrate that they hold for the goods and services offered for sale in the underground economy.

3.3.1 The Types of Goods and Services Offered for Sale on the Underground Economy

3.3.1.1 Goods

Thomas and Martin [31] mention the following goods being advertised on the IRC channel they monitored: CCNs, credentials, scam (phishing) kits and compromised hosts. Franklin *et al.* [12] on a similar channel mention the most common goods being "online credentials such as bank logins and PayPal accounts, sensitive data such as credit cards and SSNs, compromised machines, spamming tools including mailing lists and open mail relays, and scam webpages used for phishing."

Symantec in 2008 [29] tabulates the goods and services offered for sale, which we reproduce in Table 3.1. The dominance of CCNs is borne out by Thomas and Martin, Franklin *et al.*, Dhanjani and Rios [9, 12, 31].

Good or Service	Percent of offerings	Asking price range
Bank account credentials	18%	$10-$1000
Credit Card Numbers (with CCV2)	16%	$0.50-$12
Credit Cards	13%	$0.1-$25
Email addresses	6%	$0.30/MB - $40/MB
Email passwords	6%	$4 - $30
Full identities	6%	$0.90 - $25
Cashout Services	5%	8%-50% of total value
Proxies	4%	$0.30 - $20
Scams	3%	$2.5-$100/week for hosting
Mailers	3%	$1-$25

Table 3.1 Goods and services offered for sale on an underground economy IRC market [29].

3.3.1.2 Services

Thomas and Martin [31] refer to cashiers and drops as the most sought after services in the underground economy. According to Franklin *et al.* the "most common service ad are offers for the services of a cashier, a miscreant who converts financial accounts to cash" [12]. They also find that Confirmers (who answer confirmation questions on the phone) are requested. They find "a small percentage of service ads offer services such as DoS attacks, sending phishing emails, and purchasing goods with other's credit cards (a.k.a., carding)." Other services include drops (physical locations where goods can be sent). Again these findings accord well with those of Symantec as shown in Table 3.1.

3.3.2 Is this a Market for Lemons?

3.3.2.1 Asymmetry of Information

Why does the seller have better information than the buyer as to the value or quality of a set of credentials, CCN, *etc*? First and foremost, the seller knows whether he is a ripper or not. This effect probably dwarves all others.

In addition the most common goods offered for sale on the underground economy are information goods, where quality is hard to determine. The seller knows the balance or available credit limit in the account, while the buyer does not. Also, recall that the buyer requires not merely access to the information, but *exclusive access to the information*. Take for example the stolen credentials for a WellsFargo account with balance $2000. They seller knows the balance, while the buyer must take his word for it (until he gets the password). The value to the buyer might be the full account balance if the buyer can successfully drain the account completely. But this is only possible *if he is the only person attempting to do so*. Nothing prevents the seller selling the same information many times over. If the same credential is sold multiple times each buyer will be competing against an unknown number of would-be harvesters and the return that he can expect changes drastically. The same is true for each of the information goods that the underground economy studies find: if the buyer must compete to harvest the resource his expected return changes drastically. This is certainly true of CCNs and login credentials.

For any type of software application (*e.g.* scam phishing kits, keyloggers *etc*) the situation is even worse: the buyer has no way of determining quality. Anyone who purchases an application runs the risk that it carries an unannounced malicious payload. Phishing kits and keyloggers traded on the underground economy have been found to contain concealed backdoors that remit any information harvested to the author [8, 9]. Again the buyer risks putting himself in competition for the credentials he harvests with others.

Even when dishonesty is not involved some goods have unobservable quality. Mailers and proxies, for example, have useful lifetime related to how much they

have previously been used. A phisher who buys an email list and mailer tool to advertise a phishing attempt on PayPal will clearly have lower yield if the same list and tool have been used to advertise several other PayPal phishing sites that week. Proxies that have been extensively used are much more likely to find their way onto blacklists.

Each of these exploits is a question of degree. Of course a ripper also has the simple recourse of entirely failing to deliver once payment is received. For the services traded on the underground economy the uncertainty is whether the seller will perform as advertised. A cashier can fail to hand over the proceeds of a transaction and keep 100% for himself. And a drop can fail to hand over the delivered merchandise.

3.3.2.2 No Credible Disclosure

In addition to having better quality information than the buyer the seller has no credible way of disclosing this information to the buyer. A seller who does not intend to cheat, merely subsidizes those who do. Attempts to disclose quality are referred to several of the studies available. For example (from [31]): "One miscreant even provided a screen shot of a compromised Wells Fargo account, with a net total of US $21,431.18 in cash." However it is difficult to see what assurance this offers: altering the account balance on a screenshot hardly represents a challenge.

Even in the case where the seller offers the buyer a chance to verify the account balance this does not help much. The only thing the buyer can do to guarantee exclusive access is to immediately change the password (and password reset mechanisms) of the account. This might seem an attractive way of excluding any others from the account. This is not feasible however, as for most financial account this generates an email to the user informing them that the password or other information has changed.

The same is true for the services offered. A cashier who will drain an account and remit the proceeds has no credible way of disclosing whether he will perform honestly or not.

3.3.2.3 Continuum of Seller Quality or Low Seller Quality

The evidence certainly indicates that the average seller quality in the underground economy is extremely low, and cheating and dishonesty are rampant. Thomas and Martin [31] introduce us to the term ripper: a market participant who does not provide the goods or services he's been paid for. This phenomenon appears to be widespread. For example, Franklin *et al.* document a daily average of 490 credit card numbers being posted on an IRC market; however fully 22% of them failed to satisfy the Luhn checksum (*i.e.* they are no better than random 16-digit numbers). They also find evidence that various services offered by the administrator of the channel they monitored were designed to trick participants. For example, commands that check the validity, credit limit and validation number of credit cards were avail-

able: !chk, !cclimit and !cvv2. However they did not function as advertised, leading to the suggestion that they are merely a simple way for the channel administrator to steal CCNs from participants. Similarly, Dhanjani and Rios [9] demonstrate the backdoors that some phishers insert in kits so that they can harvest the fruits of other phishers' labor.

Symantec (see Figure 1 of [29]) show a screenshot of a IRC channel with six messages, two of which end with the line "Ripper #$& off" and one of which (for a cashier) promises "you can trust me 100%." Symantec also reports that "Many underground economy servers have channels specifically created by the server administrators as a direct forum to report and list current rippers to avoid. Repeat rippers can be kicked off and banned from the servers." Clearly cheating and dishonesty are a way of life on the underground economy markets, making average seller quality low. Since there is no barrier to entry, it is difficult to imagine a mechanism that would keep seller quality high.

3.3.2.4 Lack of Quality Assurance or Regulation

There are several ways to ensure the functioning of a market in the presence of quality uncertainty. Lemon laws, product warranties, and return policies are efforts to protect the buyer against a bad transaction. However, this clearly works only when there is an enforcement mechanism; and (according to [29]) "In the underground economy, buyers have no recourse to obtain refunds for unsatisfactory goods or services." Further, e-gold, the predominant payment mechanism in the underground economy, promises anonymous irreversible transactions.

The natural way to combat this is to establish a seller reputation mechanism. Indeed even legitimate markets such as eBay require a reputation system to function well. However the reputation system referred to by [12, 29–31] is very basic: "To establish a reputation and prove themselves, potential sellers are often required to provide samples of their goods for validation and verification." Usernames (nics) are either verified or un-verified, and there is no reference to a more complex seller reputation system. In fact the dedicated channels to report rippers appear the only central reputation system. And (as [29] points out): "if an advertiser is accused of being a ripper, he or she can simply switch nicknames and start anew."

Of course individual sellers may perform honestly. But in the absence of a trustworthy seller reputation system this information is diffused among many buyers. Performing honestly in a transaction will effect his reputation with a single buyer, but does not impact his overall reputation in the market. Further, the fluidity with which IRC channels set up and shut down makes complex reputation systems difficult. In addition, as Franklin *et al.* point out a simple slander attack on the reputation of a good seller is not merely possible but profitable for rippers: in assailing good reputation they can drive other sellers from the market and decrease the disadvantage caused by their own lack of reputation [12].

3.3.2.5 Summary

Thus we conclude that each of the goods and services traded on the underground economy indeed satisfy the conditions for a market for lemons. Indeed they satisfy the criteria more faithfully than used cars, since with cars quality is not entirely unobservable, and reputation and enforcement mechanisms do exist. In each of the goods and services offered for sale in the underground economy we find that dishonesty and misrepresentation is not merely possible, but is actually observed and appears very frequent.

Alternatively, suppose not. Suppose the underground economy does not operate as a lemon market and every seller is honest. In this functioning market a single participant who is willing to cheat has an endlessly profitable opportunity.

3.4 Analysis and Implications

3.4.1 Countermeasures Ought to be Easy: Lemonizing the Market

The idea of inducing market failure by increasing the quality uncertainty in the market is suggested by Franklin *et al.* [12]. They suggest counter-measures which involve generating many sybil accounts, achieving verified status for each of them, and then conducting deceptive sales. The last step involves offering no-value goods for sale at the market. For example, suppose CCNs are sold at the market at a going rate of $1.25. This may be because the sellers acquisition cost is, say $1.00, and the buyer is able to collect $1.50 on average from each account. Suppose further, there are about 1000 CCNs offered for sale each day on a certain bulletin board. By simply offering another 1000 worthless CCNs for sale, we reduce the expected value per card to $0.75. Since that is below the acquisition cost of the seller, no trade would take place at all even though willing buyers and sellers are both present. It is worth differentiating the above from a Denial of Service attack which would involve bombarding the market. Here just a handful of messages would be enough to cause failure.

This attack makes a great deal of sense. However, as Section 3.3 shows, the dishonesty and greed of the participants require little encouragement or assistance. They can and do lie, steal and cheat. They are already performing all of the actions that Franklin *et al.* suggest as counter-measures. Thus the underground economy satisfies requirements for a market for lemons perfectly, and the counter-measures to attack it appear to be already extensively practised by the market participants themselves.

3.4.2 The Ripper Tax

In effect, the uncertainty created by the presence of rippers imposes a tax on every transaction conducted in the market. Suppose, for a buyer there is a probability q that a transaction is with a ripper, and $1-q$ that it is with a legitimate seller. Thus a fraction q of all transactions result in money leaving the system without goods or services changing hands, much as happens with a tax. Both buyers and sellers share this burden [25]. That is, in the ripper-infested market, buyers pay more and sellers receive less than they would in an untaxed one. Further, the amount of market activity is reduced by the taxation [25], *i.e.*, the overall transaction amount decreases.

It is natural to wonder whether we can estimate the tax rate q. Since none of the underground economy studies [12, 29–31] observe even a single transaction closing we clearly cannot estimate the fraction of trades where one party is a ripper. However, basic economics and the asking price of goods on the underground economy both suggest that the tax rate is high. First, when a single agency, like a government, applies a tax their goal is to maximize the total tax receipts from the market. If it taxes too heavily activity drops and the return falls. However, the ripper tax is a result of many independent actors each seeking to maximize his personal return. Thus there is a Tragedy of the Commons [18]: rather than show restraint and nurture their collective resource the rippers maximize their independent profit. The result is a higher tax rate, but lower overall return than the profit maximizing rate [16]. This suggests that rippers drive the tax rate q as high as they can without extinguishing the market entirely. Second, the gap between the asking price for a CCN and its expected fraud value (*e.g.*, \$350 according to an FTC victim survey [10]) is due to banks successfully detecting fraud, the premium that the buyer demands to ensure a profit and the ripper tax. The size of the gap suggests the ripper tax must be large.

For example, if banks successfully prevent 90% of fraudulent activity the expected value of a CCN would be \$35 rather than \$350. To choose a round number let's assume choose \$3.50 as a selling price from the range given by Symantec. In a pool of CCNs for sale, a fraction $1-q$ are good, and q are offered by rippers. A buyer pays \$3.50 for CCNs and commits fraud worth \$35 on the fraction q of them that are good. Thus, if the buyer demands a 100% premium (*i.e.*, that he double his investment) to make the risk worthwhile, we get $\$35 \times (1-q) - 3.5 = 3.5$, giving $q = 0.8$. Thus, CCNs sell for \$3.50, but 80% of those are offered by rippers. If we consider 1000 CCNs sold then sellers will get $\$3.5 \times 200 = \$700$. Buyers get $\$35 \times 200 - 3.5 \times 1000 = \$3500$. Rippers get $\$3.5 \times 800 = \$2800$.

3.4.3 Formation of Firms and Alliances

Taxation of a market is one of the circumstances that Coase [7] identifies as leading to the formation of relationships and ultimately firms. The idea is that when market transactions are taxed, expensive, or uncertain it makes sense to form groups who

deal with each other regularly rather than return to the market for every resource requirement. We can readily see how this happens in the underground economy. After a transaction with a good seller it makes sense to deal with that seller again rather than brave the mixed pool of sellers and rippers. Thus, dealing with someone successfully increases the likelihood that one will deal with them again, since doing so eliminates the ripper tax from the transaction. This is corroborated by each of the studies of the underground economy. For example, Thomas and Martin find [31]: "Those who provide services in the underground economy are looking for long-term customers." Similarly Franklin *et al.* and Symantec find the desire to form partnerships is strong.

There is some evidence that integrated gangs, rather than individuals, are responsible for much online crime. In phishing, for example, the Rock-phish gang has been credited with perpetrating about 50% of all attacks [26]. Moore and Clayton find that their attacks are better organized and harder to measure. In examining a large spam campaign launched from the Storm botnet Kanich *et al.* [22] find evidence that the spam is sent on behalf of the botnet controllers, rather than sent as a service for a fee. First, the return is very low, indicating that the service could not be profitably rented for the quoted asking prices. Second, similarity between email addresses used in propagating the botnet and the spam campaign suggests the same people are behind both. Further evidence is given by the concentration of exploits in certain countries and in certain language groups. Four well organized Russian and Ukranian gangs appear to be behind much bot herding and spam campaigns [6].

3.4.4 A Two-Tier Underground Economy

The argument we have advanced suggests a two-tier system where those who are organized avoid the ripper tax, while those who frequent the IRC channels have higher costs and lower profitability. As the better organized competitors with lower costs those in the upper tier probably enjoy the bulk of the profits. That is, those who see a good return on their investment of time probably belong to gangs that form integrated chains to extract all of the value from their product without having to frequent markets where there is a risk of rippers.

It would also appear that entering the upper tier requires performing as a profitable partner to existing members of the upper tier. Thus, those who possess only commodity skills are unlikely to enter. It is hard to see why an existing alliance or gang in the upper tier would share profits for goods or services that are easily acquired. Upper tier gangs will extract all the value from any resources they control. Thus, as in other spheres, those with few skills who arrive in the underground economy are relegated to the low paying margins. If they succeed in harvesting CCNs or credentials they must sell in ripper infested markets. Further, since they compete with better organized competitors who have a lower cost basis, it appears likely that those who trade on the IRC channels struggle with profitability.

We have argued elsewhere that US phishing losses are about $60 million annually [19]. However, it is probable that the bulk of this gain is concentrated in the hands of the upper tier, while the lower tier makes only their opportunity costs. Levitt and Venkatesh [23] suggest that drug dealing is modeled as a tournament, where participants accept low-pay and high-risk for a small chance of large reward. It is interesting to wonder whether a similar phenomenon might not be at work here. We explore this further in Section 3.4.6.

3.4.5 What Can We Estimate From Activity on IRC Markets?

3.4.5.1 What Can We Say about Participants in a Lemon Market?

So why then does the market exist at all? Why does anyone offer goods for sale when they have no way of differentiating themselves from rippers? Even if commerce on IRC channels is taxed, there are various reasons why people will continue to participate in the market:

1. They need to form relationships (with a view to avoiding the ripper tax)
2. They are newcomers and are trying to get started
3. They wish to sell resources that have no value to them
4. They intend to cheat others (*i.e.* they are rippers).

First, while the underground economy servers may represent a dis-functional market it may also be the only way to get required goods or services. For many with criminal services to buy or sell, this is simply the gathering place to meet others with whom one can form mutually beneficial relations. There may be no alternative to a few unprofitable transactions with rippers to find partners with whom one can deal profitably on an ongoing basis. Second, for newcomers this looks like a particularly dangerous place, but they may know no better and have little choice. It appears that offers to help almost universally end up being an attempt to cheat or profit from the newcomer [8, 9]. Third, it certainly makes sense that participants will sell goods or services that they are unable to monetize. For example, if one has CCNs or stolen credentials that one is unable to extract value from, it makes sense to sell them to those who can, even if much revenue is stolen on the way. Also, those who have tried spamming or phishing and found it unprofitable may find it easier to sell services to others who have yet to reach that conclusion [19]. Finally, for rippers the IRC markets appear an ideal playground. But life is competitive, even for rippers: the Tragedy of the Commons [18] again suggests that rippers will overgraze the underground economy markets and drive overall returns down. The laws of economics haven't been suspended: not for those who steal, nor for those who steal from those who steal.

3.4.5.2 Activity Does not Imply Dollars

Most of the publicly available data on the underground economy documents activity [12,14,29,31]. It is almost universal to take this as a evidence of profit. We argue that this is profoundly in error. One cannot estimate the gold in the mountains from the activity at the shovel store. In none of the studies of the underground economy do we have examples of transactions actually closing. For legal, ethical and logistical reasons there is not a single confirmed instance of a sale of illicit goods documented in [12, 29–31].

Symantec [29, 30] uses measured activity to estimate the size of the underground economy. It reports the total asking price of goods offered for sale on all the IRC servers it monitored as $276 million. Of this 59%, or $163 million was CCN related. They then estimate the potential value of these CCNs as $5.3 billion, by assuming that each card suffers the median CCN fraud loss of $350 [10] rather than the $0.50 to $12 for which they are offered for sale. There are a number of problems with this approach.

First, offered for sale does not mean sold. We have no data on what percent of goods offered for sale get sold. Recall the spam campaign which achieved 27 sales for 350 million emails sent [22]. Indeed, if we applied the assumption that everything advertised was sold, we would conclude that that campaign would have yielded $8.75 billion rather than the $2731 actually achieved: a difference of six orders of magnitude! Second, asking price in a market riddled with dishonesty isn't necessarily an accurate indication of what the goods are worth. Taking the average of unverified numbers creates great opportunity for upward bias. Those who ask high prices and sell least affect the average most. More significantly, taking the average of offered sales includes the worthless goods offered by rippers. Finally, assuming that each offered-for-sale CCN results in $350 worth of fraud, rather than the $0.5 to $12 range for which it was offered seems unrealistic. This assumes that banks detect no fraud, and assumes that sellers allow others to extract more than 95% of the value of their product. While this is possible, a simpler explanation would be that CCNs are offered for $0.50 to $12 on the underground economy because, in expectation, they are worth no more than a small multiple of that (to account for the profit margin of the buyer).

Returning to the $163 million worth of CCNs that Symantec observed: if we assume that only a quarter of what is offered actually sells, and that buyers achieve a 100% premium (*i.e.*, double their at-risk money) we get a value of $82 million rather than the $5.3 billion. This is the total fraud from all of the CCNs that Symantec observed offered for sale.

The history of any goldrush reminds us that effort does not imply reward. For example, over 100000 prospectors attempted to reach the Klondike after gold was discovered in 1897 [5]. Of these fewer than 4000 actually found any gold, and a few hundred found enough to cover their costs and perhaps get rich. The total value of the gold extracted from the Klondike is estimated at $50 million, while the average prospector spent $1000 and Seattle merchants alone sold over $25 million worth of goods to those heading to the gold fields [5].

3.4.5.3 Activity Does Imply Competition

Even if we cannot estimate the dollar size of the merchandise traded on IRC markets there is a great deal we can learn from the amount of activity. First, this is an extremely competitive market. There is enormous activity from those seeking riches.

Second, there is a lot of cheating. This can itself be taken as evidence that many find the underground economy a very challenging environment. Newcomers are beset by offers of kits, tutorials and gear [29] much as those going to the Klondike were offered merchandize from those who preferred to trade than try their luck in the gold fields [5]. The extent to which cheating and rippers are a factor suggests that life in the underground economy is not as easy as it is often portrayed. If getting credentials and draining accounts worth thousands of dollars were simple why would anyone waste their time ripping by, *e.g.*, offering to sell non-existent CCNs for $0.50 each? This evidence suggests that rippers are better informed than their victims about the returns on exploits such as phishing and spam.

3.4.5.4 What Can We Say About the Goods Offered in a Lemon Market?

We argue in Section 3.4.5.2 that activity cannot be used to estimate the dollar size of the underground economy. But we can still learn much of its workings by observing activity. Anyone who chooses to buy or sell in a heavily taxed market clearly has few other options. The fact that he pays the ripper tax tells us that he has little alternative. For a seller this means that he cannot monetize the goods himself, and does not have access to someone who can do so for a smaller premium than the ripper tax. This suggests that the goods and services advertised on the underground economy are those that are easy to acquire, but hard to monetize.

3.4.6 Who are We Fighting? What are We Trying to Accomplish?

The picture that emerges is of a two-tier underground economy where the inhabitants of the lower tier are taxed by rippers and struggle to monetize their efforts. Why does this matter? If all we cared about were the direct losses from cybercrime it might not be important. Why should we care if one subset of cybercriminals get cheated by another? However, the gains enjoyed by participants in the underground economy are not an accurate measure of the size of the problem. For example, Kanich *et al.* show that a 350 million email campaign resulted in a mere $2731 in revenue for the spammer. Clearly, this gain is minor in comparison to the externalities: the value of the infrastructure required to handle and store this email, the spam filtering work required, and the time wasted by recipients. Similarly, with other forms of cybercrime.

While the inhabitants of the lower tier struggle to monetize their efforts it does not follow that they account for a small portion of the externalities. For example, the

bulk of the profits from phishing may be concentrated in the hands of a few gangs, but responsibility for the erosion of trust, cost of customer support calls, and expense of educating users and deploying stronger authentication mechanisms belong to all those who phish, not just those who make a profit. Consider two different phishers. An upper tier phisher who gets 100 credentials per million emails delivered into inboxes, and a lower tier phisher who makes gets one credential per million emails delivered. The contributions of these two to the direct costs of phishing are very different, while their contributions to the externalities are similar. This brings us to the important questions of who we are fighting and what we are trying to accomplish.

Who are we fighting? Those in the upper tier are engaged in a profitable activity, and are members of alliances or gangs. It is reasonable to expect that they will respond to economic and law enforcement pressures much as any other firm will. However, those in the lower tier appear to struggle with profitability. We can explain their persistence using the tournament model of the job market that Levitt and Venkatesh [23] apply to the drug trade. Newcomers accept low pay and high risk in exchange for a chance of a large reward.

What are we trying to accomplish? If we cared only about direct losses we would concentrate on the upper tier. We could effectively ignore the lower tier, since they gain little for their efforts. However, the evidence from spam [22] and phishing [19] is that the direct losses are minor compared to the externalities, *i.e.*, indirect losses. To reduce the externalities we must fight both upper and lower tiers.

Unfortunately, if the lower tier is largely unprofitable, and acts as a tournament job market it may be relatively impervious to economic pressures and law enforcement. Participants are striving for their chance at reward and are willing to endure difficulties, risk and loss. Thus the tools that the upper tier responds to have less influence on the lower tier. However, this does suggest a third approach: if lower tier participants are misinformed about the true likelihood of winning, *i.e.*, overestimate the rewards then it may be possible to influence them by publicizing accurate information. That is, as those who are new and inexperienced lower tier participants believe that the underground economy is a path to easy riches. Where would they get that idea? From the same place the rest of us get that idea: unreliable and exaggerated estimates repeated without scrutiny. We suggest that accurate estimates are not just interesting from a research standpoint, but can have material influence on the recruitment of our opponents.

3.5 Conclusion

The underground economy is often painted as an easy money criminal Utopia where even those without skills can buy what they need and sell what they produce. They can buy phishing kits, rent hosting services and then profitably sell the credentials they produce on IRC channels. This picture does not withstand scrutiny. The IRC markets on the underground economy represent a classic example of a market for

lemons. The rippers who steal from other participants ensure that buying and selling is heavily taxed.

Avoiding the ripper tax reduces costs and increases profitability. Those who can do so extract all the value from their resources. Those who cannot have no alternative but to trade on IRC channels where cheating is a way of life. This suggests a two tier underground economy: gangs and alliances that can extract value from their resources form the upper tier. Those who must buy resources, or who cannot monetize the credentials they steal, form the lower tier. They have no choice but to pay the tax that the rippers extract.

We find that the published estimates of the dollar value of underground economy IRC channels are exaggerated. They are derived by simply adding the unverified claims of anonymous channel participants. Those who lie most and exaggerate most affect the average most. We emphasize that the activities of the upper tier are largely invisible and probably account for a majority of the losses.

An important conclusion is that goods offered for sale on the IRC channels are hard to monetize. Those who sell there are clearly unable to monetize the goods themselves or find someone who will do so for a smaller premium than the ripper tax. Since stolen CCNs and bank credentials are a majority of the goods offered for sale this implies that getting credentials is only a first step, and by no means the most important one, in the chain of fraud.

We find that different means are necessary to fight the two tiers. The alliances and gangs of the upper tier act as businesses and will respond to economic and law enforcement pressures. Those in the lower tier are harder reach with these means. While they make little they generate very significant externalities.

Ironically defenders, *i.e.*, whitehats, security vendors and members of the InfoSec community, actively and energetically recruit their own opponents. Repeating unverified claims of cybercrime riches, and promoting the idea of easy money for all, attracts new entrants into the lower tier of the underground economy. While they may produce no profit they still generate large quantities of spam and phishing and cause significant indirect costs. There is a further irony that both upper and lower tier cybercriminals, internet users, financial institutions and the InfoSec community all have interests that are aligned on the matter have having accurate data free of exaggeration. This is so since an accurate accounting of their prospects might cause many in the lower tier to leave the underground economy. Most obviously internet users, banks and financial institutions would be better off and the InfoSec community could concentrate on a smaller if abler upper tier. Less obviously, those in the upper tier would benefit from decreased competition. Finally, those in the lower tier would benefit as they would be spared wasting their time on what, for most of them, will be a profitless endeavor. The only people who benefit from exaggerated and inaccurate underground economy estimates appear to be the rippers.

References

1. Acquisti, A., Grossklags, J.: Uncertainty, ambiguity and privacy. In: Proceedings of the Fourth Workshop on the Economics of Information Security (WEIS). Cambridge, MA (2005)
2. Akerlof, G.A.: The market for 'lemons': Quality uncertainty and the market mechanism. Quarterly Journal of Economics *84*(3), 488–500 (1970)
3. Anderson, R.: Why information security is hard – an economic perspective. In: Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC), pp. 358–365. (2001)
4. Anderson, R., Moore, T.: The economics of information security. *Science* **314**(5799), 610–613 (2006)
5. Berton, P.: Klondike: The Last Great Gold Rush, 1896-1899. Mcclelland and Stewart (1972)
6. Brady, H.: The EU and the fight against organised crime (2007). `http://www.cer.org.uk/pdf/wp721_org_crime_brady.pdf`
7. Coase, R.H.: The nature of the firm. Economica **4**(16), 386–405 (1937)
8. Cova, M., Kruegel, C., Vigna, G.: There is no free phish: an analysis of "free" and live phishihg kits. In: Proceedings of WOOT. USENIX Association, Berkeley (2008)
9. Dhanjani, N., Rios, B.: Bad Sushi: Beating Phishers at their Own Game. *Blackhat*, 2008.
10. Federal Trade Commission: Identity theft survey report (2007). `www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf`
11. Ford, R., Gordon, S.: Cent, five cent, ten cent, dollar: Hitting spyware where it teally hurt$. In: Proceedings of the New Security Paradigms Workshop (NSPW), pp. 3–10. ACM Press, New York (2006)
12. Franklin, J., Paxson, V., Perrig, A., Savage, S.: An inquiry into the nature and causes of the wealth of Internet miscreants. In: Proceedings of the ACM Conference on Computer and Communications Security (CCS), pp. 375–388. ACM Press, New York (2007)
13. Fultz, N., Grossklags, J.: Blue versus red: Toward a model of distributed security attacks. In: R. Dingledine, P. Golle (eds.) 13th International Conference on Financial Cryptography and Data Security, *LNCS*, vol. 5628, pp. 167–183. Springer, Berlin Heidelberg (2009)
14. Geer, D., Conway, D.: The 0wned price index. IEEE Security & Privacy **7**(1), 86–87 (2009)
15. Geer, D., Conway, D.: What we got for Christmas. IEEE Security & Privacy **6**(1), 88 (2008)
16. Gordon, H.S.: The economic theory of a common-property resource: the fishery. *The Journal of Political Economy* **62**(2), 124–142 (1954)
17. Grigg, I.: The market for silver bullets (2008). `http://iang.org/papers/market_for_silver_bullets.html`.
18. Hardin, G.: The tragedy of the commons. *Science* **162**(3859), 1243–1248 (1968)
19. Herley, C., Florêncio, D.: A profitless endeavor: phishing as tragedy of the commons. In: Proceedings of the New Security Paradigms Workshop (NSPW), pp. 59–70. ACM Press, New York (2008)
20. Holz, T., Engelberth, M., Freiling, F.: earning More About the Underground Economy: A Case-Study of Keyloggers and Dropzones. Reihe Informatik. TR-2008-006 (2008). `http://honeyblog.org/junkyard/reports/impersonation-attacks-TR.pdf`
21. John, J.P., Moshchuk, A., Gribble, S.D., Krishnamurthy, A.: Studying spamming botnets using Botlab. In: Proceedings of NSDI. USENIX Association, Berkeley (2009)
22. Kanich, C., Kreibich, C., Levchenko, K., Enright, B., Voelker, G.M., Paxson, V., Savage, S.: Spamalytics: An empirical analysis of spam marketing conversion. In: Proceedings of the 15th ACM Conference on Computer and Communications Security, pages 3–14. ACM Press, New York (2008)
23. Levitt, S.D., Venkatesh, S.A.: An economic analysis of a drug-selling gang's finances. Quarterly Journal of Economics **115**(3), 755–789 (2000)
24. Li, Z., Liao, Q., Striegel, A.: Botnet economics: Uncertainty matters. In: M.E. Johnson (ed.) Managing Information Risk and the Economics of Security, pp. 245–267. Springer, New York (2008)
25. Mankiw, N.G.: Principles of Economics. South-Western College Publishers (2007)

26. Moore, T., Clayton, R.: Examining the impact of website take-down on phishing. In: Proceedings of the 2nd APWG eCrime Researchers Summit, pp. 1–13. ACM Press, New York (2007)
27. Ozment, A., Schechter, S.: Milk or wine: does software security improve with age? In: Proceedings of the 15th USENIX Security Symposium, article no. 7. USENIX Association, Berkeley (2006)
28. Smith, A.: An Inquiry into the Nature and Causes of the Wealth of Nations. W. Strahan and T. Cadell (1776)
29. Symantec: Symantec Internet Security Threat Report XIII (2008). http://eval.symantec.com/mktginfo/enterprise/white_papers/b-whitepaper_internet_security_threat_report_xiii_04-2008.en-us.pdf
30. Symantec: Symantec Report on the Underground Economy XII (2009). http://eval.symantec.com/mktginfo/enterprise/white_papers/ent-whitepaper_internet_security_threat_report_xii_09_2007.en-us.pdf.
31. Thomas, R., Martin, J.: The underground economy: Priceless. USENIX ;login **31**(6), 7–16 (2006)
32. Zhuge, Z., Holz, T., Song, C., Guo, J., Han, X., Zou, W.: Studying malicious websites and the underground economy on the Chinese web. In: M.E. Johnson (ed.) Managing Information Risk and the Economics of Security, pp. 225–244. Springer, New York (2008)

Chapter 4
Security Economics and Critical National Infrastructure

Ross Anderson and Shailendra Fuloria

Abstract There has been considerable effort and expenditure since 9/11 on the protection of 'Critical National Infrastructure' against online attack. This is commonly interpreted to mean preventing online sabotage against utilities such as electricity, oil and gas, water, and sewage – including pipelines, refineries, generators, storage depots and transport facilities such as tankers and terminals. A consensus is emerging that the protection of such assets is more a matter of business models and regulation – in short, of security economics – than of technology. We describe the problems, and the state of play, in this paper. Industrial control systems operate in a different world from systems previously studied by security economists; we find the same issues (lock-in, externalities, asymmetric information and so on) but in different forms. Lock-in is physical, rather than based on network effects, while the most serious externalities result from correlated failure, whether from cascade failures, common-mode failures or simultaneous attacks. There is also an interesting natural experiment happening, in that the USA is regulating cyber security in the electric power industry, but not in oil and gas, while the UK is not regulating at all but rather encouraging industry's own efforts. Some European governments are intervening, while others are leaving cybersecurity entirely to plant owners to worry about. We already note some perverse effects of the U.S. regulation regime as companies game the system, to the detriment of overall dependability.

Ross Anderson
Computer Laboratory, Cambridge University, e-mail: Ross.Anderson@cl.cam.ac.uk

Shailendra Fuloria
Computer Laboratory, Cambridge University, e-mail: Shailendra.Fuloria@cl.cam.ac.uk

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_4,

4.1 Introduction

Modern industrial societies are highly dependent on a small number of utilities that provide power, water, and fuel. In times of conflict, attacks are carried out on enemies' generators, transformers, dams and pipelines; during the cold war, for example, the CIA inserted a Trojan into pipeline control software that the Soviets bought covertly, which caused the pumps, turbines and valves to go haywire and resulted in "the most monumental non-nuclear explosion and fire ever seen from space" in June 1982 [22]. More recently, the US-led coalition knocked out much of Iraq's generating capacity in 2003. These attacks can have serious consequences – in Iraq, for example, delays in restoring electric power were a significant factor in the discontent that led to insurrection against the occupying forces.

Terrorist groups have also targeted critical utilities. Perhaps the worst 'near miss' in recent history was an IRA attempt in 1996 to blow up the four electricity substations that supply London with much of its electricity. That project was thwarted by the police and intelligence services (it later turned out that a senior IRA commander was a British agent) but had it succeeded it would have wrecked electricity supplies to the south-east of England for many months [15]. The only comparable incident in a modern city in peacetime was a five-week outage in central Auckland, New Zealand, caused by a cascade of cable failures in 1998. This led to 60,000 of the 74,000 employees in the area having to work from home or from relocated offices, while most of the 6,000 apartment dwellers in the area moved out for the duration [14]. A power outage such as that planned by the IRA, which would have blacked out millions of people and businesses accounting for perhaps a third of Britain's GDP, would have done immense economic damage.

In the late 1990s, some writers started to point out the vulnerability of industrial control systems to online sabotage. Utility control systems have traditionally been designed for dependability and ease of safe use. They used completely private networks and thus their designers gave no thought to authentication or encryption. These networks were typically organized with a star topology, with many sensors and actuators connected to a control center. Common protocols such as DNP and Modbus enable anyone who can communicate with a sensor to read it, while anyone who can send data to an actuator can give it instructions. But private networks are expensive, and the prospect of orders-of-magnitude cost reductions led engineers to connect control systems to the Internet. The result was that many industrial control systems became insecure without their owners realizing this.

The wake-up call came ten years ago when it was realized that critical control systems might be disrupted by sending carefully chosen commands to the right IP address [7]. The concerns have mainly focused on the energy and water sectors, although very similar systems are in use in railways, manufacturing and elsewhere, and there are separate but comparable issues with telecoms. At the same time, in the late 1990s there was mounting hype about 'information warfare' whose mavens predicted that the combination of computer- and network-based attacks with propaganda would enable combatants to dominate the 'information battlespace' and gain an advantage comparable to that given by air power in previous generations [10].

After 9/11, government agencies and others started thinking systematically about vulnerabilities that might be exploited by hostile states and sub-state groups to do damage and cause alarm. One of the early fruits of this program was a series of publications in 2003 that collated information on previous incidents of online sabotage. Poster events included both directed malice, such as a wireless attack on a sewage facility in Queensland, Australia, in 2000 by a disgruntled former employee, and the unplanned effects of less directed malice, notably the shutdown of the Davis-Besse nuclear plant in Ohio in 2003 after some of its systems were infected with the Slammer worm. A database of incidents compiled by the British Columbia Institute of Technology revealed that in 2003 there had been 34 confirmed incidents worldwide of online sabotage, with a further 11 pending investigation [8]. A survey of control systems by the Idaho National Laboratory from 2004-6 revealed numerous vulnerabilities, and from 2006 there has been a growing number of publications describing threats to control systems [13]. For example, the CIA claimed in January 2008 that a cyber-attack had caused a multi-city power outage at an unspecified location outside the USA [19].

As far as we know, no-one has ever been killed by a cyber-terrorist, and this has limited the attention given by the media to the problems. Some people have even remained skeptical about whether online attacks could do real damage. So in March 2007, the Department of Energy's Idaho National Laboratory made a video demonstrating the 'Aurora vulnerability' in which a series of 'on' and 'off' commands are sent to a generator, timed in such a way as to bring it out of phase and thus destroy it. The video was released to the press in September 2007; in it, a large generating set shudders, emits smoke, and then stops [18]. This helped make clear to legislators that the confluence of the private but internally open systems used in industrial control, with open networking standards such as TCP/IP, was creating systemic vulnerabilities.

SCADA security – the protection of systems designed for Supervisory Control and Data Acquisition – thus become a hot topic. The combination of the clear societal importance of a dependable energy and water supply, the evident vulnerability of existing systems, the salience of 'cyber-terrorism' and the societal sensitization to terrorism since 9/11 have led to increasing amounts of money and regulatory effort being devoted to it. This paper is a first attempt to set out the security-economics issues that arise. It follows a talk on security economics given by the first author at the SCADA Security Scientific Symposium in January 2009 and discussions with the participants there.

4.2 Critical Infrastructure: Externalities of Correlated Failure

The first question we might ask is why the government needs to intervene at all. Surely a utility should be sufficiently motivated to protect its own assets against saboteurs – whether old-fashioned ones using dynamite, or new-fangled ones using network hacks?

We already have two common models of market failure leading to information security failure. In platforms like PCs, the combination of network effects, switching costs and low marginal costs lead to dominant-firm markets with a huge first-mover advantage; in the resulting market races, platform vendors appeal to complementers rather than users, leading to locked-in users and defective security [5]. With mobile phones, a complex supply chain leads to the chip IP owner, chip foundry, software platform vendor, network operator and application vendors – all trying to dump risk and liability on each other while the end users have little power [3].

Industrial control systems have both lock-in and complex supply chains. A utility that builds a plant such as a power station or oil refinery is typically locked into the control-system vendor for at least 25 years; the vendor for its part typically supplies the software for the central control function, plus the systems integration, while purchasing a wide range of equipment (cabling, sensors, actuators and indeed whole subsystems) from other vendors.

First, the lockin here has nothing to do with network effects; it's physical. The real assets of the North American energy sector are worth over a trillion dollars; control systems at major sites amount for $3 – 4 billion, while remote field devices add a further $1.5 – 2.5bn. Absent a catastrophic attack, this investment will be replaced only when it is fully depreciated. The closest model of which we are aware in the security economics literature is the study by Lookabaugh and Sicker of the U.S. cable-TV industry [16]. There, companies that buy a set-top box technology are locked in for a comparable period. The study found that while the financial effects of lockin were generally negotiated away, the effects on innovation could not be, and that this was a factor in cable TV losing ground to other channels of video distribution such as the Internet.

Second, the complex supply chains don't work in quite the same way as with mobile phones. On the one hand, there is a standards problem, and this is less tractable because relationships in the top tier of the industry are less structured. For example, on one project we might find ABB being the lead contractor, and buying subsystems from Honeywell and GE; on another project, Honeywell might lead while ABB and GE subcontract. The many smaller firms that supply specialist sensors, actuators and so on sell into numerous projects with different prime contractors. Thus, while it was possible for Nokia or ARM to push certain security technologies and standards in the mobile-phone world, it's harder in the world of control systems.

But perhaps the largest difference between the world of industrial control and the world of mobile phones (or PCs) is that the customer is far from powerless. The typical purchaser of critical infrastructure is a big utility or energy company, which has a real liability if a plant blows up. So why can't security just be left to them?

We suggest that a useful way to view this is the *large externalities of correlated failure*. If a small terrorist group – a latter-day Timothy McVeigh – were to blow up a single oil refinery, that might cost $1bn: say $500m of damage and $500m of lost profits during rebuilding. The oil company and its insurers could surely cope. However, if a more organized terrorist group – say Al-Qaida – were to blow up six oil refineries, then chaos and petrol rationing would ensue, with significant damage to the economy. For example, Britain suffered a strike by fuel-tanker drivers in 2001

that caused major disruption for weeks; the loss of six oil refineries might have a comparable impact but for a year or more, leading to social costs in the tens or even hundreds of billions.

The oil company does not internalize the social costs of this, so will make the fence high enough only for a $1bn single-incident loss. If the additional risk of a $100bn multiple-incident loss is to be dealt with, the state may have to step in. Correlated failure can take many forms. It can result from simultaneous targeted attacks, whether physical attacks as planned by the IRA or cyber-attacks; it could also result from untargeted attacks, such as the Slammer worm that shut down the Davis-Besse nuclear plant; there could be a simultaneous failure, as was feared might happen due to the "millennium bug"; and there are also cascade failures, where a failure of one part of a network shifts more load suddenly to others, causing a series of trips. The Auckland failure was of this type, and they have a long history. Early power systems were independent and served limited areas; interconnecting them meant that local generator failures could be covered more easily, but the net effect was that failures became rarer but larger. For example, the Great Northeastern Blackout of 1965 left more than 25 million people in Ontario and the Northeastern USA without electricity for almost 12 hours [9]. With electricity, too, the social costs of power failure are much higher than the revenue lost by the power company itself. Security of supply is thus a legitimate public interest.

(In passing, we note that the argument for state intervention is similar in some respects to the case for financial regulation. The isolated failure of a single bank would be of little consequence; it's the risk of correlated failure that rightly worries governments. And correlated failures impose large externalities; Lehman's collapse may have cost its CEO Dick Fuld a few hundred million dollars, but it could cost the world economy over a trillion dollars.)

4.3 Regulatory Approaches

Many governments now have programs for critical national infrastructure protection. By no means all do; for example, the French government leaves pretty well alone. But even among those governments that do intervene, there is great diversity of approach. This may create an interesting natural experiment for security economists to observe.

The UK has espoused light-touch regulation. The Centre for the Protection of National Infrastructure (CPNI) is a part of the Security Service (MI5) and operates by bringing together security managers in particular sectors to share experiences and become more discerning customers; these "buyers' clubs" can exert more pressure (and better-directed pressure) on the control system vendors than individual utilities could acting alone.

The USA, on the other hand, has gone for regulation, at least in the electricity sector. The North American Electric Reliability Corporation (NERC) is a self-regulatory organization but subject to oversight of the US Federal Energy Regula-

tory Commission (FERC) and the Government of Canada. Its mission is to ensure the reliability of the bulk power system in North America. Ultimate oversight in the USA is by the Department of Energy and the Department of Homeland Security.

NERC approved a set of standards for Critical Infrastructure Protection (CIP) in June 2006; they come into force in 2009 for every firm in North America that acts as a Reliability Coordinator, Balancing Authority, Transmission Operator, Generator Operator or Load Serving Entity in the bulk power system. NERC-CIP 001 deals with sabotage reporting; it requires responsible firms to keep proper records and report all sabotage events (and disturbances due to them) to the FBI or the RCMP. NERC CIP 002 through 009 cover cyber security.

4.4 Security or Reliability?

NERC CIP 002 is about 'Critical Cyber Asset Identification'. Each responsible entity must first identify critical assets and then those cyber assets essential to their operation. Among the critical assets is any generating plant with a 'black start' capacity. This means that it can be brought up to power even if the grid is down. In case of large scale blackouts black start generators are used to bootstrap the power grid. Hydro power stations are a good example of plant with an intrinsic black start capability; the operator merely has to turn a valve to allow the water into the turbines, and the plant will spin up. Nuclear power stations on the other hand do not by default have such a capability; they need an external power source to be safely brought up to criticality. In the middle lie fossil-fuel generators, which may or may not have black-start capability depending on whether or not they have auxiliary diesel generators. An alternative black-start strategy is for a plant to have the ability to remain operating at reduced power levels while disconnected from the grid.

At the Electric Power 2008 conference, it transpired that plant managers were removing black start capability in order to not have to pay for NERC CIP compliance [23]. This carries a clear cost in terms of system-wide reliability. Some transmission operators were removing IP connectivity from their networks, thereby escaping NERC CIP, while leaving dial-up, Bluetooth and other serial communications into their networks vulnerable. In fact, one of our informants described NERC CIP as 'a giant exercise in avoidance'!

It might be more charitable to say that the regulatory regime needs some tuning. In the short term, this may involve intervention at other levels; for example PJM, a regional transmission organization that coordinates wholesale electricity movement from New Jersey down to North Carolina and as far east as Ohio, and operates power markets among more than 500 firms, is considering allowing NERC CIP compliance costs for black start facilities to be recoverable [20]. But in the medium-to-long term, it is not advisable to have continental and regional regulators pulling in different directions.

The lesson to be learned is that security and reliability should be treated together; the proper target of the regulatory process is the sum of the two, namely depend-

ability. The electricity should continue to come out of the wall socket, regardless of the attempts of either Murphy or Satan to interrupt the supply.

4.5 Cross-Industry Differences

In North America, the electricity industry may be closely regulated, but oil and gas are almost totally unregulated, at least at the level of the control systems themselves. In these industries, the pressure comes from the major companies themselves who exert pressure primarily through the tendering and contracting process. There is indirect regulation through Sarbanes-Oxley, which has given some impetus to their information security strategy.

The oil and gas companies also have much stronger risk management. While failures of electricity supply tend to be merely inconvenient (unless they go on for a long time as in Auckland), explosions at oil and gas facilities tend to be expensive, in terms of lives, dollars and publicity. For example, an explosion at BP's Texas City refinery in March 2005 killed 15 workers and injured over 170 others. BP has paid $1.6 billion compensation to victims and has offered to pay a $50m fine. Its CEO retired early. This is by no means an isolated incident; explosions, spills, and other accidents happen regularly costing serious amounts of money. As a result, large oil companies have long embedded safety and security procedures driven by formal risk-management processes [1]. (In fact BP has taken the lead within the industry in preaching the gospel of SCADA security.)

4.6 Certification and Lifecycle Management

The collision between the proprietary world of industrial control systems and the open world of IP-based networking was a root cause of the current problems with SCADA security. The Internet offers huge cost savings over proprietary networks, and – as in other applications such as banking – there was first a rush to use the new technology to save money, then a realization that a lot would have to be spent on security in order to deal with the suddenly increased risk of remote attacks. Control systems engineers and vendors are therefore now coming into contact with traditional information security mechanisms, such as patch management and Common Criteria evaluations. A number of tensions are becoming evident.

The security-economics literature has many papers on the costs and incentives that drive lifecycle management [2]. However, common platforms either get routinely patched every month (PCs) or else replaced frequently (mobile phones). Control systems may remain in use for decades, and many of their components were never designed for remote upgrade. The costs of taking down (say) a nuclear power plant to patch components may also be very substantial, while some systems require 99.999% availability – which translates into less than 6 minutes downtime per

annum. The upshot is that control systems are patched late or not at all. Patch management has thus become contentious, with some firms believing that vulnerability information should not be published, and arguing in favour of a private CERT or even just reporting to the FBI/RCMP as mandated by NERC CIP. (This appears to be particularly the case with firms from a defence background, while firms whose SCADA business evolved from a civil engineering or computing business tend to favour the normal CERT approach.)

Matters are made more complex by the question of what to certify. In respect of legacy systems that cannot feasibly be patched, there used to be a get-out: an 'unless technically infeasible' clause in CIP. That is now being removed, and legacy systems are being protected by firewalls of various kinds. There, a 'normal' approach of frequent upgrades and CERT notification of vulnerabilities may apply to the firewall itself; there is the separate question of the rules applied by the firewall to protect the vulnerable devices behind it. The Department of Homeland Security has taken a step into this debate by issuing recommended practice for patch management of control systems according to which responsible entities must establish a patch management program dealing with hardware inventory, network mapping, software libraries and operational procedures such as patch testing and incident response [12]. This allows the asset owner to customize their plan to their circumstances, but not to just leave patch management in the 'too hard' file. However, it gives little guidance about prioritization. The difficulty of establishing good security metrics pervades this field, as it does others; the value-at-risk approach based on annualized loss expectancy does not give hard numbers unless there's adequate loss history, and the proxies used when applying security economics to traditional IT (insurance markets, stock markets and vulnerability markets) give less or no information to the control systems engineer. At least in traditional IT, we are starting to gather statistics on attacks, even although we don't have as many statistics as we'd like [6]; but there have been too few documented cyber-attacks on control systems to give us much guidance.

The move towards Common Criteria certification of protection systems and components will also raise familiar issues. Although control systems security is fundamentally about integrity and availability rather than confidentiality, there is still a multilevel element: the plant safety system should be protected from errors in (or attacks on) the control system, while the control system must in turn must be protected from the everyday systems used by office staff. Multilevel security is hard, and providing high levels of assurance is also hard. At the lower levels of Common Criteria assurance, evaluations are performed by commercial licensed evaluation facilities (CLEFs) – that is, by companies that compete for the vendor's business, giving the vendor every assurance to pick the CLEF that will give its products the easiest ride [4].

What's more, full Common Criteria certification is so slow and expensive that there will be every incentive to resort to shortcuts. The UK banks, for example, have PIN entry devices "Common Criteria evaluated" which means that they were evaluated by a CLEF, but outside the Common Criteria scheme. Such products turned out to be pathetically insecure [21]. The control systems community do not seem to realize how hard security certification can be, and the costs – especially when

layered on top of existing safety certification processes – could be very substantial. At present, U.S. regulators are mulling over whether to require control systems to undergo Common Criteria evaluation. NIST produced a Protection Profile for industrial control systems as early as 2004 [17] . This isn't the place for detailed technical discussion; we merely warn that there are significant policy issues that need to be thought through before such a step is taken. It is likely to be more expensive, and less helpful, than one might naively think.

And there are many tensions that engineers have still not begun to explore. For example, ease of safe use is a priority in control systems design, and security usability is known to be hard. Will we see conflicts between security and safe usability? As a typical plant operator earns less than $40,000, the 'Homer Simpson' problem is a real one. How do we design security that Homer can use safely?

4.7 The Roadmap

Much of the last ten years of control systems security work has been aimed at fixing the vulnerabilities that arose when previously isolated systems were heedlessly connected to the Internet. For many firms that has involved purchasing large numbers of firewalls and encryption devices so as to ensure that the traditional private networks were isolated from the Internet by an "electronic security perimeter" (as NERC CIP 005 puts it). They have thus been reconstituted as virtual private networks. However maintaining this perimeter is hard, and many incentives drive towards 'deperimeterization' (an ongoing debate in the network security community). Component vendors helpfully include new modes of communication; a transformer may now come with Bluetooth connectivity and its own web server, so that the engineer doesn't have to get out of his truck in the rain to take meter readings and adjust parameters. As fast as the security engineers can close down unauthorized access points, innovators open them up.

There is thus a growing consensus on the need to move towards a more systematic approach. Control systems should migrate to using protocols that have appropriate security measures built in to support authentication and resist service-denial attacks. There is just no feasible alternative to using commercial-off-the-shelf components in control systems, and the consequences of this have to be dealt with.

The U.S. Departments of Energy and Homeland Security therefore launched in January 2006 a Roadmap to Secure Control Systems in the Energy Sector [11], based on a 2005 workshop with asset owners and operators. Its vision is that within ten years control systems throughout the U.S. energy sector will be able to survive an intentional cyber assault with no loss of critical function in critical applications. It is not limited to engineering new control systems, but encompasses the continuing protection of surviving legacy systems, understanding strategic threats better, training, information sharing and other support activities. It focuses on critical assets, just like NERC CIP (and this does raise the issue of what happens if a worm like Blaster takes out a lot of unprotected "non-critical" systems, whose cumulative con-

tribution is critical). A significant number of technical research projects have been funded at various universities and national laboratories. A significant roadmap goal is to sustain the security improvements that this research will make possible. The roadmap acknowledges nine challenges:

1. Limited resources are available within businesses to address security needs.
2. Cyber security is a difficult business case.
3. Limited knowledge, understanding and appreciation of control systems security risks inhibit sector.
4. Insufficient sharing of threat and incident information among government and industry entities.
5. Effective security-oriented partnerships between government and industry have been difficult to establish.
6. Poor coordination among government agencies creates confusion and inefficiencies.
7. New regulation may impose requirements beyond the technical capability of legacy systems.
8. Highly educated staff with broad skill sets is needed to manage future operations.
9. Increasing sophistication of tools used by hackers.

About five of these nine fall with the classical remit of information security economics. It might therefore be appropriate for more of the research budget to be directed towards security economics research rather than purely technical projects. The security engineering community already knows how to do things like crypto, protocols, and access controls; what we don't know how to do is to ensure sustainable implementation and effective use of these technologies in different business environments.

4.8 Conclusions

Security is hard. Control systems are hard too. Control systems security will be harder; but most governments now accept that it has to be tackled. Modern societies depend completely on utilities such as electricity, oil, water and sewage, and these systems have become vulnerable to online attack.

In this paper we have looked at the state of play some ten years after this first became an issue, and some three years after the U.S. government took major policy initiatives in the form of the NERC CIP standards and the Roadmap. It is by now clear that control systems security is at least as much a security-economics problem as it is a technical one. Yet the issues are interestingly different from those studied so far by security economists. The lockin is physical rather than based on network effects; a case for government intervention may be made because of the large externalities of correlated failure; existing regulations have led companies to game the system, to the detriment of dependability; established patch management practices

conflict with control system realities; a move to Common Criteria certification could be hugely expensive; and different regulatory approaches in the USA and Europe, as well as between different U.S. industries, have created a large natural experiment for security economists to study.

Acknowledgements The second author's research is funded by ABB. The contents of this article do not necessarily express the views of ABB.

References

1. American Petroleum Institute: Security vulnerability assessment methodology for the petroleum and petrochemical industries, second edition (2004). http://www.npradc.org/docs/publications/newsletters/SVA_2nd_Edition.pdf
2. Anderson, R.: Security economics resource page (2010). http://www.cl.cam.ac.uk/~rja14/econsec.html
3. Anderson, R.: Security Engineering – A Guide to Building Dependable Distributed Systems. Wiley (2008)
4. Anderson, R.: Security Engineering – A Guide to Building Dependable Distributed Systems, chapter 26. Wiley (2008)
5. Anderson, R.: Why information security is hard – an economic perspective. In: Proceedings of the 17th Annual Computer Security Applications Conference (ACSAC), pp. 358–365. (2001)
6. Anderson, R., Böhme, R., Clayton, R., Moore, T.: Security economics and European policy. In: M.E. Johnson (ed.) Managing Information Risk and the Economics of Security, pp. 55–80. Springer, New York (2008)
7. Byres, E.J.: Network secures process control. Tech Magazine, Instrumentation Systems and Automation Society (1998)
8. Byres, E.J., Lowe, J.: The myths and facts behind cyber security risks for industrial Ccntrol systems. BCIT (2003)
9. CBC Digital Archives: The great northeastern blackout of 1965. http://archives.cbc.ca/economy_business/energy/topics/874/
10. Denning, D.: Information Warfare and Security. Addison-Wesley (1999)
11. Department of Homeland Security: Roadmap to secure control systems in the energy sector. Department of Energy (2008). http://www.controlsystemsroadmap.net/
12. Department of Homeland Security: Recommended practice for patch management of control systems (2008). http://csrp.inl.gov/Documents/PatchManagementRecommendedPractice_Final.pdf
13. Fink, R., Spencer, D., Wells, R.: Lessons learned from cyber security assessments of SCADA and energy management systems. US Department of Energy (2006)
14. Gutmann, P.: Auckland's power outage, or Auckland – your Y2K test site (1998). www.cs.auckland.ac.nz/~pgut001/misc/mercury.txt
15. Hoge, W.: Britain convicts 6 of plot to black out London. New York Times, 3 July (1997)
16. Lookabaugh, D., Sicker, T.: Security and lock-in. In: L.J. Camp and S. Lewis (eds.) Economics of Information Security, pp. 225–246. Kluwer Academic Publishers (2004)
17. Melton, R., Fletcher, T., Early, M.: System protection profile – industrial control systems. NIST (2004). www.isd.mel.nist.gov/projects/processcontrol/SPP-ICSv1.0.pdf
18. Meserve, J.: Sources – staged cyber attack reveals vulnerability in power grid. CNN, 26 Sep (2007). http://edition.cnn.com/2007/US/09/26/power.at.risk/index.html
19. Paller, A.: CIA confirms cyber attack caused Multi-city power outage. SANS Newsbites **10**(5) (2008)
20. PJM Media: Black Start Service Working Group – MRC Update (2009). www.pjm.com/Media/committees-groups/working-groups/bsswg/20090217/20090217-mrc-update-01-15-09.pdf

21. Drimer, S., Murdoch, S.J., Anderson, R.: Thinking inside the box: system-level failures of tamper proofing. In: IEEE Symposium on Security and Privacy, pp. 281–295. IEEE Computer Society (2008)
22. Safire, W.: The farewell dossier. New York Times, 2 February (2004)
23. Weiss, J.: Electric Power 2008 – is NERC CIP Compliance a Game? Control Global Community (2008)

Chapter 5
Internet Multi-Homing Problems: Explanations from Economics

Richard Clayton

Abstract Companies seeking to ensure that their Internet connection is resilient often purchase services from multiple providers. This leads them inexorably towards having their IP address range visible in the global routing table, increasing the resource usage of every Internet router. Since this is essentially 'free', yet impacts the cost and stability of every router in the world, this is a classic 'tragedy of the commons'. There is little prospect of change in the IPv4 world, but there is a chance to fix the problem as IPv6 is rolled out. Unfortunately, SHIM6, the engineering solution chosen to solve this issue in IPv6, will only be effective if universally adopted, and there are no short-term incentives to prefer SHIM6 over a duplication of the IPv4 arrangements. Incentives could be artificially introduced by requiring payment for adding multi-homed address space to the global routing table — a naïve estimate of the actual cost being $77 000 per routing prefix. However, it would be almost impossible to ensure the substantial revenues involved are correctly redistributed to those bearing the costs.

5.1 Introduction

The increasing reliance of all sizes of business on Internet connectivity is leading them to seek resilient methods of ensuring that they are never disconnected. Ironically, this resilience is creating instability within the Internet, and, for reasons that economists will instantly recognize, current attempts at solutions are failing to be effective.

The growth of email use in companies has been extraordinarily rapid. For example in the UK, a 1998 survey [19] found only a quarter of small companies using email (and in two thirds of them, only 10% of employees used email regularly). By 2002 a survey [13] of marketing and procurement managers in the auto/electrical

Computer Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0FD, UK.
e-mail: richard.clayton@cl.cam.ac.uk

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_5,

component manufacturers, financial services and telecommunications industries didn't even mention if any company wasn't using email — it was just assumed that within this industry sector they would. The 2002 survey was more concerned to show that email was now second in importance to the telephone for both buyers and suppliers. Usage has continued to grow, and access speeds have become faster, so that by 2006, an OFCOM survey of SME businesses found that 84% had an Internet connection, and only 20% of those were still using dialup.

Companies are now increasing their dependence on the Internet by migrating their telephone usage to VoIP (Voice over IP) services, so that their voice traffic shares the same link as their Internet traffic. Recent surveys, such as the 2008 annual OFCOM Communications Market report, show VoIP usage remaining very low with just 20% of users making one or more calls a month. However, this is mainly measuring Skype usage by individuals, whereas the companies being considered in this paper would purchase integrated telecoms products, for which there are few reliable statistics.

As companies discover that they cannot operate without a working Internet connection, they will insist upon resilience. The obvious solution, to purchase connectivity from more than one Internet Service Provider (ISP), turns out to be complicated, as will now be explained.

5.2 How Internet Routing Works

As is well understood, machines connected to the Internet have a unique 'IP address'. When machines communicate, routers inspect each of the packets they forward to pick out the destination IP address and send the packet over an appropriate link to a router that is, in some sense, 'closer' to where the packet is to be finally delivered.

Internet address space is allocated to ISPs in a hierarchical manner by the five Regional Internet Registries (RIRs), ARIN, RIPE, LACNIC, APNIC and AFRINIC. The ISPs are also allocated AS (Autonomous System) numbers by the RIRs, which are used to group together their allocations of address space for which they will have a consistent routing policy. The ISPs operate routers which communicate with their neighbors using BGP (the Border Gateway Protocol). These routers learn which 'routes' their neighbors are aware of, where a route consists of a 'route prefix' (the first n bits of a block of IP address space, along with the value of n) and an 'AS path' which indicates the AS's which must be traversed to reach the AS that owns the address block.

In the absence of any overriding local configuration, a router chooses which neighboring router to send a packet to on the basis of two rules: first it picks the 'most specific' route prefix (the one with largest value of n, representing the smallest enclosing address block). The router then picks the shortest AS path from amongst competing advertisements of that prefix. The reason for selecting the shortest path is the obvious one of getting packets to their destination as efficiently as possible.

The reason for the 'most specific' rule is to simplify route announcements; an ISP can announce a large address block such as a /16 (where the prefix length n is 16), without having to split this up into separate chunks if a subset of the address space, such as a single /19 (n = 19, one eighth the size), is to be routed differently.

For a multi-homed company to fully benefit from the resilience of having multiple connections to the global Internet, it must use a fixed set of IP addresses, and the traffic will then arrive over whichever path is shortest and still working. From the description above, it can be seen that for a customer to use the same set of IP addresses with two ISPs, it is necessary for this block of address space to be announced by both providers.

There isn't strictly any necessity for the customer to have their own AS, but this is generally seen as the 'clean' way to operate. It has the advantage to the customer that they can more easily change providers, it simplifies configuration for all concerned, and it permits remote systems to check some security properties of the announcement.

Therefore, in practice, for a customer to be multi-homed they will need to obtain an AS of their own; operate a BGP-speaking router (or ask a provider to run it for them); and announce their route prefix to their connectivity providers, so that it will become known to the rest of the world. Hence, an entirely local decision to arrange for resilience has, of necessity, a global impact because the route prefix will be recorded in the 'global routing table' that each and every router must construct to know where to send packets.

5.3 The 'Global Routing Table'

The size of the global routing table has been a matter of concern for many years. Routers need to keep the table in memory for instantaneous access; which has proved to be a problem for older router architectures where adding memory is expensive or even impossible past a certain limit. Furthermore, inter-router traffic grows along with the size of the table.

There is a specific concern about apparently unnecessary entries, where for example a provider splits some address space in two, and advertises two adjacent /19 blocks rather than a single /18. The CIDR report [6] tracks these occurrences, and at present the global routing table would reduce by 37% if all possible aggregations occurred.

Aggregation is of course impossible if address space is fragmented, e.g.: when a new allocation of address space to an ISP is not adjacent to their existing space. Fragmentation may also occur by choice, because the ISP wants to avoid congestion by splitting the traffic to different parts of their network over multiple ingress paths. Nonetheless a great deal of fragmentation is unnecessary and aggregation is often possible. Social pressure, exemplified by the weekly publication of the CIDR report, has helped to reduce the number of unnecessary announcements. The importance of this social pressure was remarked upon in a 2001 survey paper [10], where the

observation was made that there are visible dips in the upward trend immediately after IETF meetings where the issue of routing table size was discussed.

Growth of the routing table has usually been exponential [10], and the current trend is a growth of about 25% per annum, with the May 2009 size being just under 300 000 prefixes. The growth is caused by new allocations of IP address space (as new people connect to the Internet), traffic engineering schemes to balance the load and avoid congestion, and route prefixes that are only present to permit multi-homing.

A 2005 study by Meng et al. found that around 45% of prefixes were 'covered', viz: they were more specific prefixes for other routes; and they ascribed 44% of these to multi-homing; i.e. around 20% of the entire global routing table is present solely because of multi-homing [14]. Furthermore, Bu et. al found that the number of multi-homing prefixes (along with prefixes that were present for load balancing reasons) was growing faster than the routing table as a whole [7].

There has been a similar growth in AS number allocations, with about 31 000 currently in active use, and another 15 000 allocated but not yet in use on the public Internet [9]. Growth is presently a steady 5 000 or so per annum. Since AS numbers were originally 16-bit values, this would have meant exhaustion in 2011 or so, and so the BGP protocol has been re-engineered to permit the use of 32-bit AS numbers [20] and support for this will be universal by the beginning of 2010.

Further evidence of the role of multi-homing can be seen by examining the amount of address space advertised per AS. Since AS numbers are generally allocated in order (albeit they are passed to the RIRs in lumps which are then used up at different rates), the higher the AS number the more recently it has been issued. Additionally, most of today's ISPs have existed for many years (albeit seldom under the same name, or management).

Therefore we would expect ISPs to have low AS numbers and large amounts of address space, but higher AS numbers will have been allocated to multi-homed companies who use a small amount of address space. Examining a scatter plot of the address space announced by each AS (see Fig. 5.1) we see that our prediction is borne out, and most of the high AS numbers (past 20 000) have very small amounts of address space, whereas many of the low AS numbers (particularly below 5 000) have considerably more.

Besides the impact on the size of the global routing table, multi-homing companies share a further unfortunate characteristic in that they are more volatile. Meng et al. observed [14] that covered prefixes (i.e. the category into which multi-homed companies fall) were more likely to be announced and then later withdrawn. Each time a route prefix appears or disappears, then all of the world's routers have to recalculate their version of the global routing table, a resource intensive task. When many prefixes are announced or withdrawn over a short period of time it can be several minutes before the routers catch up with the changes and are routing packets normally again. Thus the existence of the extra routes is contributing to overall instability and adversely affecting 'availability' world-wide.

Economists will not find it hard to see parallels with other scenarios. Individual ISP customers choose whether or not to become multi-homed by considering a local

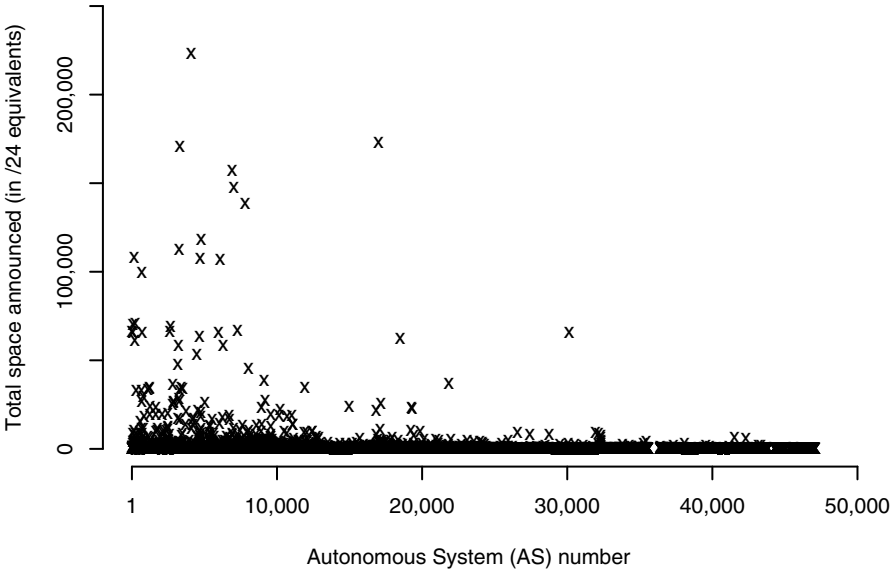


Fig. 5.1 Size of IPv4 address space announced (in /24 (256-address) equivalents) plotted against the AS number making the announcement.

cost/benefit analysis, rather than assessing the cumulative impact on the size of the global routing table, or the need to re-engineer the entire BGP infrastructure to cope with 32-bit AS numbers. This is essentially Hardin's 'Tragedy of the Commons' played out in a high-tech setting [8].

5.4 IPv6

Although it is too late within the IPv4 protocol to prevent local multi-homing decisions having global impact, one might hope that this mistake will not be repeated in IPv6. But the outlook is gloomy.

The problem facing IPv6 is well understood in the community. In August 2003, RFC3582 [1] "Goals for IPv6 Site-Multihoming Architectures" explained the need for multi-homing and set out some clear goals, including scalability ("A new IPv6 multihoming architecture should scale to accommodate orders of magnitude more multihomed sites without imposing unreasonable requirements on the routing system.") and limited cooperation ("A multihoming strategy may require cooperation between a site and its transit providers, but should not require cooperation (relating specifically to the multihomed site) directly between the transit providers"). The IPv4 multi-homing system was assessed against the RFC3582 considerations by Abley et al. in RFC4116 and found wanting [2].

Furthermore, Savola and Chown [17] provided some indications of the sort of scalability required, calculating (extremely simplistically) that if there were 1000 multi-homed firms per million people this would result in a million extra routing prefixes if the current multi-homing scheme was continued. They also drew attention to the risk that major network failures could result in bursts of 100 000 simultaneous BGP updates — a considerable workload.

Their paper went on to survey the new architectures being proposed in 2005. They distinguish:

- Host-centered proposals, where the hosts have multiple IP addresses, one for each link to the Internet. The hosts must arrange to communicate these addresses to the other end of connections, who then select which address to use.
- Modifications to the transport layer to allow dynamic changes to IP addresses within the TCP protocol (or the replacement of TCP with some other protocol such as SCTP). They did not believe there was much enthusiasm for this.
- Use of the 'Mobile IPv6' mechanisms to permit the link to the Internet to change. This posed some difficulty, not least because a key security mechanism of Mobile IPv6 is that when bindings change a check is made to ensure that this is agreed to by communicating with the old address — but if the link to the old address has just failed then this is impossible.
- Schemes that break the binding between identification and location. The Host Identity Protocol (HIP) is one such, using cryptographic hashes to link identifiers at the transport layer and address values, however this requires too many changes to be viable. Another, LIN6, has been patented and this has prevented serious consideration.
- Schemes that propose geographic allocation of IPv6 addresses. These fix the aggregation problem because most customers would multi-home with geographically close providers. However, the Internet isn't wired up in this way, and it is unclear how country level links, carrying significant volumes of traffic, would be funded.

As can be seen, the assessment made of these proposals was basically just testing their engineering elegance; with the addition of a small amount of commonsense thinking about how Internet peering actually works.

Around the same time, Lear documented the issues that ought to be considered in RFC4219 [12]. He set out 45 questions, all of which related to technical aspects of possible solutions. He failed to ask what the prospects were of getting a solution deployed in the real-world, perhaps because it was widely believed that there would be no actual choice about that.

The proposal that eventually emerged from amongst the various competing ideas to be taken forward was SHIM6, a host-oriented scheme.

5.4.1 SHIM6

In the SHIM6 design, connections are made using the TCP protocol in the normal way, but if more than a few packets are exchanged (and so the overhead appears to be worthwhile), the multi-homed host will tell the other end of the connection about any other IPv6 addresses on which it can be reached. If the connection subsequently fails, then the other end will use these fallback values, and tag the packets to indicate this has happened.

The higher stack levels will be unaware of the changed IPv6 address values because the receiver detects the tag and fixes up the packets to contain the original address, hiding the link failure. The 'fixing-up' layer is implemented as an add-on within the network stack's IP layer, hence the name, which is not an acronym, but is chosen because 'shim' is a common jargon word for modules that add functionality to a network stack layer.

The SHIM6 protocol is complex, not least because it must be secure against fraudulent announcements of IPv6 addresses that are not valid, and are not an appropriate way to make contact. The main description covers 124 pages [15], along with another 61 pages of related documents [4, 5]. Admittedly, some of the pages are filled with justifications for architectural choices and reasons why parts of the design are the way that they are, but it is still a significant undertaking to implement the protocol. For comparison, the size is two-thirds that of the description of the NFSv4 distributed file system protocol (RFC3530) which supports traditional file access while integrating support for file locking and the mount protocol, along with strong security (and its negotiation), compound operations, client caching, and internationalization [18].

This implementation complexity is compounded by the documents having remained at the 'work-in-progress' Internet-Draft stage right up until June 2009, when they finally became stable 'Standards Track' RFCs. This strongly suggests that SHIM6 will not be widely implemented and universally deployed in the near future, if at all.

5.4.2 The Lack of Incentives for SHIM6 Deployment

Unfortunately, the way that SHIM6 works means that if it is to provide any resilience, then both ends of a connection must be using it. Thus, for its benefits to be fully enjoyed by a multi-homed site, it must be universally deployed. Naturally, there is a clear incentive for the multi-homed site to upgrade their machines to use the new protocol. However, the incentive for others is entirely absent, which means that even if SHIM6 turns out to be straightforward to deploy, there is no obvious reason for people to bother.

Hence, especially in the short term, we must expect multi-homed IPv6 sites to use the same multi-homing scheme as they would have used in IPv4, viz: obtaining their own AS number, and adding their route prefix to the global routing table. Since

these sites will now no longer derive any special benefit from SHIM6 there will no longer be any incentive — even for them — to deploy it.

The ISPs are unlikely to be especially keen for their customers to deploy SHIM6. At present, ISPs can perform crude 'traffic management' on their customers by artificially extending the AS path as they relay customer BGP routing announcements. This has the effect of causing traffic to flow preferentially via other providers, and hence it can be a useful way of dealing with temporary congestion. However, if the customer is using SHIM6 then there is no customer specific announcement to tinker with. The effect of creating such an announcement will be to make it the 'most specific' route to the customer so that, no matter how long the AS path, all of the traffic will flow through the ISP and increase the congestion. Thus SHIM6 removes some traffic engineering 'control knobs' from ISPs, thereby reducing their incentive to recommend the protocol.

With no encouragement to be expected from ISPs, no advantages for early adopters, and the likelihood that those who might benefit from SHIM6 having to settle for another approach entirely, it is difficult, at the time of writing this paper, to see the protocol catching on.

5.4.3 Cooperating ISPs

Although, as discussed above, RFC3582 [1] ruled out solutions that require cooperation between transit providers, this could in fact offer a way forward.

In practice, multi-homed companies will be purchasing service from a small number of ISPs in their geographic region. These ISPs could cooperate by arranging that all of the multi-homed customers they shared with a particular competitor were placed within a single block of address space whose prefix was announced by both ISPs. When connectivity via one ISP fails, the other ISP (where there was no problem) would then announce a more specific route for the customer, so that all of the traffic flows through the working connection.

Whilst there were no connectivity problems this would markedly reduce the number of prefixes in the global routing table, and the extra routes added in the event of local failures would not be a huge burden. However, IP address space management would be far from simple — in regions where there were dozens of competing ISPs there would have to be hundreds of different blocks of shared address space.

So although this approach could conceivably be made to work, there would be considerable costs involved in arranging the necessary cooperation between the ISPs. In addition, the scheme would almost certainly require customers who changed suppliers to renumber to another block of IP address space. Since renumbering is of itself disruptive, this might suit the ISPs (because there would be a disincentive for customers to leave) but it must be presumed that the customers would not choose such an arrangement if others were on offer.

Hence although cooperation might be desirable, without creating some disincentives to the existing method of multi-homing, it is most unlikely to be adopted.

5.5 Discouraging Growth in the Global Routing Table

One way to prevent unjustified growth in the global routing table would be to be charge people for entries. Provided that the charge was correctly set, this could fairly recompense those whose resources are being consumed by companies choosing to become multi-homed in the current manner. In fact, there are existing mechanisms which could be used for this purpose, because adding a route is not quite as free as has been suggested so far.

The Regional Internet Registries (RIRs) currently fund their activities by charging members for their services. For example, RIPE NCC (the RIR for Europe, the Middle East and parts of Central Asia) splits their membership up by size, from 'large' though 'medium' to 'small', charging €5 500/annum to the large members, and €1 300/annum to small ones. The size is determined by a complex formula that assesses how many AS numbers and blocks of IP address space have been allocated, and how long ago this allocation was initially made.

Therefore, should a company wish to become multi-homed, they could join RIPE in their own right — which would cost them €2 300 in the first year and €1 300 thereafter. However, if they were to obtain their space via an existing member then that member might well pay nothing more by becoming a little 'bigger', but even if the new customer pushed them over a charging boundary, the amortized cost over all of their customers would only be a handful of Euro each.

So there is a small financial disincentive to creating new multi-homed sites. However, the actual worldwide cost of coping with the extra prefix is substantially more than a few thousands of Euros. We can estimate what this cost might be by calculating the total current cost of providing routing, and dividing this down by the 300 000 route prefixes currently in the global routing table. Unfortunately, this estimate can only be made very roughly, because of a lack of detailed numbers.

One rough and ready approach is to consider the topmost tier of network providers, those who do not have 'transit providers', but only mutual peering relationships. There are currently 13 such, each of which will have around 10 000 routers costing say $100K each (i.e. $13 billion of kit between them). The next tier down, which have complete meshes within regions, are about 10 times as many, albeit around 10 times smaller, but with cheaper hardware their routers cost them in total around $8 billion. Finally there are the stub systems, around 30,000 of these, but with just a handful of $30K routers each: for roughly another $2 billion.

Hence the total infrastructure cost can be estimated to be very roughly $23 billion. This is in line with estimates of yearly sales of $12.8 billion [11], given that routers need regular replacement as traffic (and the global routing table) grows. Dividing this down gives a cost per prefix of $77 000.

Of course, this is only one way of calculating the cost of adding a route prefix. The actual cost of any particular prefix is either zero (the general case where it makes no difference) or occasionally the cost of an entire new router (when an old one can no longer cope). Furthermore, new routers may be purchased anyway to handle greater amounts of traffic — and being newer designs they may cope with bigger routing tables as a matter of course.

Hence other calculations are certainly possible. But the real difficulty in trying to take this approach is not how much should be charged, but the lack of any obvious way to distribute this money to subsidize the people purchasing and running the routers. If the money is equally shared 'per router' then if the $77 000 figure is correct, by purchasing an AS and a cheap router you would actually get given money! If routers are not counted equally then money should flow to tier 1 providers from the multi-homed edge systems; but it would be extremely hard to prevent them 'gaming' the system by misrepresenting how many routers are actually needed and how much subsidy they should receive.

The conclusion must be that there doesn't seem to be any practical way of charging for routes at the present time; but the disparity between the straw man figure of $77 000 and the few thousand Euro that is the absolute maximum that would currently be paid, underlines the point that multi-homed customers are consuming expensive resources but are not having to pay anything like the full cost.

5.6 Related Work on the Economics of Protocols

Ozment and Schechter specifically looked at the issue of bootstrapping the adoption of Internet protocols, their focus being specifically on security protocols [16]. They developed a formal model, and considered strategies that might lead to protocol adoption.

Only a few of their strategies would work for SHIM6. "Global Mandate" would correspond to having some way of fining people who did not deploy the protocol, which would be unrealistic. "Partial Mandate" is inapplicable because there is no 'tipping point' after which deploying SHIM6 would be an obvious choice. "Bundling" is also inapplicable at present because SHIM6 does not give any other benefits — although if there was more commonality with the 'Mobile IPv6' protocols that might change. Their "Facilitating Sub-network Adoption" strategy might be viable if multi-homed companies were able to work with the subset of the whole Internet with whom they wanted to have reliable long duration connections; that is, they don't need the whole Internet to use SHIM6, just certain parts of it. "Coordination" also seems inapplicable, but "Subsidization" might well be the best way forward — those who stood to lose most from a ever growing IPv6 global routing table could invest in ensuring that SHIM6 was incorporated into standard network stacks, and hence became widely adopted.

The real problem is that SHIM6 may make engineering sense (albeit, given its complexity, that could be debated), but the economics of its deployment has hardly been considered within the IETF. In contrast, within the totally unrelated area of email spam control, economic arguments have come to be seem as absolutely key when evaluating proposals.

It is extremely common for new anti-spam solutions to be proposed which would only work if universally deployed, which have no benefits for early adopters, which assume that spam senders would not change their behavior, or that senders of le-

gitimate email would be delighted to pay extra for the privilege. Proposals with such failings are routinely dismissed by the anti-spam community and no substantial work put into experimenting with them.

This type of security economics analysis is widely used within forums such as the IETF Anti-Spam Research Group (ASRG). It is not presently described in any formal academic papers, but, as is the way of these things, has been quite beautifully encapsulated in the widely circulated "Why Your Anti-spam Solution Won't Work" [3] which, although written to amuse, is of immense practical use in summarizing what is wrong with a new proposal. Almost none of its points are technical. The emphasis is on economic, legal and philosophical objections — as well as the occasional medical issue, since imaginative new methods for killing spammers are seldom painful enough.

5.7 Conclusions

As uninterrupted access to the Internet becomes central to the day-to-day operation of companies, they are seeking ways to make that access more resilient. Purchasing connectivity from multiple ISPs gives resilience, but to fully realize the benefits when one of the connections fails, it is necessary for every router in the world to learn of the existence of their particular block of IP address space. The cost of this is out of all proportion to what is actually being paid by the company — a modern day 'Tragedy of the Commons'.

SHIM6, the engineering fix for this within the upcoming IPv6 protocol is complex, has only been finalized very recently, and offers no special benefits to early adopters. There is little reason to believe that it will be rapidly and universally deployed. This means that the current exponential growth of the global routing table in IPv4 is likely to be replicated in IPv6.

Security Economics has already begun to permeate the way in which we evaluate other protocols, such as anti-spam schemes. It is clearly well past time that proposals for new network layer protocols were considered in a similar manner. One way of achieving this would be for the IETF to require an 'Economics Considerations' section within all standards track RFC documents. Sections on 'IANA Considerations' and 'Security Considerations' are already mandatory.

Social pressure has had a significant effect on the growth of the global routing table so far. This may continue to be the most effective (and by far the cheapest and simplest) mechanism to rely upon. The way forward may be for multi-homing of small customers using global routing announcements to cease to be seen as a legitimate engineering solution.

It can only be a matter of time until a major ISP does a deal with a competitor to offer multi-homing to ten thousand of their biggest business customers, with a managed BGP-speaking router and a block of address space bundled into their offering. When that happens, they may agree to cooperate in announcing the address space as set out in Sect. 5.4.3 above. If not, and their initiative is popular enough in the

marketplace to grow the global routing table by 30% almost overnight; we may see a rapid change away from current laissez faire attitudes.

Acknowledgements The author benefited from many helpful discussions on this topic, especially with Ross Anderson, Jon Crowcroft, Chris Hall and Andrew Moore.

References

1. Abley, J., Black, B., Gill, V.: Goals for IPv6 site-multihoming architectures. IETF RFC 3582 (2003)
2. Abley, J., Lindqvist, K., Davies, E., Black, B., Gill, V.: IPv4 multihoming practices and limitations. IETF RFC 4116 (2005)
3. Anonymous: Why your anti-spam solution won't work (2004). `http://craphound.com/spamsolutions.txt`
4. Arkko, J., van Beijnum, I.: Failure detection and locator pair exploration protocol for IPv6 multihoming. IETF RFC 5534 (2009)
5. Bagnulo, M.: Hash-based addresses (HBA). IETF RFC 5535 (2009)
6. Bates, T., Smith, P., Huston, G.: CIDR report. `http://www.cidr-report.org`
7. Bu, T., Gao, L., Towsley, D.: On characterizing BGP routing table growth. Computer Networks **45**(1), 45–54 (2004)
8. Hardin, G.: The tragedy of the commons. Science **162**(3859), 1243–1248 (1968)
9. Huston, G.: The 16-bit AS number report. `http://www.potaroo.net/tools/asn16/`
10. Huston, G.: Analyzing the Internet BGP routing table. Internet Protocol Journal **4**(1), 2–15 (2001)
11. Infonetics Research: Great year for service provider router market, but 4Q showed signs of downturn. Press Release (2009). `http://www.infonetics.com/pr/2009/router-switch-market-highlights.asp`
12. Lear, E.: Things multihoming in IPv6 (MULTI6) developers should think about. IETF RFC 4219 (2005)
13. Leeka, S., Turnbulla, P., Naudé, P.: How is information technology affecting business relationships? Results from a UK survey. Industrial Marketing Management **32**(2), 119–126 (2003)
14. Meng, X., Xu, Z., Zhang, B., Huston, G., Lu, S., Zhang, L.: IPv4 address allocation and the BGP routing table evolution. SIGCOMM Computer Communication Review **35**(1), 71–80 (2005)
15. Nordmark, E., Bagnulo, M.: Shim6: Level 3 multihoming shim protocol for IPv6. IETF RFC 5533 (2009)
16. Ozment, A., Schechter, S.E.: Bootstrapping the adoption of Internet security protocols. In: Fifth Workshop on the Economics of Information Security WEIS2006 (2006)
17. Savola, P., Chown, T.: A survey of IPv6 site multihoming proposals. In: Proceedings of the 8th International Conference on Telecommunications (ConTEL 2005). IEEE, pp. 41–48 (2005)
18. Shepler, S., Callaghan, B., Robinson, D., Thurlow, R., Beame, C., Eisler, M., Noveck, D.: Network file system (NFS) version 4 protocol. IETF RFC 3530 (2003)
19. Sillince, J.A.A., Macdonald, S., Lefang, B., Frost, B.: Email adoption, diffusion, use and impact within small firms: A survey of UK companies. International Journal of Information Management **18**(4), 231–242 (1998)
20. Vohra, Q., Chen, E.: BGP support for four-octet AS number space. IETF RFC 4893 (2007)

Chapter 6
Modeling the Security Ecosystem - The Dynamics of (In)Security

Stefan Frei, Dominik Schatzmann, Bernhard Plattner, Brian Trammell

Abstract The security of information technology and computer networks is effected by a wide variety of actors and processes which together make up a security ecosystem; here we examine this ecosystem, consolidating many aspects of security that have hitherto been discussed only separately. First, we analyze the roles of the major actors within this ecosystem and the processes they participate in, and the the paths vulnerability data take through the ecosystem and the impact of each of these on security risk. Then, based on a quantitative examination of 27,000 vulnerabilities disclosed over the past decade and taken from publicly available data sources, we quantify the systematic gap between exploit and patch availability. We provide the first examination of the impact and the risks associated with this gap on the ecosystem as a whole. Our analysis provides a metric for the success of the "responsible disclosure" process. We measure the prevalence of the commercial markets for vulnerability information and highlight the role of security information providers (SIP), which function as the "free press" of the ecosystem.

6.1 Introduction

With the ongoing deployment of information technology in today's economy and society, comprehending the evolution of information security at large has become much more than the mere understanding of the underlying technologies. There is

Stefan Frei
Communication Systems Group, ETH Zurich, e-mail: frei@techzoom.net

Dominik Schatzmann
Communication Systems Group, ETH Zurich, e-mail: schatzmann@tik.ee.ethz.ch

Bernhard Plattner
Communication Systems Group, ETH Zurich, e-mail: plattner@tik.ee.ethz.ch

Brian Trammell
Hitachi Europe, ICTL Secure Systems Team, Zurich e-mail: trammell@tik.ee.ethz.ch

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_6,

a growing realization that security failures are caused as often by bad incentives as by bad design or neglected implementation: Insecurity often results from what economists call an *externality*, a side-effect of using information technology, like environmental pollution [2]. E.g. vulnerabilities in software impose costs on the whole society of users, while software vendors get all the profits. Whenever a new vulnerability is discovered, various parties with different and often conflicting motives and incentives become engaged in a complex way. These players and their interactions form what we call the *Security Ecosystem*. The security impact resulting from the interplay of the actors of the security ecosystem cannot be understood and managed unless we can better measure these risks. The goal of this paper is to develop metrics that help to obtain a better understanding of the state and the evolution of today's security environment from a global perspective. Our method to give insight into the dynamics and the prevalence of important processes of the security ecosystem is the analysis of the *Lifecycle of a Vulnerability*, based entirely on publicly available data from various sources. In the following we define the lifecycle of a vulnerability and introduce a model to describe the main players and their interactions in the security ecosystem. The sequence of events in the vulnerability lifecycle measures the main processes governing the security ecosystem. To support the understanding of these complex processes we revisit the key elements of the "disclosure debate", look at "vulnerability markets", and analyze the motivations of vendors and cyber-criminals. Finally we show how the security ecosystem can be described and analyzed quantitatively using statistical analysis of the vulnerability lifecycle.

6.2 Related Work

After years of providing more and more security features, a realization emerged that a pure technical point of view is not sufficient to understand the ever evolving security landscape [2]. According to [34], the *security ecosystem* describes the activities of creating, preventing, dealing with, and mitigating insecurity in the use of information technology. The economics of information security is *cross-disciplinary* as much as *interdisciplinary* according to Pfleeger [39]. Quantitative measurements of the security ecosystem typically focused on partial analysis of individual events. In "The new school of information security" Shostack and Stewart observe that until today there exist no aggregated long-term indicators or indexes to better understand how the security ecosystem functions [47]. Research on the economic consequences of cyber attacks has been dealing primarily with microanalysis of specific events, technologies or targeted organizations [39]. In 2004, Cavusoglu and Arora examine how a disclosure policy affects the time for a vendor to release a patch [5, 16]. Kannan and Telang study whether market-based mechanism for vulnerability disclosure lead to a better social outcome [22]. The lure of money is changing the computer security playing field, and we must reexamine our assumptions in the face of financially motivated attackers. In 2004 Thomas et al. highlight that fraud is likely to

be as prevalent in the online environment as in the conventional environment [51] and Maillart et al demonstrated in 2008 that the largest possible ID losses per event grow faster-than-linearly [27]. The convergence of criminals and technically savvy crackers is on the way [25].

6.3 Methodology

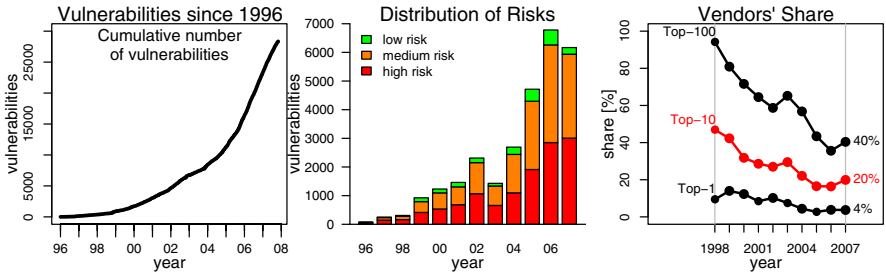


Fig. 6.1 Vulnerability disclosures 1996-2007 and share of the top-N vendors with the most vulnerabilities.

In this research, we analyze the state and the evolution of the security ecosystem over the last twelve years based on an empirical dataset of more than 27,000 vulnerabilities disclosed between 1996 and 2008. We examine the prevalence of different sequences of events in the vulnerability lifecycle for a large set of vulnerabilities, normalized to the time of vulnerability disclosure. Normalization with respect to the time of disclosure is an obvious approach as this is the first point in time the vulnerability becomes known to the public. To create a comprehensive vulnerability database we download, parse, and correlate the information of well over 200,000 individual security bulletins of various sources. Due to the inaccessibility, privacy or unavailability of data, only certain aspects of the security ecosystem can be measured from the outside. It is unlikely that cyber-criminals will ever share data about their operations, and software manufacturers are reluctant to publish data about their internal vulnerability handling processes. The data for this research is gathered exclusively from publicly available sources.

Phase 1 - Data Collection We do not attempt to take all possible information sources into consideration, rather than being exhaustive we choose a set of sources based on criteria such as independence, accessibility, and available history of information. Thus, we processed all security advisories from *US-CERT* [53], *SecurityFocus* [49], *IBM ISS X-Force* [19], *Secunia* [43], *Vupen* [15], *SecurityTracker* [44], *iDefense's* (VCP) [21], and *TippingPoints* (ZDI) [52]. For exploit information we analyzed *Milw0rm* [31], *Packetstorm* [1], *SecurityVulns* [45], and *Metasploit* [17]. Finally we imported the content of the National Vulnerability Database (NVD), the Open Source Vulnerability Database (OSVDB) [37], and the CVE database [33].

Phase 2 - Parsing We processed the data gathered in Phase 1 to extract the *date of publication*, all *CVE identifiers* and all *cross references (URLs)* to other security sources. From the NVD we derive the mapping of vulnerability to vendor/product name and risk rating (high, medium, low). This information is fed into our vulnerability database.

Phase 3 - Data Correlation In the database we correlate the raw data collected in the previous phases. CVE identifiers are used for the correlation of vulnerability information from different sources. To capture cases where the CVE identifier is missing in an advisory, we used cross references in NVD and CVE documents (where a CVE is always assigned by definition). The output of this step is a set of unique vulnerabilities identified by their CVE identifier and a set of related advisories from different sources providing the specific vulnerability lifecycle data.

Vulnerability Data Before we proceed with the analysis, we look at the total number of vulnerabilities in our database and their distribution among vendors and risk classes. In Fig. 6.1 left we plot the cumulative number of vulnerabilities disclosed since 1996 and in the center we plot the number of disclosures by year and risk rating. The information plotted is based on the content of our vulnerability database. Consistently, most vulnerabilities are classified as either "high" or "medium" risk, and up to 2006 we see a steady increase in the number of vulnerabilities disclosed per year. The distribution of these vulnerabilities among the affected vendors is depicted in Fig. 6.1 (right), and Fig. 6.2. Only a few vendors account for most vulnerabilities published in a given year and we observe a skewed distribution similar to a power law distribution. This fact is shown in Fig. 6.1 (right) where we plot the combined share of the *top-N* vendors (affected by vulnerabilities) per year since 1998 for $N \in \{1, 10, 100\}$. E.g. only $N = 10$ (or 0.04%) of the 2,491 vendors of vulnerable software in 2007 are responsible for 20% of the reported vulnerabilities in that year. Fig. 6.2 lists the names of the *top-10* vendors from 2002 to 2007. From this analysis we observe that most of the vulnerabilities published in any given year affect well known commercial and open-source software vendors. These vendors produce the majority of software products in daily use at home and within business. As a result most of the vulnerabilities disclosed are of relevance to the majority of users.

6.4 Vulnerability Lifecycle

Our method to give insight into the dynamics of the security ecosystem is the analysis of the vulnerability lifecycle shown in Fig. 6.3. The sequence of events in the vulnerability lifecycle is used to measure the main processes governing the security ecosystem. We first define what we consider to be a security vulnerability and introduce the events of the vulnerability lifecycle followed by the identification of specific risk exposure phases defined by the sequence of these events.

What is a Vulnerability? The lifecycle of a vulnerability cannot be modeled without a precise definition of the term *vulnerability*. However, defining vulnerabilities is

a delicate undertaking that depends significantly on the parties involved and their intent. For example, whether a specific software flaw is considered *a defect*, *a feature*, or *a vulnerability* differs whether you talk to a researcher, the vendor, or different users of the software. In the field of information security, many competing definitions of a vulnerability have been proposed [26, 38]. As we are mainly interested in accurately reflecting the processes of the security ecosystem, we delegate the decision on what counts as a vulnerability to the Common Vulnerabilities and Exposures (CVE) consortium [33]. CVE is a *de facto* industry standard that has achieved wide acceptance in the security industry, academia, and a number of government organizations since its launch in 1999. According to CVE, a vulnerability is a mistake in software that can be directly used by an attacker to gain access to a system or network [32]. For this research, we consider only vulnerabilities listed in the CVE database, thereby delegating the decision on what counts as a vulnerability to the CVE editorial board:

Definition 6.1. For this research, only a security issue with an assigned CVE identifier is considered a **vulnerability**.

This definition explicitly does not try to define technical properties of security issues, as we are interested in capturing the real-world impact of security issues in order to shed light on the processes of the security ecosystem. Given the high acceptance of the CVE process in academia and industry we assume that any security issue *of relevance* will eventually get a CVE number assigned.

2002	2003	2004	2005	2006	2007
Microsoft	Microsoft	Microsoft	Microsoft	Microsoft	Microsoft
Cisco	Sun	Gentoo	Apple	Apple	Apple
HP	Apple	Red Hat	Linux	Oracle	IBM
Sun	IBM	Apple	Mozilla	Mozilla	Oracle
Oracle	Red Hat	Linux	Sun	Linux	PHP
IBM	SGI	SuSE	IBM	IBM	Sun
SGI	HP	SGI	Oracle	Sun	Cisco
Apache	Apache	Mozilla	Red Hat	Cisco	Mozilla
FreeBSD	Cisco	Mandrake	Ethereal	Joomla	HP
Mozilla	Linux	Sun	SuSE	Novell	Linux

Fig. 6.2 List of the top-10 vendors by number of vulnerabilities in their products. Source: NVD

Vulnerability Lifecycle Events The lifecycle of a vulnerability $v \in V$ (with V denoting the set of vulnerabilities listed by CVE) can be divided into phases between distinctive events. Each phase reflects a specific state of the vulnerability and an associated risk exposure for the users of the software affected. To capture these phases we define the events *creation*, *discovery*, *exploit availability*, *disclosure*, *patch availability*, and *patch installation* for each vulnerability, as shown in Fig. 6.3. With some restrictions, the exact sequence of these events varies among individual vulnerabilities.

Time of creation (t_{creat}) Vulnerabilities are typically created by accident as the result of a coding mistake, often involving the mismanagement of memory. If a

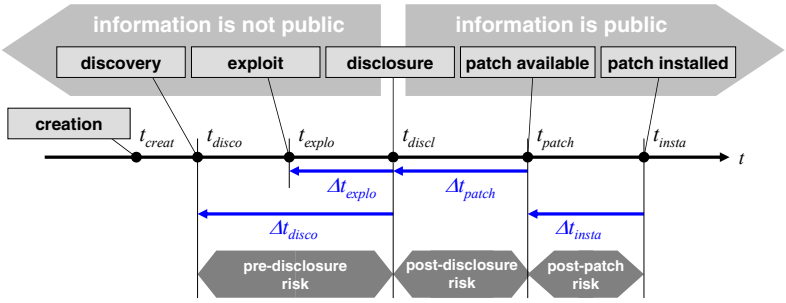


Fig. 6.3 The lifecycle of a vulnerability defined by distinctive events. The exact sequence of events varies between vulnerabilities.

vulnerability remains undetected in the code throughout the development and testing phases, chances are it will make it into generally available code that is then released [18]. In this research we consider only vulnerabilities discovered *after* the release of the software. The time of vulnerability creation is typically unknown by definition, however it may be determined in retrospect, after the discovery or disclosure of the vulnerability. If the creation of a vulnerability is malicious and thus intentional, discovery and creation time coincide [3]. In this paper we do not further investigate the time of vulnerability creation.

Time of discovery (t_{disco}) The *time of discovery* is the earliest time a software vulnerability is recognized to pose a security risk. Vulnerabilities do exist before they are discovered, but prior to the discovery of the vulnerability the underlying defect is not recognized to pose a security risk. Usually the time of discovery of a vulnerability is not publicly known until *after* its disclosure.

Time of exploit availability (t_{explo}) An exploit is a piece of software, a virus, a set of data, or sequence of commands that takes advantage of a vulnerability in order to cause unintended or unanticipated behavior to occur in software or an embedded device. Proof-of-concept code or exploits provided within security research and analysis tools are also deemed exploits[1]. Typically, it is a trivial exercise for criminals to turn such code into a working exploit. The *time of exploit* is the earliest time an exploit for a vulnerability is available.

Time of public disclosure (t_{discl}) The purpose of *disclosure* is to make security information available to the public in a standardized, understandable format. Disclosure is an important event in the security ecosystem. In the literature, definitions of *disclosure* range from ”made public to wider audience”, ”made public through forums or by vendor”, ”reported by CERT or Securityfocus”, or ”made public by anyone before vendor releases a patch” as in [3, 4, 35]. To normalize this set of definitions, we define the disclosure time as follows:

[1] E.g. *Metasploit*, a tool for developing and executing exploit code to aid in penetration testing and IDS signature development.

Definition 6.2. The **time of disclosure** $t_{discl}(v)$ of a vulnerability v is the first time a vulnerability is described on a channel where the *information disclosed* and the *information channel* publishing the vulnerability satisfy the following requirements:

1. *Free Access:* The disclosed vulnerability information is available to the public for free.
2. *Independence:* The vulnerability information is published by a widely accepted and independent source.
3. *Validation:* The vulnerability has undergone analysis by security experts such that risk rating information is included.

These requirements ensure the quality of vulnerability information threefold: From the security perspective only a free and public disclosure of vulnerability information can ensure that all interested, affected, or concerned parties get the relevant security information (*free access*). *Independence* is a prerequisite for unbiased and complete information, while the *validation* requirement builds confidence in the quality of the information delivered. The mere discussion of a potential flaw in a mailing list or vague information from a vendor therefore does not qualify. We call viable sources of vulnerability information *Security Information Providers (SIP)*, which we discuss in detail in Section 6.5. Furthermore, only an information source not dependent on a vendor or government is unbiased and ensures a fair dissemination of security critical information[2]. This implies the use of several sources to determine the time of disclosure, as many of the organizations that publish security information are associated with vendors or governments. In combination, these three requirements ensure that the disclosure date reflects the first time when trusted, widely understandable information about a new vulnerability is publicly available to everyone concerned. Correlation using CVE identifiers allows to handle dissimilar publication dates from diverse sources: The publication date of the first SIP (as listed in the Appendix) reporting a given vulnerability is used as the disclosure date t_{discl} for a vulnerability.

Time of patch availability (t_{patch}) The *time of patch availability* is the earliest time that the vendor releases a patch that provides protection against the exploitation of the vulnerability. Unfortunately, software vendors typically cannot make security patches available instantly after the discovery of new vulnerabilities or exploits. While some vendors publish patches as soon as these are available, others publish patches on a predefined schedule to ease the planning of patch installation (e.g. monthly or quarterly scheduled release of new patches). We analyze the patch release performance of various software vendors in detail in Section 6.6. In many cases a patch may be available before public disclosure (e.g. the DNS vulnerabilities of 2008 and service pack roll-ups for new operating systems).Fixes and patches offered by third parties are not considered as a patch, we deem the vendor as the only authoritative source to provide patches for its software. The complexity of patches varies from simple configuration fixes to extensive changes in the foundation of

[2] In the following of this paper we use the term *vendor* to name the manufacturer of the software for *commercial products*, *freeware*, and *open-source software* alike

the software. Other security mechanisms such as signatures for intrusion prevention systems or anti-virus tools are not considered as patches.

Time of patch installation (t_{insta}) Software users can only benefit from the correction of a vulnerability after a patch is installed on their systems. The processes leading from patch availability to patch installation vary considerably among different kinds of software users. Hence, the time to patch installation is not a specific point in time for a vulnerability, it can only be given as a distribution for a specific population of users (e.g. corporate or home users).

6.4.1 Risk Exposure Times

Between the discovery of a vulnerability and its elimination through the installation of a patch, a system is potentially at risk. This exposure period can be separated into three phases: the "pre-disclosure", the "post-disclosure" and the "post-patch" phase as shown in Fig. 6.3. We analyze the relation and evolution of these periods to distinguish and understand important processes in the security ecosystem.

Pre-disclosure phase (Δt_{disco}) During the time from discovery to disclosure Δt_{disco}, only a unknown group is aware of the vulnerability. This group could be anyone from lone hackers to cyber-criminals likely to misuse their knowledge. On the other hand, this group could also consist of researchers and vendors working together to provide a patch for the identified vulnerability. We call the risk exposure arising from this period as "pre-disclosure" risk because the vulnerability is known to have a security impact whereas the public has no access to this knowledge.

$$\Delta t_{disco}(v) = t_{disco}(v) - t_{discl}(v) \tag{6.1}$$

Post-disclosure phase (Δt_{patch}) During the time from disclosure to patch availability Δt_{patch} the user of the software waits for the vendor to release a patch. We call the risk exposure arising from this period the "post-disclosure" risk because the public is aware of this risk but has not yet received remediation from the software vendor/originator. However, users of the vulnerable software can assess their individual risk and implement a workaround based on the information provided with the disclosure of the vulnerability.

$$\Delta t_{patch}(v) = t_{patch}(v) - t_{discl}(v) \tag{6.2}$$

Post-patch phase (Δt_{insta}) The time from patch availability to patch installation Δt_{insta} is called the "post-patch" risk. The duration of this period is typically under direct control of the user of the affected software or embedded device. Typically, business and private users face different challenges to timely patch installation. Installing a patch or changing security-relevant configuration settings on a mission-critical business system is a non-trivial task for a typical enterprise. Further, we found considerable delays of patch installation timing of end-users' Web browsers

in [10, 12, 13], mostly attributed to the degree of automation available for patch installation. Note that an ever-increasing number of embedded control devices are deployed in support of our networked society, many of which cannot be patched by their users.

$$\Delta t_{insta}(v) = t_{insta}(v) - t_{patch}(v) \tag{6.3}$$

Exogenous vs. Endogenous We designate "pre-disclosure" and "post-disclosure" phases as *exogenous*, since the operator of the vulnerable system cannot exert direct influence on the length of these periods. The length of these phases can only be **influenced on a macro perspective** through the interplay of the processes in the security ecosystem, as shown in Fig. 6.4 and discussed in Section 6.5. Likewise, the nature of the "post-patch" phase is *endogenous* as the operator of the system determines the time when the patch is installed.

6.5 The Security Ecosystem

In this section we introduce and discuss the *major players* and *main processes* in security ecosystem followed by a review of the "disclosure debate" which is central to understand these processes and the incentives. In the last decade, the number of players and their roles and interactions within the security ecosystem have evolved considerably. A variety of legislative and social issues directly influence the processes of vulnerability research, detection, publication, and response. Vendors, developers, customers, cyber-criminals, and the security community have divergent perspectives on the impact of vulnerabilities. The processes and interactions between these actors are driven by the continuous discovery of new vulnerabilities and the subsequent constant need of the public (the software users) for security information and patches. In Fig. 6.4 we model the main processes in the security ecosystem, starting with the discovery of a new vulnerability on top and the public disclosure of vulnerability information at the bottom. The flow of vulnerability information from the discoverer to the public can take several paths, each describing a different process with implications for the resulting risk exposure. The boxes *Discovery*, *Exploit*, *Patch*, and *Disclosure* in our model identify important events in the security ecosystem that can be related to events in the vulnerability lifecycle as introduced in Section 6.4. Examination of the exact sequence of vulnerability lifecycle events for a large sample of vulnerabilities allows us to identify the prevalence of particular processes and the dynamics of the security ecosystem.

6.5.1 Major Players

We start the discussion of the security ecosystem model with the introduction of its major players, namely the *discoverer*, commercial-, and underground *vulnerability*

markets, *cyber-criminals*, the software *vendors*, *security information providers*, and the *public*.

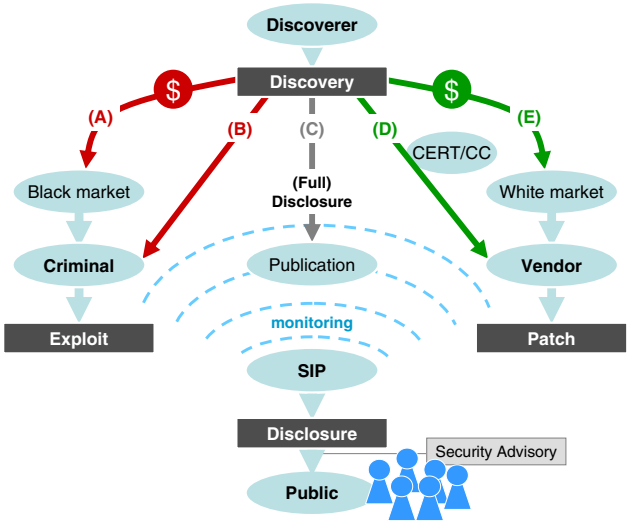


Fig. 6.4 Main processes of the security ecosystem and relation to vulnerability lifecycle events.

6.5.1.1 Discoverer

The *discoverer* of a vulnerability is an individual or organization (e.g. the vendor, independent researcher, cyber-criminal, government agency) that discovers a new vulnerability. How the discoverer proceeds with this information depends on his intrinsic motivation and the incentives offered by the environment. Whatever the choice, it ultimately impacts the risk exposure time of the public. There are many different motivations to direct the discoverer of a vulnerability:

- malicious intent for profit, Path (A) or Path (B)
- altruism, Path (C), Path (D)
- recognition or fame, Path (C)
- forcing unresponsive vendors to address a vulnerability, Path (C), Path (D), or Path (E)
- curiosity and the challenge of vulnerability analysis, Path (C)
- political motives, Path (A) or Path (B)

It is important to note that the number of third party software vulnerability discoveries has not declined over the last decade, as shown in Fig. 6.1, despite massive efforts of the security and software industry.

6.5.1.2 Vulnerability Markets

Information about security vulnerabilities can be a valuable asset. Vulnerability information is traded in both the underground "black market" and the commercial services "white market". While a market for vulnerabilities has developed, vulnerability commercialization remains a hotly-debated topic tied to the concept of vulnerability disclosure. Responsible disclosure fails to satisfy security researchers who expect to be financially compensated, while reporting vulnerabilities to the vendor with the expectation of compensation might be viewed as extortion [11]. On the other hand, cyber-criminals not bound by legal or ethical considerations are willing to invest considerable amounts in suitable vulnerability information. H. D. Moore[3] claims that he was offered between $60k and $120k for critical vulnerabilities in Microsoft products as reported in [6, 28, 30]. Researchers that intend to sell a vulnerability face the possibility that the same vulnerability is discovered, patched, and published independently. This threat of independent discovery pressures them to sell the vulnerability to the quickest bidder instead of the highest one. Factors that determine the market price of a vulnerability are:

- *Exclusivity of information*. This is the key factor, once the vulnerability becomes widely known the value of the information tends to zero.
- *Security impact*. The higher the security impact, the higher the value of the vulnerability.
- *Product popularity*. A vulnerability affecting a popular product has a higher value.

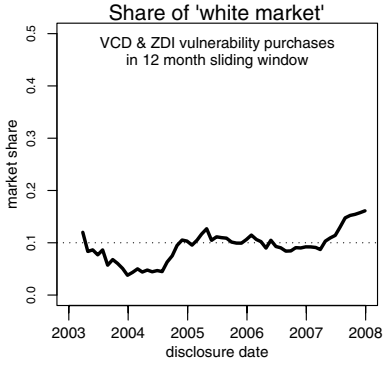
Black Market The black market has developed around the illegal or malicious use of the vulnerability information. Sellers are not driven by ethical considerations. The black-market trade is not openly advertised, and the information is used in a way that generally increases the risk exposure of the public. The lack of trust between sellers and buyers potentially exposes both parties to fraud. Due to the nature of the market accurate information on the number and type of trades completed is not systematically available. Only specific investigations provide some insight into the inner workings, e.g. by Symantec's "Underground Economy Report" [50].

White Market Players in the white market offer commercial services and openly advertise their vulnerability handling policies. Demonstrating and ensuring that buyers and sellers don't have malicious intent is a major challenge for the players in the commercial vulnerability market. White market buyers typically purchase vulnerability information to protect their customers before the vulnerability becomes public knowledge, and inform the vendor of the affected software. Such buyers advertise their ethics and ask security researchers to accept lower compensation with the promise that the information will be used for benevolent purposes [28]. Incentives for the buyers are:

[3] H. D. Moore founded the Metasploit project, an open platform for developing and testing exploit code.

- Publicity generated from disclosing newsworthy vulnerabilities drives interest in their commercial services.
- Providers of intrusion detection and prevention systems include additional protection, which customers might perceive as an advantage.
- They provide the information as a paid service to their customers.

Today, the two primary players in the commercial vulnerability market are *iDefense*, which started their vulnerability contributor program (VCP) in 2003, and *TippingPoint*, with their zero-day initiative (ZDI) started in 2005.TippingPoint's ZDI receives an average of about 40 new vulnerabilities per month, and buys about one out of 10. Vulnerability prices are not disclosed but ZDI runs a "frequent-flyer" style program that can pay out bonuses as high as $20k to top researchers. Together, VCP and ZDI published 793 vulnerabilities affecting 192 different vendors since their start in March 2003 to December 2007. In the same period a total of 8,111 vulnerabilities were published for the same group of 192 vendors, including the 793 bought by VCP and ZDI. We normalize the number of "white market" vulnerabilities with



	Vendor	Total CVEs	Market CVEs	Market Share
1	Microsoft	866	65	7.5%
2	Apple	534	40	7.5%
3	Computer Associates	94	37	39.4%
4	IBM	356	32	9.0%
5	Novell	127	23	18.1%
6	Symantec	127	16	12.6%
7	Gentoo	117	15	12.8%
8	Sophos	26	15	57.7%
9	Ipswitch	57	14	24.6%
10	Sun	332	14	4.2%

Fig. 6.5 Share of commercial vulnerability purchase programs in 12 month moving window (left). Top-10 vendors for which the "white market" brought vulnerabilities from 2003 to 2007 (right)

respect to the total number of vulnerabilities disclosed for the group of affected vendors in the same period to estimate the prevalence of the "white market", Path (E). Using a 12 month sliding window approach, we calculate the share of the "white market" within the group of vendors for which VCP and ZDI already bought vulnerabilities, shown in Fig. 5(a). We observe an almost constant share of about 10% of these commercial programs since the end of 2004 and a rise to over 15% starting in 2007. In Table 5(b) we list the top 10 vendors for which the "white market" bought vulerabilities. We find that the share of vulnerabilities bought varies considerably between vendors, e.g. 4.2% of Sun's and 57.7% of Sophos vulnerabilities follwed Path (E). These numbers shed a first light to what extent "white markets" contribute to the vulnerability ecosystem. Fig. 5(a) shows the prevalence of Path (E), which at the same time provides a *minimum estimate* of the number of vulnerabil-

ities *not* discovered by the vendors themselfs. For example, between March 2003 and December 2007 in average 7.5% of the vulnerabilities affecting Microsoft and Apple were processed by either VCD or ZDI, while other vendors achieved higher shares.

6.5.1.3 Criminal

Any individual or organization misusing vulnerability information for its own profit regardless of motivation is denoted as *criminal* in the model of Fig. 6.4. This can be anyone from an individual hacker to cyber-criminals or government agencies. In this context *misuse* stands for any operation on the targeted system that the user of the system neither approved nor is aware of. Criminals develop or buy exploit material in order to make use of a vulnerability, and typically install malicious software to spy on the user, launch further attacks, and build botnets. Security vulnerabilities in widely used software prove to be a formidable instrument in the hands of cyber-criminals to either enable or expand their business.

6.5.1.4 Vendor

The vendor is the originator of the software affected by a vulnerability. We use the term vendor for commercial products, freeware, and open-source software alike. It is up to the vendor to produce and release a patch once he becomes aware of a vulnerability in his software. In Section 6.6 we measure the zero-day patch share as a metric to measure the performance of vendors' patching and security communication processes.

6.5.1.5 Security Information Provider (SIP)

In the face of a rapidly evolving and hostile environment, businesses and private users alike are in constant need of accurate and validated security information to assess their risk exposure and to protect their systems. However, for the majority of businesses and users it is infeasible and prohibitively costly to monitor, understand and validate all the possible primary information sources in order to extract the security information relevant for them. Several private and government organizations specialize in collecting and publishing security information. Some of these organizations run security research labs, sell security tools, or provide paid security and consulting services. These organizations efficiently monitor the primary sources of security information, validate the content found, and publish their findings as *security advisories* which describe vulnerabilities in a standardized format. These organizations have an important role in the security ecosystem and we denominate them *Security Information Providers (SIP)*. This monitoring of the (in)security environment by SIPs is depicted by dashed curves in Fig. 6.4. Through SIP services, the

public has systematic access to independent, validated, timely, and understandable security information. The availability of trusted security information from SIPs has an important impact on the *behavior* and *incentives* on the actors in the security ecosystem. The combined effect of the efforts of SIPs is a major pillar building the incentives for the actors in the security ecosystem [48]: Collectivity, the role of security information providers in the security ecosystem is comparable to the role of the **free and independent press** in an open society: Issues addressed by them can hardly be ignored, hidden or downplayed.

6.5.1.6 Public

All users, individuals, or organizations, that use software affected by a vulnerability comprise the public. These users typically are in need of accurate and validated security information to assess their risk and to protect their systems until a patch is released by the vendor.

6.5.2 Processes of the Security Ecosystem

Whether ethical or mischievous parties first get information about a new vulnerability impacts the risk exposure of software users. After the discoverer finds a new vulnerability we distinguish five principal paths, denoted Path (A) to Path (E), to proceed as depicted by solid arrows in Fig. 6.4.

6.5.2.1 Path (A) and Path (B)

Cyber-criminals discover security vulnerabilities through their own research or by purchasing the needed information from *black markets* for vulnerabilities [40, 54], represented by Path (A) and Path (B) respectively. For a vulnerability following Path (A) or Path (B) we typically observe the following sequence of events:

$$Discovery \rightarrow Exploit \rightarrow Disclosure \rightarrow Patch \tag{6.4}$$

$$t_{disco}(v) < t_{explo}(v) < t_{discl}(v) < t_{patch}(v) \tag{6.5}$$

The time of vulnerability discovery is likely not available as criminals typically do not share information about their operations. The vendor can only start developing a patch after the vulnerability is actively exploited. Cyber-criminals basically have two options to take advantage of an exploit: *stealthy exploitation* or *full scale exploitation*:

In case of *stealthy exploitation*, cyber-criminals use the exploit only against a few, carefully-selected, high-profile targets, and actively avoid detection to extend the time they can profit from the unknown vulnerability [36]. This phenomenon

is known as "customized malware". However, as described in Section 6.5.3, it is not possible to keep security information secret forever. Eventually, information about the vulnerability spreads to a wider audience. When the disclosure of the vulnerability or the release of a patch is imminent, cyber-criminals may maximize their return of investment by moving on to *full scale exploitation* of the exploit.

In case of *full scale exploitation*, cyber-criminals release the exploit against a large population of targets to take advantage of a greater proportion of unprotected systems. With the higher percentage of compromised systems comes the greater risk of exposure of their activity, which eventually exposes the vulnerability to detection and subsequent disclosure. SIPs and other organizations monitor the (in)security scene, exploit archives, and research malicious activity:

- Anti-virus vendors or providers of managed security services (MSS) capture a sample of the exploit for analysis.
- Hoenypots and honeynets capture a sample of the exploit for analysis [24]
- Vendors capture a sample of the exploit through their error reporting mechanisms [29] (usually if the exploit crashes on certain configurations).

These activities lead to the timely disclosure of the underlying vulnerability. Thus, Path (A) and Path (B) favor the malicious use of vulnerability information resulting in an increase of security impact and exposure to risk for users: a decrease of social welfare given the ubiquitous use of computer and communication technologies in our society.

6.5.2.2 Path (C)

The discoverer publishes information about the vulnerability on a suitable channel (e.g. in a security conference or on a security mailing list[4]). Following Path (C), the vulnerability information is available to all interested parties at the same time: the criminals, the vendor, and the public. SIPs monitoring the security landscape spot this information and report it in a new security advisory. However, usually writing an exploit based on vulnerability information is less complex and faster than writing and releasing a patch. In the extreme case of *full disclosure*, the discoverer includes proof-of-concept code and exploit material. A discoverer following Path (C) is typically not financially motivated. He either decides to publish the vulnerability firsthand, or he does so because the vendor was not responsive. We discuss these options in Section 6.5.3. For a vulnerability following Path (C) we typically observe the following sequence of events:

$$Discovery \rightarrow Disclosure \rightarrow Exploit \rightarrow Patch \tag{6.6}$$

$$t_{disco}(v) < t_{discl}(v) < t_{explo}(v) < t_{patch}(v) \tag{6.7}$$

[4] FullDisclosure and BugTraq are two well known security mailing lists

6.5.2.3 Path (D) and Path (E)

The discoverer reports the vulnerability either directly to the vendor, Path (D), or through a commercial vulnerability market, Path (E). In case the vulnerability affects several vendors the discoverer can do so using the services of a CERT/CC[5]. The discoverer and the vendor then typically follow the responsible disclosure process described in Section 6.5.3: the vulnerability information is kept secret until the vendor has a patch ready for release. If the vendor is not responsive or uncooperative, the discoverer might fail over to Path (C). When the patch is ready, the discoverer publishes his advisory at the same time as the vendor releases the patch. Criminals can only start with the development of an exploit after a patch is available. For a vulnerability following Path (D) or Path (E) we typically observe the following sequence of events:

$$Discovery \rightarrow \left\{ \begin{matrix} Disclosure \\ Patch \end{matrix} \right\} \rightarrow Exploit \tag{6.8}$$

$$t_{disco}(v) < t_{discl}(v) = t_{patch}(v) < t_{explo}(v) \tag{6.9}$$

Path (E) is an option for a financially motivated discoverer who does not want to sell the vulnerability in the underground where misuse is very likely. The prevalence of commercial vulnerability markets is shown in Fig. 5(a). Path (D) and Path (E) are more favorable for public risk exposure, as the vendor gets the information about the vulnerability before mischievous parties do. On the other hand, cyber-criminals have also refined their ability to analyze vulnerability information from vulnerability disclosures and reverse engineering of patches. Recent research demonstrated the potential of automated exploit generation based on a patch [9]. Cyber-criminals quickly create exploits upon the availability of such information.

6.5.3 The Disclosure Debate

Appreciation of vulnerability disclosure concepts and the accompanying incentives of the players involved is a prerequisite to understand the processes of the security ecosystem. The disclosure debate discusses the question of how to handle information about security vulnerabilities in order to minimize the security impact for the society:

- On the one hand, public disclosure of security information enables informed consumer choice and inspires vendors to be truthful about flaws, repair vulnerabilities and build more secure products [11]. This is the *security through transparency* stance of Kerckhoff [23].
- On the other hand, vulnerability information can give attackers (not sophisticated enough to identify a vulnerability on their own) the very information they

[5] CERT Coordination Center

need to exploit a security hole in a computer or system and cause harm. This is the *security through obscurity* stance[6].

The process of *responsible disclosure* evolved as a middle way between the opposing stances found in the disclosure debate. It has evolved and become a accepted way to handle security information [35].

Full disclosure is a security philosophy that holds that the details of security vulnerabilities should be available to everyone in a timely fashion. Before the systematic publication of software vulnerabilities, vendors typically would not bother to spend the time and money to fix vulnerabilities, believing in the security of secrecy [7, 11, 25, 41, 46]. Public disclosure or the threat of disclosure give vendors a strong incentive to fix the problem quickly. It is inevitable that cyber-criminals get the information alike with the public disclosure. This disadvantage is more than compensated by providing benign users the information needed to defend their systems as there is no way to assure that cyber-criminals do not already possess the same vulnerability information.

Responsible Disclosure Process The key insight from the disclosure debate is that secrecy mainly prevents people from assessing their own risks, which contributes to a false sense of security [42]. The process of *responsible disclosure* evolved as a middle course between the extremes of *full disclosure* and *security through obscurity*: The researcher discloses full information only to the vendor, expecting that the vendor will start the process to develop a patch, as in Path (D) or Path (E). In return, the vendor is expected to expeditiously issue a patch and give credit to the researcher for his discovery. The vendor is well incentivized to collaborate, as the discoverer can revert to *full disclosure* Path (C) if the vendor becomes unresponsive or the vulnerability is reported through other channels. In the last phase the discoverer will coordinate the publication of his advisory with the vendor's publication of the vulnerability information and the patch. An increasing number of vendors and security organizations adopted some form of *responsible disclosure* over the last decade [7, 8, 20].

6.6 The Dynamics of (In)Security

In this section, we focus on the evolution of the dynamics between security *(availability of patches)* and insecurity *(availability of exploits)*, based on the vulnerability lifecycle normalized to the time of disclosure. The intimate relation between the vulnerability lifecycle events and the processes in the security ecosystem are depicted in Fig. 6.4. The availability of an exploit poses a security threat, whereas the availability of a patch neutralizes this threat if the patch gets installed on the vulnerable system. Assuming that both the exploit and the patch work as intended by the respective originator, the resulting security risk for software users will depend

[6] also often referred to as *bug secrecy*

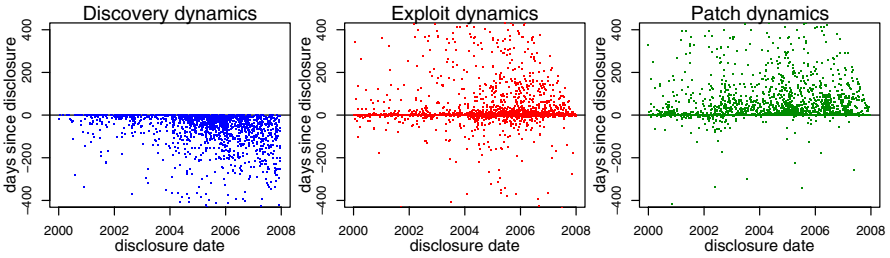


Fig. 6.6 Scatter plot of time of vulnerability discovery (left), exploit availability (center), and patch availability (right) by disclosure date.

strongly on the timing or dynamics of the availability of these. We measure the current state and identify global trends. For all vulnerabilities we know the time of the vulnerability disclosure $t_{discl}(v)$ taken from the fastest SIPs reporting this CVE with a resolution of one calendar day. Fig. 6.7 shows the number of vulnerabilities for which we found the time of discovery $|V_{disco}|$, time of exploit availability $|V_{explo}|$, and the time of patch availability $|V_{patch}|$ for every year from 2000 to 2007. The absolute number of vulnerabilities disclosed in a given year (100%) is visibile in Fig. 6.1. In the following of this section we individually discuss the dynamics of vulnerability *discovery*, *exploit availability*, and *patch availability* and describe the data sources used to build V_{disco}, V_{explo}, and V_{patch}. We examine the vulnerability

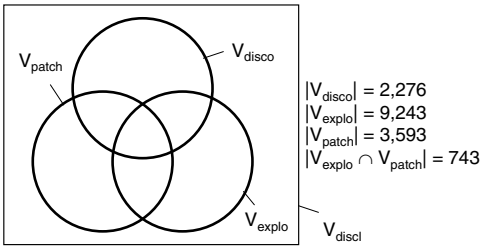


Fig. 6.7 Number of observed events within all vulnerabilities disclosed from 2000 to 2007.

lifecycle by looking at how the time of the events $\alpha \in E = \{disco, explo, patch\}$ relate to the respective disclosure time $t_{discl}(v)$ of the vulnerability. For all vulnerabilities from 2000 to 2007 and each type of event, we present a scatter plot, the associated distribution function, and yearly summaries to evaluate the evolution and identify trends. Normalization of the vulnerability lifecycle events with respect to the disclosure time is key to evaluate the aggregated dynamics of thousands of vulnerabilities. We build Δt_{disco}, Δt_{explo}, and Δt_{patch} as follows:

$$\Delta t_\alpha(v) = t_\alpha(v) - t_{discl}(v) \qquad \alpha \in E,\ v \in V_\alpha \tag{6.10}$$

Essentially $\Delta t_\alpha(v)$ represents the number of days event $\alpha \in E$ happened *before* or *after* the disclosure of vulnerability v:

$$sgn(\Delta t_\alpha(v)) = \begin{cases} -1 & \alpha \text{ occurs } \textit{before} \text{ disclosure} \\ 0 & \alpha \text{ occurs } \textit{at} \text{ disclosure} \\ 1 & \alpha \text{ occurs } \textit{after} \text{ disclosure} \end{cases}$$

Δt_{disco} is an estimator of the "pre-disclosure" risk and Δt_{patch} is an estimator of the "post-disclosure" risk period as introduced in Section 6.4.1.

Scatter plots We first use scatter plots of Δt_α to visualize the distribution and the evolution of events $\alpha \in E$ over the last eight years. In the scatter plots of Fig. 9.4 each point $P_\alpha(v)$ of event α is built according to

$$P_\alpha(v) \rightarrow (x,y) \quad \begin{cases} x = t_{discl}(v) \\ y = \Delta t_\alpha(v) \end{cases} \quad \alpha \in E,\ v \in V_\alpha \tag{6.11}$$

In all scatter plots, the x-axis is the calendar day of the disclosure of vulnerability v. The y-axis represents the time difference of event α to the disclosure of vulnerability v.

Distribution function To further analyze the dynamics, we plot and discuss the cumulated distribution $\mathscr{P}_{\leq}(X \leq x)$ of the same data used to generate the scatter plots. The $ecdf_\alpha(x)$ of event $\alpha \in E$ is

$$\begin{aligned} \mathscr{P}_{\leq}(X \leq x) &= ecdf_\alpha(x) \\ &= \left| \{ v \in V_\alpha \mid \Delta t_\alpha(v) \leq x \} \right| \end{aligned} \tag{6.12}$$

In Fig. 6.8, Fig. 6.9, and Fig. 6.10 we plot the $ecdf_\alpha(x)$ for discovery, exploit, and patch availability for the range of $x = \pm 400$ days around disclosure. These plots give insight in to the aggregated dynamics of the vulnerability lifecycle.

6.6.1 Discovery Dynamics

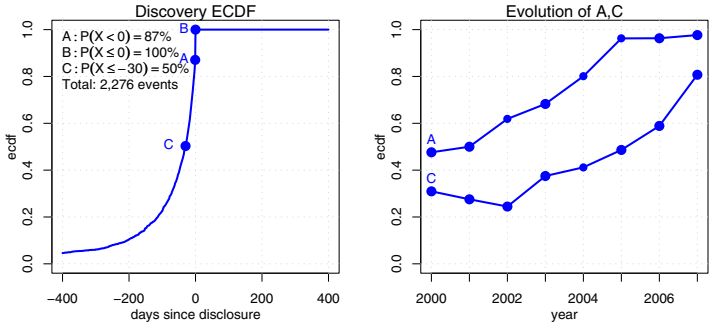


Fig. 6.8 Empirical cumulated distribution of the *discovery time* (left) and yearly evolution of selected points in the ecdf (right).

Usually the time of discovery of a vulnerability is not publicly known until *after* its disclosure. Indeed, for many vulnerabilities the time of discovery will never be known or reported to the public, depending on the motives of the discoverer. Cyber-criminals - and most software vendors - won't provide information about their vulnerability discoveries to the public. However, there are a few sources from which we can derive the time of vulnerability discovery. One source is the Open Source Vulnerability Database (OSVDB); the security bulletins of commercial vulnerability markets are another source. When *iDefense* or *TippingPoint* buy a vulnerability, they record the time of purchase or the time at which they notified the vendor of the affected software. Upon public release, this date can be retrieved from the disclosure timeline of the security advisory. Using this methodology we determined the time of discovery $t_{disco}(v)$ for a subset $V_{disco} \subset V$ of all vulnerabilities. Further, as the disclosure of a vulnerability implies its discovery we can state

$$t_{disco}(v) \leq t_{discl}(v) \qquad \forall\, v \in V_{disco} \tag{6.13}$$

Using Eq. 6.1 we can calculate $\Delta t_{disco}(v)$, a minimum estimator for the "pre-disclosure" risk. The true "pre-disclosure" risk period is always longer than what we can estimate based on publicly available data. In Fig. 6.8, the values for $x < 0$ show the distribution of the "pre-disclosure" risk from 2000 to 2007. For $x \geq 0$ $\mathscr{P}_{\leq}(X \leq x)$ equals 1 as disclosure implies discovery (Eq. 6.13). In Fig. 6.8 we plot the values for (A) $\mathscr{P}_{\leq}(X < 0)$ and (C) $\mathscr{P}_{\leq}(X < -30)$ for each year. The rise of (A) since 2000 points out that over time we observe more events with $t_{disco} < t_{discl}$ compared to $t_{disco} \leq t_{discl}$. The course of line (C) $\mathscr{P}_{\leq}(X < -30)$ shows that since 2000 more than 24% of the vulnerabilities were known to insiders more than 30 days before disclosure. In 2007 this share rose to 80% of the vulnerabilities. The course of line (C) is a minimum estimator of the "pre-disclosure" risk, of which one part is desirable - as it partially measures the success of the responsible disclosure process. However, for most vulnerabilities (mostly the ones discovered and abused by cyber-criminals) we never learn the discovery date. E.g. we only know the discovery date for 12% percent of the vulnerabilities patched in the last 5 years. We therefore consider our measurement of the "pre-disclosure" risk as a minimum estimator for the amount of time any privileged party has access to security critical information. This clearly shows the potential of the abuse of vulnerability information, especially as we have no data on vulnerability discoveries made by cyber-criminals or traded on the "black market". We conclude that vulnerabilities are systematically known to insiders (good and bad) well before the public learns about it.

6.6.2 Exploit Availability Dynamics

From the public exploit archives listed in Section 13.4 we can find the time of exploit availability for a subset $V_{explo} \subset V$ of all vulnerabilities. These exploit archives report the date when the exploit was published. The actual number of exploits avail-

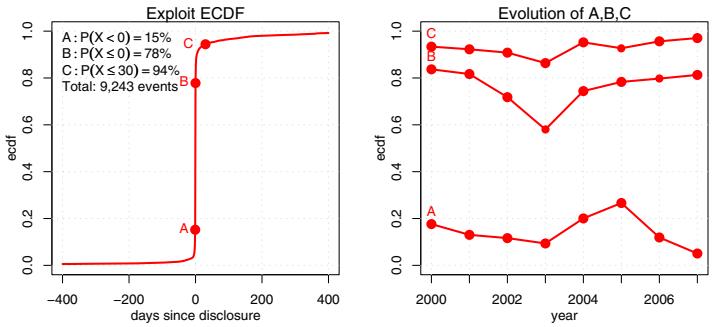


Fig. 6.9 Empirical cumulated distribution of the *exploit availability time* (left), yearly evolution of selected points in the ecdf (right).

able on these exploit archives is larger than $|V_{explo}|$ as we exclude exploits that cannot be correlated to a given CVE. Cyber-criminals use their exploit material for profit and have no incentive to publish their material on public exploit archives. Eventually, some of the exploits used exclusively by cyber-criminals make their way into exploit archives (as an exploit, proof of concept, test for patch). However, these postings are delayed. On the other hand, cyber-criminals monitor exploit archives and quickly enhance their repository of malware, should they find material previously unknown to them. As a result, we can only estimate the extent of yet undisclosed exploit information available to cyber-criminals at any time. V_{explo}, based on the content of public exploit archives, is therefore a *minimum estimate* for the true number of exploits available to cyber-criminals at a any given date. The time of exploit availability is $t_{explo}(v)$ with $v \in V_{explo} \subset V$. The scatter plot in Fig. 9.4 (center) shows the distribution of these exploits from 2000 to 2007. We observe that exploits are available both *before* and *after* the disclosure of the vulnerability, with an increasing density of exploit availability close to the disclosure day as of 2004. The plot of the cumulated distribution $\mathscr{P}_{\leq}(X \leq x)$ of Fig. 6.9 (left) quantifies the high dynamics of exploit availability close to the vulnerability disclosure. The sudden rise of $\mathscr{P}_{\leq}(X \leq x)$ from 15% before disclosure to 78% at disclosure from 2000 to 2007 quantifies the so called zero-day exploit phenomena [25]. A zero-day exploit is an exploit that takes advantage of a vulnerability at or before the day the vulnerability is disclosed. In other words, the vendor and the public have zero days to prepare for the security breach. The plot on Fig. 6.9 (right) shows that the zero-day exploit availability is above 70% for the last eight years with the only exception of 58% in 2003. Several mechanisms lead to the very high exploit availability at the time of disclosure. The combined effect of prior vulnerability knowledge and rapid analysis of disclosed vulnerability information (as discussed in Section 6.5.2.1) is readily seen by the increased activity at the disclosure day, and measured with a zero-day exploit availability of close to 80% since 2003. We cannot distinguish these mechanisms due to the limited scope and resolution (one calendar day) of publicly available information. Further, exploit availability reaches 94% 30 days after disclosure. Cyber-criminals systematically take advantage of users failing to

install patches quickly, or not having the latest patches installed. We analyzed and measured Internet users' discipline of patching their Web browsers in [12, 13].

6.6.3 Patch Availability Dynamics

A vendor typically reports the date when a new patch is released together with the patch bulletin or security advisory. To measure the dynamics of patch releases we download, parse, and correlate patch release bulletins of the seven vendors *Adobe, Apache, Apple, Microsoft, Mozilla Foundation, Oracle,* and *RedHat*. We chose these vendors to cover major players of the industry and with respect to the distribution of vulnerabilities among vendors as of Fig. 6.2. Using the release date posted in these vendor bulletins we determine the time of patch availability $t_{patch}(v)$ for a subset of vulnerabilities $V_{patch} \subset V$. Fig. 6.7 shows the number of vulnerabilities for which we have patch information available through the analysis of these seven vendors. The scatter plot in Fig. 9.4 (right) shows the distribution of the availability of these patches from 2000 to 2007. We observe that patches are mostly available

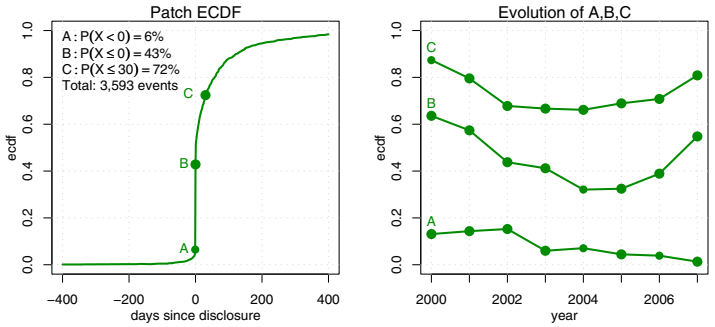


Fig. 6.10 Empirical cumulated distribution of the *patch availability time* (left), yearly evolution of selected points in the ecdf (right).

at or *after* the disclosure of the vulnerability. The plot of the cumulated distribution $\mathscr{P}_{\leq}(X \leq x)$ of Fig. 6.10 (left) quantifies the dynamics of patch availability close to vulnerability disclosure. Essentially, Δt_{patch} reveals the performance of the software industry in providing patches, a measure of the "post-disclosure" risk introduced in Section 6.4.1 and estimator of Path (D) and Path (E). Patch availability 30 days before the time of disclosure is at 2%. There are only few vulnerabilities found for which a patch already exists before the disclosure. The sudden rise of $\mathscr{P}_{\leq}(X \leq x)$ from 6% one day before disclosure to 43% at disclosure from 2000 to 2007 quantifies what we call the *zero-day patch* phenomena. The fraction of zero-day patches can be interpreted as a measure of the *responsible disclosure process*, implying Path (D) or Path (E) in our security ecosystem model. Before a patch is ready for publication the vendor needs time to analyze the vulnerability, develop,

test, document, and finally release the patch. Typically, a vendor is unable to release a patch within twenty-four hours of vulnerability discovery. Thus, to achieve a zero-day patch the vendor needs early notification of the vulnerability, typically through the responsible disclosure process Path (D), which includes contributions by the white market Path (E). The rise of $\mathscr{P}_{\leq}(X \leq x)$ for $x > 0$ measures how fast vendors react to vulnerability disclosures. Patch availability increases from 46% at disclosure to 72% at 30 days after the disclosure (equalling 28% unpached vulnerabilities 30 days after disclosure). This is a low number compared to the exploit availability of 94% 30 days after disclosure. Further, 13% of the vulnerabilites are still unpatched 90 days after the disclosure.

To determine how the risk of a vulnerability affects the patch performance we separately analyze the data for the three risk classes “high”, “medium”, and “low”. The results indicate that patch performance of “low” risk vulnerabilities consistently lags behind the performance of “high” and “medium” risk vulnerabilities, especially after disclosure. At disclosure we measure $\mathscr{P}_{\leq}(X \leq 0)$ to be 45%, 43% and 34% for “high”, “medium”, and “low” risk vulnerabilities repsectively. After disclosure we measure $\mathscr{P}_{\leq}(X \leq 30)$ to be 77%, 72% and 56% for “high”, “medium”, and “low” risk vulnerabilities repsectively. From these observations, we assume that the risk class of a vulnerability marginally effects the patch release performance in the sense that patches for “high” and “medium” risk vulnerabilities are prioritized against patches for “low” risk vulnerabilities. If the technological complexity of a fix to vulnerability were the dominant parameter to determine patch performance, then our measurements would lead to the conclusion that “low” risk vulnerabilities are generally more complex to fix than “high” or “medium” risk vulnerabilities, which we consider unlikely. We rather assume that work flow processes and priorization (and with it incentives) are at least as important as technical complexity to determine patch performance. Note that the discovery of a vulnerability by the vendor itself is also considered as responsible disclosure. An appropriately motivated employee discovering a vulnerability could also choose to offer this information to cyber-criminals instead. The share of zero-day patches indicates the sum of vulnerability discoveries by the vendor and vulnerabilities reported to the vendor through the “responsible disclosure” process. Applying these results to our model of the processes in the security ecosystem, Fig. 6.4, we conclude that between 6% and 43% of the vulnerabilites of the analyzed vendors followed the process Path (D) or Path (E). A detailed analysis of Microsoft and Apples zero-day patch performance is published in [14].

6.6.4 (In)security Dynamics

6.6.4.1 The Gap of Insecurity

An interesting aspect of our analysis is the direct comparison of the exploit and patch availability distributions and their trends over the last five years. For this we analyze

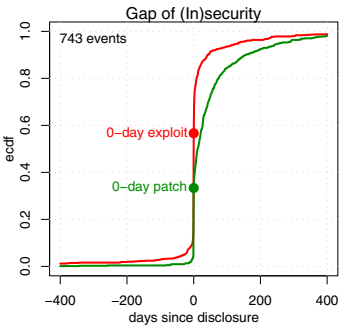


Fig. 6.11 Direct comparison of patch availability vs. exploit availability.

the cumulated distribution of $\Delta t_{patch}(v)$ for all vulnerabilities $v \in V_{patch}$ together with the cumulated distribution of $\Delta t_{explo}(v)$ for all $v \in V_{explo}$. Through vendor Web sites we have systematic access to *all patches* published by a given vendor and V_{patch} contains *all patches* published by our seven vendors. However, not all exploits are made available on public exploit archives, as explained in Section 6.6.2, so the distribution of $\Delta t_{explo}(v)$ is a *lower estimate* of the exploit availability. True exploit availability is always faster. Fig. 6.11 shows that exploit-availability continuously

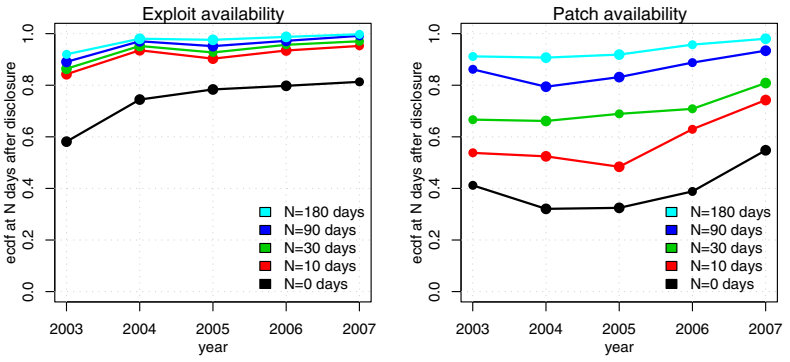


Fig. 6.12 Evolution of exploit availability and patch availability at $N \in \{0, 10, 30, 90, 180\}$ days after disclosure.

exceeds patch-availability for the full range ± 400 days around the day of disclosure. Exploit availability also consistently exceeds patch availability in every single year since 2000. This gap, which quantifies the difference between exploit- and patch-availability, is an indicator of the risk exposure and its development over time. This systematic gap also stresses the importance for the availability of independent and timely security information, the role of SIPs explained in Section 6.5.1.5. In Fig. 6.12 we plot distinct points at 0, 10, 30, 90 and 180 days of the cdf of Δt_{explo} and Δt_{patch} to visualize their evolution over time. Generally, both exploit and patch availability were increased over the last five years. With the exception of 2005, exploit availability increased steadily since 2003, and we observe a greater rise closer to the disclosure day. Exploit availability 30 days after disclosure continuously exceeds

90% since 2004. We observe high exploit dynamics within 10 days of disclosure; thereafter exploit availability rises only very slowly. We attribute this observation to the following causes:

- Exploits already known to cyber-criminals *before* public disclosure of the vulnerability.
- Increased capability to generate exploits either through reverse-engineering of patches or based on disclosed vulnerability information.
- Automated attack tools for Web application vulnerabilities that can actually *discover and exploit* a vulnerability. It is only afterward that the consultant/user of the tool realizes that the vulnerability exists - and then informs them that they need to fix it.

We cannot distinguish these causes based on our data, so we measure the aggregate effect. Note again that our data is a minimum estimate of the true availability of exploits. On the other hand, also patch availability increases almost steadily over the last years, although starting from a lower level than exploit availability. Closer to the disclosure, patch availability first dipped around 2005 and then caught up in the last three years. Again, patch availability is always lower than exploit availability at any day. Patch availability 90 days after disclosure does not surpass exploit availability 10 days after disclosure. We attribute patch availability performance to two different processes:

Patch release at zero-day: The release of a patch at the same day as the public disclosure of the vulnerability implies the vendor had early notification of the vulnerability ("responsible disclosure"), Path (D) or Path (E). A vendor is typically not able to analyze vulnerability information, then develop, test, and release a patch in less than a day. However, whether a vendor receives early notification from vulnerability discoverers is only partially under control of the vendor. This is to a high degree an exogenous factor that the vendor can only control in the long term, by establishing a trust relationship with the security community.

Patch release after disclosure: The time needed to release a patch upon knowing the vulnerability is under control of the vendor, a endogenous factor. Here we measure what a vendor *can do*, and what he is *willing to do* given technological complexity to fix the software, and economic incentives or constraints.

We believe that a good relationship with the security community can provide a higher share of early notifications of vulnerabilities which benefits a vendor in the following ways:

- Within responsible disclosure the vendor has more control of the time available to develop and release a patch than under the pressure of an already published vulnerability. This will typically result in a more efficient allocation and use of available resources of the vendor.
- A higher share of zero-day patches will be perceived as a better service to the customer.

Further, the systematic gap between patch and exploit availability underlines the role and importance of SIPs. During these periods, software users are exposed to risk of

exploit without already having received remediation from the vendor. It is during this time that security information on the threats is most important. The observed trend toward increased patch availability *at* and *after* the public disclosure indicates that the processes involved to release patches (technological, economic, incentives) have not yet reached saturation. A detailed analysis of Microsoft and Apples patch release performance since 2002 was published in [14]. Continued measurements using the methodologies presented in this chapter should be able to identify the limits of such processes at macroscopic scale.

Limitations The presented analysis is a first attempt at making the processes in the vulnerability ecosystem measurable. As there exists no systematic access to data on cyber-criminals operations, such an analysis comes with limitations. The *zero-day patch* share implies Path (D) or Path (E), however without excluding prior discovery through cyber-criminals. While we measured the extent of the *zero-day exploit* phenomena, the one day resolution of our data does not allow to distinguish between exploits that were derived from patches from exploits available before disclosure. Given the skewed distribution of vulnerabilities per vendor, the analysis must be viewed in the context of the specific vendors measured.

6.7 Conclusion

We introduced a model of the security ecosystem to capture its major players and processes. This is the first model of the security ecosystem that consolidates hitherto separately discussed aspects of the security processes. On the basis of the model we analyzed and discussed the roles and incentives of the players involved, backed with empirical data of more than 27,000 vulnerabilities. We enumerated the options of vulnerability discoverers, and visualized the security impact of their choices. For the first time we estimated the success of the "responsible disclosure process" backed with measurements, using the zero-day patch share as a metric. Our measurement revealed that commercial vulnerability markets cannot be neglected; on average they handle between 10% and 15% of the vulnerabilities of major software vendors. We found that exploit availability has consistently exceeded patch availability since 2000. This systematic gap between the availability of exploits and patches highlights the rapid dynamics around the day of vulnerability disclosure and the all-important role of *security information providers (SIP)* within the security ecosystem. The complexity and delay of installing patches paired with the fact that we can only provide an minimum estimate for exploit availability stresses the need for third party protection *and* timely availability of security information to the public. Our measurement methods are based entirely on publicly available information and provide a useful tool to measure the state of the security ecosystem and its evolution over time.

References

1. Packetstorm Security. http://packetstormsecurity.org
2. Anderson, R., Moore, T.: The Economics of Information Security. Science **314**(5799), 610–613 (2006). http://dx.doi.org/10.1126/science.1130992
3. Arbaugh, W.A., Fithen, W.L., McHugh, J.: Windows of vulnerability: A case study analysis. Computer **33**(12), 52–59 (2000). DOI http://doi.ieeecomputersociety.org/10.1109/2.889093
4. Arora, A., Krishnan, R., Nandkumar, A., Telang, R., Yang, Y.: Impact of vulnerability disclosure and patch availability – an empirical analysis. In: R. Anderson (ed.) Workshop on the Economics of Information Security (WEIS). Cambridge, UK (2004)
5. Arora, A., Telang, R., Xu, H.: Optimal policy for software vulnerability disclosure. In: Workshop on the Economics of Information Security (WEIS) (2004)
6. Boehme, R.: Vulnerability markets. what is the economic value of a zero-day exploit? In: Private Investigations (Proc. of 22nd Chaos Communication Congress). CCC (2005). DOI http://doi.acm.org/10.1145/1162666.1162671
7. Chambers, J.T., Thompson, J.W.: Niac vulnerability disclosure framework. Department of Homeland Security DHS (2004)
8. Christey, S., Wysopal, C.: Responsible vulnerability disclosure process (2002). http://tools.ietf.org/html/draft-christey-wysopal-vuln-disclosure-00
9. David, B., Pongsin, P., Dawn, S., Jiang, Z.: Automatic patch-based exploit generation is possible. In: IEEE Security and Privacy, 2008, pp. 143–157 (2008)
10. Duebendorfer, T., Frei, S.: Why Silent Updates Boost Security. Tech. Rep. 302, TIK, ETH Zurich (2009). http://www.techzoom.net/silent-updates
11. Electronic Frontier Foundation EFF: Coders' Rights Project Vulnerability Reporting FAQ
12. Frei, S., Dubendorfer, T., Ollmann, G., May, M.: Understanding the web browser threat. Tech. Rep. 288, ETH Zurich (2008). http://www.techzoom.net/papers
13. Frei, S., Duebendorfer, T., Plattner, B.: Firefox (In)Security Update Dynamics Exposed. Computer Communication Review **39**(1) (2009)
14. Frei, S., Tellenbach, B., Plattner, B.: 0-day patch - exposing vendors (in)security performance. BlackHat Europe (2008). http://www.techzoom.net/papers
15. FrSIRT: French Security Incident Response Team. http://www.frsirt.com
16. Hasan Cavusoglu, H.C., Raghunathan, S.: Emerging issues in responsible vulnerability disclosure. In: WITS (2004)
17. H.D. Moore: The Metasploit Project. http://www.metasploit.com
18. IBM Internet Security Systems: The Lifecycle of a Vulnerability. www.iss.net/documents/whitepapers/ISS_Vulnerability_Lifecycle_Whitepaper.pdf (2005)
19. IBM Internet Security Systems - X-Force: X-Force Advisory. http://www.iss.net
20. IBM Internet Security Systems - X-Force: Responsible vulnerability disclosure process (2004). http://documents.iss.net/literature/vulnerability_guidelines.pdf
21. iDefense: Vulnerability Contributor Program. Http://labs.idefense.com/vcp
22. Kannan, K., Telang, R.: An economic analysis of market for software vulnerabilities. In: Workshop on the Economics of Information Security (WEIS) (2004)
23. Kerckhoffs, A.: La cryptographie militaire. Journal des sciences militaires **IX**, 5–83 (1883)
24. Leita, C., Dacier, M., Wicherski, G.: SGNET: a distributed infrastructure to handle zero-day exploits. Tech. Rep. EURECOM+2164, Institut Eurecom, France (2007)
25. Levy, E.: Approaching zero. IEEE Security and Privacy **2**(4), 65–66 (2004). DOI http://doi.ieeecomputersociety.org/10.1109/MSP.2004.33
26. Lindner, F.F.: Software security is software reliability. Commun. ACM **49**(6), 57–61 (2006). DOI http://doi.acm.org/10.1145/1132469.1132502
27. Maillart, T., Sornette, D.: Heavy-tailed distribution of cyber-risks (2008). URL http://www.citebase.org/abstract?id=oai:arXiv.org:0803.2256
28. McKinney, D.: Vulnerability bazaar. IEEE Security and Privacy **5**(6), 69–73 (2007). DOI http://doi.ieeecomputersociety.org/10.1109/MSP.2007.180
29. Microsoft: Windows Error Reporting. Http://technet.microsoft.com/en-us/library/bb490841.aspx

30. Miller, C.: The legitimate vulnerability market: Inside the secretive world of 0-day exploit sales. In: Workshop on the Economics of Information Security (WEIS) (2007)
31. Milw0rm: Milw0rm Exploit Archive. http://www.milw0rm.com
32. MITRE : CVE Vulnerability Terminology 3. http://cve.mitre.org/about/terminology.html
33. MITRE: Common Vulnerabilities and Exposures (CVE). http://cve.mitre.org
34. Oborne, M.W.: The Security Economy. OECD, Paris : (2004). ISBN 92-64-10772-X
35. OISA Organization for Internet Safety: Guidelines for Security Vulnerability Reporting and Response. http://www.oisafety.org/guidelines/
36. Ollmann, G.: The evolution of commercial malware development kits and colour-by-numbers custom malware. Computer Fraud & Security **2008**(9), 4 – 7 (2008). http://dx.doi.org/10.1016/S1361-3723(08)70135-0
37. OSVDB: Open Source Vulnerability Database. Http://www.osvdb.org
38. Ozment, A.: Improving vulnerability discovery models. In: QoP '07: Proceedings of the 2007 ACM workshop on Quality of protection, pp. 6–11. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1314257.1314261
39. Pfleeger, S.L., Rue, R., Horwitz, J., Balakrishnan, A.: Investing in cyber security: The path to good practice. The RAND Journal **Vol 19, No. 1** (2006)
40. Radianti, J., Gonzalez, J.J.: Understanding hidden information security threats: The vulnerability black market. Hawaii International Conference on System Sciences **0**, 156c (2007). DOI http://doi.ieeecomputersociety.org/10.1109/HICSS.2007.583
41. Schneier, B.: Locks and Full Disclosure. IEEE Security and Privacy **01**(2), 88 (2003)
42. Schneier, B.: The nonsecurity of secrecy. Commun. ACM **47**(10), 120 (2004)
43. Secunia: Vulnerability Intelligence Provider. http://www.secunia.com
44. SecurityTracker: SecurityTracker. http://www.SecurityTracker.com
45. Securityvulns: Computer Security Vulnerabilities. http://securityvulns.com/
46. Shepherd, S.A.: Vulnerability Disclosure. SANS InfoSec Reading Room (2003)
47. Shostack, A., Stewart, A.: The new school of information security. Addison-Wesley (2008)
48. Stefan Frei and Martin May: Putting private and government CERT's to the test. In: 20th Annual FIRST Conference, June 22-27, 2008, Vancouver, Canada (2008)
49. Symantec: SecurityFocus. http://www.securityfocus.com/vulnerabilities
50. Symantec: Report on the Underground Economy (2008)
51. Thomas, B., Clergue, J., Schaad, A., Dacier, M.: A comparison of conventional and online fraud. In: CRIS'04, 2nd Int. Conf. on Critical Infrastructures, Oct 25-27, 2004 - Grenoble
52. TippingPoint: Zero day initiative (zdi). http://www.zerodayinitiative.com/
53. US-CERT: US-CERT. http://www.us-cert.gov/aboutus.html
54. Whipp, M.: Black market thrives on vulnerability trading. PCpro (2006). http://www.pcpro.co.uk/news/84523

Chapter 7
Modeling the Economic Incentives of DDoS Attacks: Femtocell Case Study *

Vicente Segura, Javier Lahuerta

Abstract Many of the Internet security incidents are caused by agents which act moved by economic incentives. When that is the case, it is possible to model attacker's incentives by applying economics principles and, if we can collect appropriate data, we can use the model to have a better understanding of the risk imposed by these threats. This paper presents a simple model that represents the economic incentives for launching DDoS attacks against a specific telecommunications service. In addition, some data has been collected in order to quantify some of the variables of the model. Finally, some simulations have been performed to have a better knowledge of the risk of suffering this kind of attacks and propose solutions to mitigate it.

7.1 Introduction

Risk analysis and management methodologies provide procedures and techniques for identifying and estimating security risks, identifying possible countermeasures and estimating how they reduce risks. Since now, these methodologies have proved to be useful for the systematic identification of risks. However, their usefulness for risks quantification is a very polemic topic. As Bruce Schneier says [12], one of the main reasons can be the scarcity of available data to estimate the variables in which the risk calculation models are based.

Most of the time, collecting suitable data in order to make reliable risk estimations is quite difficult or has unacceptable costs. Sometimes, risk analysts face these difficulties using qualitative scales for risk estimation. Although it does not provide

Vicente Segura, Javier Lahuerta
Department of Network and Services Security, Telefonica I+D, e-mail: vsg@tid.es

* This material is based upon work supported by the SEGUR@ Project, funded by the Centre for the Development of Industrial Technology (CDTI) of the Spanish Ministry of Science and Innovation.

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_7, © Springer Science+Business Media, LLC 2010

either a clear knowledge of risks magnitudes or a return-of-investment analysis of countermeasures, it enables the prioritization of risks. Anyway, when using qualitative or quantitative measures, we need procedures or techniques to estimate risk factors in a systematic and objective way. Similar results should be reached by different analysts when analyzing the same scenario.

This paper describes the details and results of a work in which we have modeled the economic incentives behind DDoS attacks against a specific telecommunications service. We think this is an example of systematic procedure that can be used to estimate risk factors when there is an economic motivation. It seems that it is what happens in an increasing percentage of cases due to the rising specialization of cybercrime [13]. In this situation, attacker's behavior is rational and even predictable. Therefore, it is possible to model the conditions that influence the attacker's behavior and, if we can collect data about our particular scenario we can estimate some risk factors, such as the probability of being impacted by some threats.

This paper focuses on a specific service which some European mobile operators will start to provide soon and which is already available in USA and in some Asian countries [4]. This service extends and improves the mobile coverage inside the home of its customers by means of a device called femtocell that links to the operator core network through a broadband line, typically an ADSL line. Femtocells are connected to the operator core network through a device called security gateway. The service architecture has a radical difference with respect to traditional architecture of mobile operator networks. They have been typically quite isolated from the Internet, but now the security gateway is the linking point between femtocells and the operator core network and it must be accessible from external networks that, in some of the proposed deployments, include the Internet.

The rest of the paper is organized as follows. Section 2 focuses on describing some of the previous works developed by different researchers that have some relation with this paper. Section 3 develops the economic model. Section 4 applies the model to the telecommunications service and it is composed of three subsections: data collection, analysis of collected data and use of the model to assess the economic incentives under different assumptions. Finally, section 5 concludes this paper.

7.2 Background and Related Work

Most of the research works about DDoS attacks and the main instrument to carry them out, botnets, have mainly considered technical aspects and have tried to understand how the botnets work, its topologies and their use. In the last years, some of the works have started to complement this technical approach with an economic analysis whose main objective consists in modeling the economic incentives of the attackers and in finding strategies to mitigate risks.

This work belongs to this second group of works. This section describes some of the research works developed up to now and compares them with the one described in this paper.

The study by Liao et al. has some similarities to ours [10]. They propose an economic model for representing the incentives of both actors: the attacker who rents a botnet for launching a DDoS attack and the botnet master who owns that botnet. The attacks will only happen if both obtain benefits. Their goal is to model how the introduction of virtual bots affects the benefits. However, they do not apply the model to any specific scenario using real data. Our work focuses in applying the model to a specific scenario and we have collected some data, such as the cost of hiring a service for launching DDoS attacks.

In another study [5], Franklin and Perrig analyze data collected from the underground markets and propose two possible techniques that hinder transactions. Both try to damage sellers' image so that it increases uncertainty and distrust among possible buyers. Our work focuses in a specific threat, DDoS attacks, and in the prices for launching them, but both works uses data collected from underground markets in order to apply an economic model and propose solutions to mitigate risks.

Another similar work developed by Ford and Gordon [6] focuses on analyzing the revenue generated by malicious code. They do not focus on a single threat but consider the whole set of malicious activities that can provide revenue to botnet controllers such as adware, confidential data sales or renting botnets for launching DDoS attacks. On the other hand, it is a theoretical model and they do not try to apply it to a real scenario.

Last but not least, Friess and Aycock analyze the business case of using botnets for collecting and selling personal information [7]. Although we analyze the use of botnets for other malicious activities, there are some similarities between both works because they develop the business case in order to identify possible defense strategies. In our work, we also use our model to identify how the different defense strategies reduce attack profits.

7.3 The Model

In order to collect data from Internet underground markets we have been analyzing advertisements published by cybercriminals and we have been talking to them through instant messaging clients. The main conclusion we can extract from this activity is that most of them are specialized in concrete activities. Of course, it is possible that someone builds its own botnets in order to launch DDoS attacks and extort victims on her own although it seems that it is not very common. The simplest and less risky way for someone who wants to extort a victim is to hire this service. There are many cybercriminals that are specialized on launching this kind of attacks. They just need an IP and some dollars.

Therefore, our model assumes that there is a high specialization of activities in the underground market and that if someone wants to extort an organization, just needs to hire this service to a botnet master.

The attacker hopes to obtain some revenue by extorting the victim. As a result, the expression for modeling the profit will be as follows:

$$Profit = E - C > 0; \tag{7.1}$$

where E represents the attacker's revenue and C the cost of hiring the DoS service.

This is a general expression for any affected service as we are not considering the particular features of the service yet. Then, the next step consists in analyzing our particular service and identifying the variables that influence in some way the components of the equation 7.1.

The figure 11.1 shows the general scenario of the telecommunication service analyzed in this paper. Residential femtocells are connected to the mobile operator core network through a device called security gateway. Mobile operators' infrastructures have more than a security gateway and each one will have a different IP address. Thus, the attacker will need to launch different DDoS attacks for each of them. We assume that the attackers will arrange the attacks independently for each target gateway. Therefore our model represents the incentives for launching a DDoS attack against a specific security gateway.

We assume that the security gateway links with a total set of "n" femtocells through IPSEC tunnels that encrypt the data that travels between the peers. In addition, the security gateway is able to process up to "r" Gbps of data.

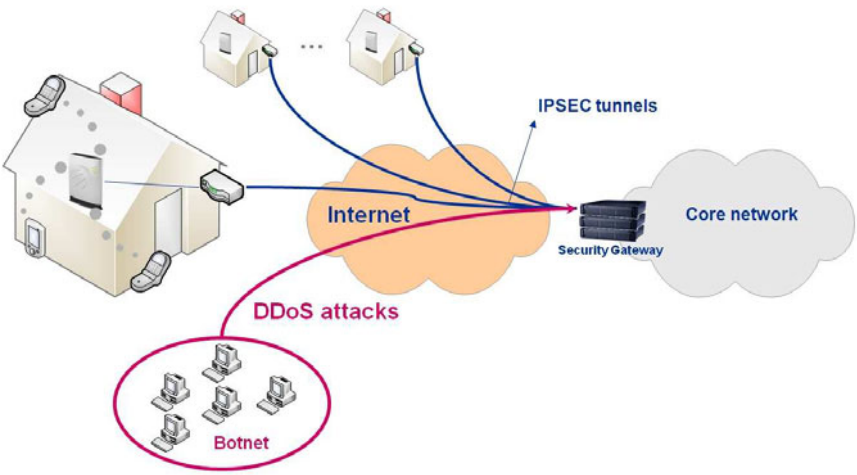


Fig. 7.1 Scenario of the telecommunication service: extended and improved coverage of mobile networks in the residential environment with femtocells.

It would be possible to leave homes without service by launching a DDoS attack that depletes the data processing resources of the security gateway, i. e. an attack of "r" Gbps or higher. The typical DDoS attack against a gateway that sets IPSEC tunnels with several clients is an IKE_SA_INIT flood. Therefore, we assume that any attacker which wants to extort the network operator will have to launch an IKE_SA_INIT flood attack of this bandwidth or higher.

Regarding extortion revenue, we must say that its modeling is much more difficult due to the scarcity of data about this kind of malicious activity. It seems reasonable to assume that the amount asked by attackers should depend on the annual revenue of the mobile operator. It is at least what has happened with online gambling site extortions, one of the more affected businesses by DDoS attacks. The amount asked to them have ranged from 10 000$ to 40 000$ depending on their annual revenues [8, 9]. But asking for an amount of dollars does not mean that the victims pay it. In fact, we hope that mobile operators do not give in to blackmail, as experts and police authorities have recommended, because the less victims giving in to blackmail, the less economic incentives for attackers. However, many news warning about the high percentage of organization that ended up giving in to blackmail have been published [9, 11]. Therefore, at least until mobile operators demonstrate clearly to cybercriminals that they are not going to follow the same way, we think that it is possible that attackers hope that could happen something similar as with online gambling sites. Based on that all, we propose to model the extortion with the following equation:

$$E = \alpha \cdot f(R) \tag{7.2}$$

where α means the percentage of victims that the attacker hopes that will give in to blackmail and $f(R)$ is a function of the victim's annual revenue.

The annual revenue depends on the number of femtocells per security gateway and the average revenue per femtocell. Thus, we can rewrite the equation as follows:

$$E = \alpha \cdot f(nAR) \tag{7.3}$$

where n is the number of femtocells per security gateway and AR is the average annual revenue per femtocell.

We have not been able to infer the function that relates the extortion with the revenue because we have not found concrete data about online gambling site's revenue and the amount of dollars that they have been asked for. However, we know by the references mentioned before that the amount of the extortion was between 10 000$ and 40 000$ during 2004. We have compared these amounts with the annual revenue of some of the online gambling sites [1, 3, 14] and we have seen that they are approximately 1,000 times smaller. Thus, we will use the following equation for simulation:

$$E = \alpha \cdot k \cdot n \cdot AR \tag{7.4}$$

where k is 0.001.

7.4 Application of the Model

We have arranged this section in three subsections. The first one explains how we have collected the data for applying the model. The second one shows the results of a regression analysis that allowed us to estimate the cost of renting a botnet as a function of the bandwidth and the duration of the DDoS attack. The last one uses the results of former subsections to assess attacker's incentives.

7.4.1 Data Collection

In this section we describe how we collected data for estimating both the revenue of attackers and the cost of hiring the DDoS attack service.

7.4.1.1 Extortion Revenue

In the previous section we stated that the attacker's revenue depends on these factors:

- α, the percentage of victims that will give in to blackmail,
- k, a constant whose value is 0.001,
- $R = n \cdot AR$, the annual revenue.

Regarding the first factor, we have not found data or surveys to assess its value. The lack of legislation that forces victims to communicate if they have given in to blackmail and the fear to bad reputation are two reasons that can explain this absence of data.

We consider that there are 20,000 femtocells per security gateway, as it is the number of tunnels that can provide some of the typical devices used in these architectures, such as the Alcatel-Lucent VPN Firewall Brick 1200.

Finally, a business case study from Analysis Research [2] provides an estimate of the average revenue per femtocell. In this study they consider 4 different customer profiles based on the number of handsets per home and the indoor coverage quality. We take the optimistic estimation because it is the worst case for us as it gives the largest incentives for launching the attack. It considers that the service will provide an additional monthly revenue of 28$ per femtocell. Therefore, the annual revenue per security gateway is:

$$R = n \cdot AR = 20,000 \cdot 28 \cdot 12 = 6,720,000\$/year \tag{7.5}$$

And we have the following equation for the extortion revenue:

$$R = \alpha \cdot k \cdot n \cdot AR = \alpha \cdot 6720\$ \tag{7.6}$$

7.4.1.2 Cost of Hiring the DDoS Attack Service

There are plenty of forums in the Internet where one can access to hire this kind of services. This has been our main source of information in this work. We have looked for advertisements where cybercriminals offer this service and we have talked to them through instant messaging applications, mainly the ICQ client, in order to know how the price of the service changes with its particular features. It seems that the cost depends mainly on the bandwidth of the attack and its duration. The figure 11.2 shows the translation of one of the advertisement in a Russian forum.

Part of one of the conversations with a service provider in which we asked for the price of a service with a specific duration and bandwidth can be seen in figure 11.3.

The set of collected prices is shown in the table 7.1.

7.4.2 Regression Analysis for the Cost Function

We have performed a regression analysis using the data showed in the previous table in order to determine the cost of the service as a function of its bandwidth and its duration. The function type that adapts better is a Cobb-Douglas one:

$$C = L \cdot A^{\gamma} \cdot t^{\beta} \tag{7.7}$$

where:

- C, is the cost of hiring the service,

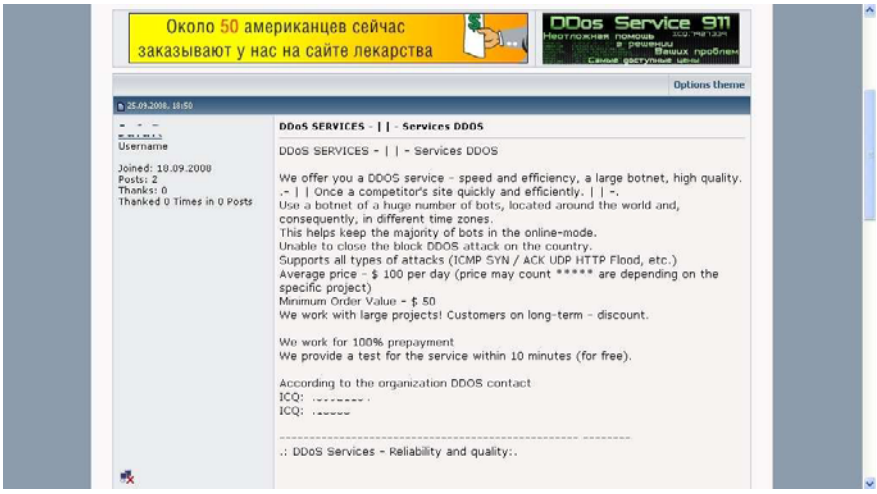


Fig. 7.2 A DDoS Service advertisement in a Russian forum.

Table 7.1 Cost of the service for different bandwidth and duration.

Price ($)	Duration (h)	Bandwidth (Mbps)
20	2	45
30	6	45
50	12	45
70	24	45
75	24	100
250	24	1000
100	24	1000
600	168	1000
900	24	4750
1000	24	4750
5500	168	4750
6000	168	4750
400	5	5000

- L, is a constant,
- A, is the bandwidth depleted by the attack in Mbps,
- t, is the duration of the attack in hours,
- γ, is the cost elasticity of the bandwidth and,
- β, is the cost elasticity of the duration.

In order to apply a linear regression, first we need to transform the equation:

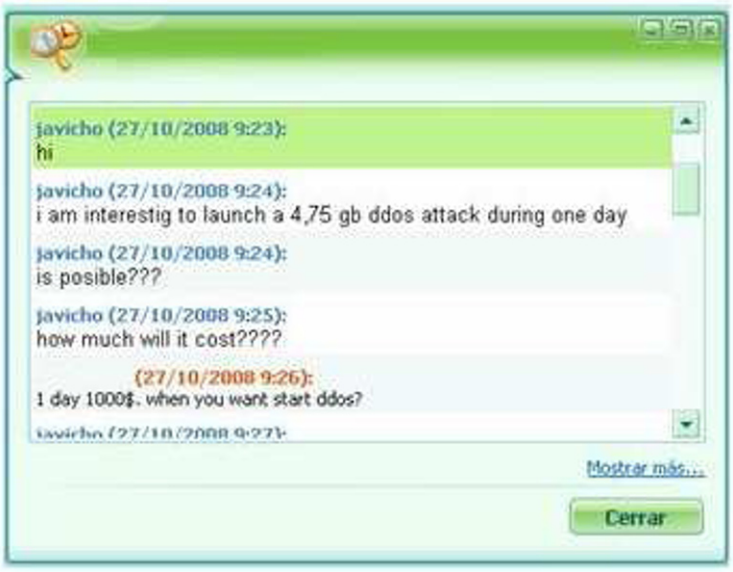


Fig. 7.3 An example of conversation with a DDoS service provider.

Table 7.2 Results of the regression analysis ($R^2 = 0.915$; adjusted $R^2 = 0.898$).

Variable	Value	Standard error	t-statistic	P-value
K	-0.036765	0.568531	-0.064668	0.9497
γ	0.586899	0.095469	6.147523	0.0001
β	0.590331	0.146111	4.04028	0.0024

$$ln(C) = K + \gamma \cdot \ln(A) + \beta \cdot \ln(t) \tag{7.8}$$

We have obtained the following results:
Therefore, the resulting function is as follows:

$$ln(C) = -0.0367 + 0.5869 \cdot \ln(A) + 0.5903 \cdot \ln(t) \tag{7.9}$$

Once transformed to its original form:

$$C = 0.9640 \cdot A^{0.5869} \cdot t^{0.5903} \tag{7.10}$$

Both the absolute value of γ and β t-statistics are greater than 2. Thus, we can assure that there is a strong relation between the two independent variables and the dependent variable. In addition, both R-squared and adjusted R-squared are greater than 0.8 which means that the function represents the relation between the cost and the independent variables with acceptable accuracy.

7.4.3 Use of the Model to Estimate the Economic Incentives for Launching DDoS Attacks

If we substitute in equation 7.1 the expressions for the revenue and costs of the attacker, we obtain the following expression for the profit:

$$Profit = E - C = \alpha \cdot 6720 - 0.9640 \cdot A^{0.5869} \cdot t^{0.5903} \tag{7.11}$$

It seems reasonable to think that the greater the profits the greater the probability of DDoS attacks. On the other hand, potential attackers lose economic incentives when the profit is zero or lower.

The remainder of this section contains three simulations that we have performed using the equation 7.11. For each one we have made some assumptions that are represented with a constant value for some of the variables (α, A and t). Then, we can analyze what values must have the rest of the variables in order to nullify the profit which can be used to identify strategies to protect against the attacks.

7.4.3.1 Simulation 1

In our scenario, the security gateway is able to resist DDoS attacks of up to 4750 Mbps. In addition, we assume that the attacker needs to hire a 24 hours service in order to force the victim to pay. Then, the profit function is as follows:

$$Profit = E - C = \alpha \cdot 6720 - 0.9640 \cdot 4750^{0.5869} \cdot 24^{0.5903} \tag{7.12}$$

To nullify the incentives, the following condition must be met:

$$\alpha \leq \frac{0.9640 \cdot 4750^{0.5869} \cdot 24^{0.5903}}{6720} = 0.1347 \tag{7.13}$$

That means that if the percentage of victims that does not give in to blackmail is 13.47

7.4.3.2 Simulation 2

Let's assume now that a 20% of victims pay and that the duration of hired attacks must be 24 hours. We can calculate the resistance that our infrastructure must have in order to nullify economic incentives:

$$Profit = 1344 - 0.9640 \cdot A^{0.5869} \cdot 24^{0.5903} \tag{7.14}$$

In figure 11.4 we can see how the profit decreases as bandwidth increases. It also can be seen that there is a bandwidth that nullifies economic incentives: 9320 Mbps.

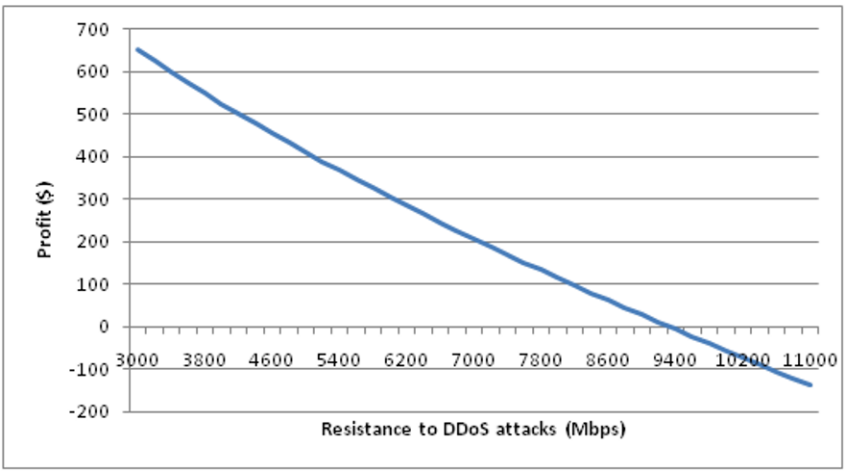


Fig. 7.4 Attacker's profit as a function of the resistance to DDoS attacks.

With a resistance of 4750 Mbps as the one we had initially, launching DDoS attacks would be profitable for an attacker. At that point, the victim has at least two possible strategies in order to reduce the probability of being the victim of these attacks:

- to deploy a security gateway that is resistant to DDoS attacks of 9320 Mbps or higher or
- to design a network architecture that increases the cost of the attacks. Instead of allowing every Internet IP addresses to reach the security gateway, only IP addresses from customers of this service should be allowed. Thus, successful attacks should be launched by bots installed on customer's PCs. It would set a special requirement to botnets that would increase significantly the cost of hiring successful DDoS attacks.

7.4.3.3 Simulation 3

Let's assume again that the duration of attacks must be 24 hours. Then, we can derive the relation between the bandwidth and the percentage of victims that pay that nullify the incentives:

$$Profit = \alpha \cdot 6720 - 0.9640 \cdot A^{0.5869} \cdot 24^{0.5903} = 0 \tag{7.15}$$

$$A = \left(\frac{6720}{0.9640 \cdot 24^{0.5903}}\right)^{1.7039} \cdot \alpha^{1.7039} \tag{7.16}$$

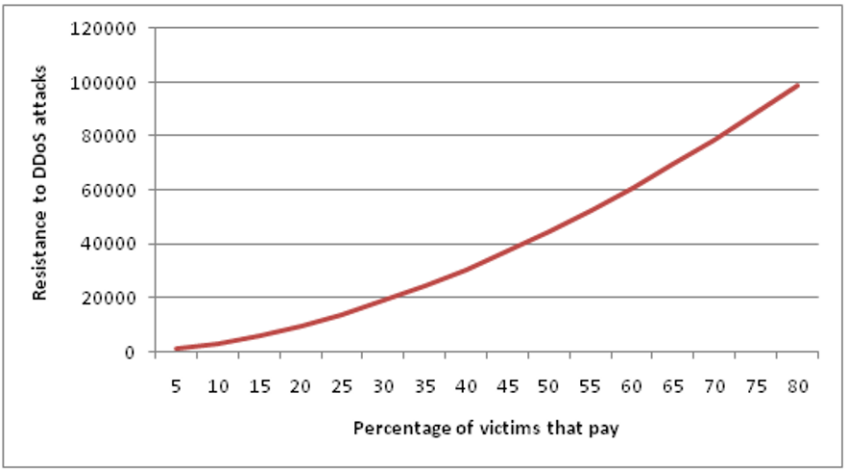


Fig. 7.5 Resistance to DDoS attack that is needed in order to nullify incentives.

The relation between A and α has been represented in figure 11.5. It shows that the resistance increases quickly with the percentage of victims giving in to blackmail.

7.5 Conclusion

The economic motive seems to be one of the main pushing forces of the Internet security incidents. When it happens, attackers' behavior is rational and, even, predictable. Under some assumptions, it is possible to model the conditions that will influence the attacker's behavior and, if we are able to collect data for our model, we can estimate the probability of the attacks.

This paper has focused on applying a simple economic model to a real scenario based on a telecommunication service that many mobile operators will start providing in the short term and that some of them are already providing. This model represents the incentives of a potential attacker for launching DDoS attacks. To apply the model we have collected data from two sources. On one hand, we have searched Internet underground markets to collect prices of hiring DDoS attack services. On the other hand, we have used existing information about past extortions against online gambling sites to estimate the amounts of money that attackers could demand in our service.

The analysis performed in this work is a first attempt to estimate the incentives for launching DDoS attacks based on objective data. We know that the proposed model can be not complex enough to cover some possible situations such as the victims agreeing not to pay up any more, but we preferred to keep the model simple enough so as to apply it to real scenarios using available data. We think the model can be further refined and the technique followed in this work can be used to assess the factors that have influence on risks when there is an economic motivation behind the incidents. In our opinion, that could complement current risk analysis methodologies.

Applying this technique can be hard because of the scarcity of the necessary data. In our case, we have had to contact cybercriminals willing to provide the service, explain them the features of the attack and try to obtain the price of the service. This has not been easy because they tended to mistrust us, especially after using the same ICQ identifier to ask prices to different people. We think that they are well organized and warn each other when there is someone behaving suspiciously.

But understanding better how they are organized and collecting data about the underground markets in an easier and more frequent way would allow us to know and assess more reliably the risks of cybercrime for any service which depends on the Internet. There is plenty of work to do in this field. We hope that our work has contributed somehow to it.

References

1. BETFAIR: Annual Report (2004). http://corporate.betfair.com/key-data/5-year-financial-summary.html
2. Brydon A., Heath M.: Femtocells in the consumer market: business case and marketing plan. Analysis Research (2007)
3. BWIN: Annual Report (2004).
4. Femtoforum: Femtoforum newsletter February (2009). http://www.femtoforum.org/newsletters/newsletter04/index.html
5. Franklin, J., Paxson, V., Perrig, A., Savage, S.: An inquiry into the nature and causes of the wealth of Internet miscreants. In: Proceedings of the ACM Conference on Computer and Communications Security (CCS), pp. 375–388. ACM Press, New York (2007)
6. Ford, R., Gordon, S.: Cent, five cent, ten cent, dollar: Hitting spyware where it teally hurt$. In: Proceedings of the New Security Paradigms Workshop (NSPW), pp. 3–10. ACM Press, New York (2006)
7. Friess N., Aycock J.: Black market botnets. In: MIT Spam Conference. Cambridge, MA (2008)
8. IDG News Service: Super Bowl fuels gambling sites' extortion fears. IT World, 28 January (2004). http://www.itworld.com/040128gamblingsites
9. Ilett, D.: Expert: Online extortion growing more common. CNET, 8 October (2004). http://news.cnet.com/Expert-Online-extortion-growing-more-common/2100-7349_3-5403162.html
10. Li, Z., Liao, Q., Striegel, A.: Botnet economics: Uncertainty matters. In: M.E. Johnson (ed.) Managing Information Risk and the Economics of Security, pp. 245–267. Springer, New York (2008)
11. Pappalardo, D., Messmer, E.: Extortion via DDoS on the rise. Computerworld, 15 May (2005). http://www.computerworld.com/networkingtopics/networking/story/0,10801,101761,00.html
12. Schneier, B.: Does risk management make any sense? (2008). http://www.schneier.com/blog/archives/2008/10/does_risk_manag.html
13. Symantec: Symantec Internet Security Threat Report XIII (2008).
14. UNIBET: Annual Report (2004). http://www.unibetgroupplc.com/corporate/templates/KeyFigureList.aspx?id=113

Chapter 8
The Privacy Jungle: On the Market for Data Protection in Social Networks

Joseph Bonneau and Sören Preibusch

Abstract We have conducted the first thorough analysis of the market for privacy practices and policies in online social networks. From an evaluation of 45 social networking sites using 260 criteria we find that many popular assumptions regarding privacy and social networking need to be revisited when considering the entire ecosystem instead of only a handful of well-known sites. Contrary to the common perception of an oligopolistic market, we find evidence of vigorous competition for new users. Despite observing many poor security practices, there is evidence that social network providers are making efforts to implement privacy enhancing technologies with substantial diversity in the amount of privacy control offered. However, privacy is rarely used as a selling point, even then only as auxiliary, non-decisive feature. Sites also failed to promote their existing privacy controls within the site. We similarly found great diversity in the length and content of formal privacy policies, but found an opposite promotional trend: though almost all policies are not accessible to ordinary users due to obfuscating legal jargon, they conspicuously vaunt the sites' privacy practices. We conclude that the market for privacy in social networks is dysfunctional in that there is significant variation in sites' privacy controls, data collection requirements, and legal privacy policies, but this is not effectively conveyed to users. Our empirical findings motivate us to introduce the novel model of a privacy communication game, where the economically rational choice for a site operator is to make privacy control available to evade criticism from privacy fundamentalists, while hiding the privacy control interface and privacy policy to maximize sign-up numbers and encourage data sharing from the pragmatic majority of users.

Joseph Bonneau
Computer Laboratory, University of Cambridge, e-mail: jcb82@cl.cam.ac.uk

Sören Preibusch
Computer Laboratory, University of Cambridge, e-mail: sdp36@cl.cam.ac.uk

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_8,

8.1 Introduction

In the past decade, social networking sites have become a mainstream cultural phenomenon [27]. Social networking has become one of the most popular activities on the web, with the top sites boasting hundreds of millions of users, and social networking sites representing 16 of the world's 100 most-visited sites [1]. Their popularity amongst the younger generation is even higher, with studies finding more than 80% of American university students active social network users [8, 48], commonly spending at least 30 minutes every day on social networks [52]. The ubiquity of social networking in youth culture has been likened to an addiction [30].

Social networks have also obtained a poor reputation for protecting users' privacy due to a continual flow of media stories discussing privacy problems [44]. Popular media angles include the disclosure of embarrassing personal information to employers [19, 37] and universities [76], blackmail using photos found online [68, 72], social scams [12, 39, 49], and user backlash against newly introduced features [75, 80].

Despite the focus of the English-speaking media on Facebook, MySpace, and occasionally Bebo, and the common perception of an oligopolistic market, there is a flourishing supply of social networking services, with dozens of large general-purpose sites competing alongside thousands of niche sites. In our study, at least 25 different services were found to be the most popular social network in at least one country [1].

There is also a common misconception that privacy violations occur routinely because the generation of (mostly younger) social networking users fundamentally do not care about privacy. This is contradicted by studies where most social network users do express an interest in privacy [8,23,31,42]. Given the plethora of competing sites, the functional similarity of most social networks, and users' stated concern for privacy, market conditions appear prime for sites to compete on the basis of privacy. This was our overarching research question as we conducted—to the best of our knowledge—the largest and most comprehensive field study in the academic literature of the global social network market. Past studies have focused on studying users of social networks; our study is unique in that we compare the sites themselves. We have attempted to collect as large a sample of sites as possible, focusing on what is offered and promoted in the way of privacy rather than on user behavior.

Our contribution is threefold. First, we report the results of a thorough analysis of the privacy supply in the social networking market (Section 8.4). Our data supports some common assumptions, such as a generally low quality of privacy policies, usability problems, and poor security practices. It also provides some surprises such as promotion of photo-sharing being far more common than game-playing, and a huge diversity of privacy controls available in different networks which is not effectively conveyed to users.

Second, we aggregate our data into overall privacy and functionality scores for each site, and use these to find which general factors may influence a site's privacy practices (Section 8.5). Again, we find interesting results, such as niche sites offering significantly less sophisticated privacy controls than general-purpose sites,

positive correlations between privacy and the age, size, and popularity of a site. Privacy and functionality aren't strong correlated, but sites that promote on privacy are often found having less favorable privacy practices. We also find evidence that sites with better privacy are growing ahead of the market, while those that mention their privacy are falling behind.

Finally, we propose a novel economic model to explain the observed under-supply and under-promotion of privacy as a rational choice by the competing social networking providers. Our model assumes the existence of consumers with varying degrees of privacy concern. We conjecture that websites seek to maximize their desirability to both populations by not raising privacy concerns for the majority of users, while minimizing criticism from the privacy-sensitive. We explore this, along with other economic explanations, in Section 8.6.

8.2 Related Work

Given the broad aims of our study, there is a large body of relevant prior research. Social networks have been an active research area in several academic disciplines in recent years. Sociologists have studied them from an ethnographic perspective, examining why they have become popular and what motivates individuals to participate [23, 27, 29, 30, 85]. Others have used surveys to examine users' attitudes towards social networks, in particular with regards to information sharing and disclosure [8, 31, 42, 48]. User studies have also been performed by automatically analyzing crawled profiles [8, 48, 52, 55]. Computer scientists have performed more quantitative studies of social graph formation, using web crawlers to study the size and link structure of social graphs [25, 41, 54, 63].

Security and data protection in social networks has recently become an active research area recently. Many researchers have outlined the potential threats and risks associated with using social networking services [17, 74]. White-hat studies have identified many security flaws due to implementation errors in social networks [20, 21, 34]. Security researchers have also taken a constructionist approach. Several interfaces have been proposed for allowing users to more easily manage privacy [11, 57, 67, 77], a few of which we saw beginning to be deployed in sites we analyzed. Some have proposed new architectures which can provide stronger privacy guarantees [10, 24, 35, 36, 71], while others have recommended implementing privacy-preserving front-ends for existing social networks [43, 60].

Privacy issues specifically arising from the graph of friendship links have been studied as well, and have identified graph privacy as a major issue. Social context was shown to make phishing attacks much more successful [46]. Several studies have indicated that knowledge of the social graph can enable accurate inference of private data [56, 90, 91]. It has been shown that it is impossible in practice to "anonymize" a social graph by removing names due to the amount of unique structure contained within it [15, 38, 65]. Social graph privacy has also been shown to be

very fragile in that obtaining a relatively small amount of information enables many useful calculations to be made [21, 22, 28, 53, 64].

Privacy has also been extensively studied from an economics perspective. The problem has been formally modeled as an economics trade-off between disclosing information and gaining access to desirable online services [45, 70, 82]. Researchers have utilized surveys to gauge user attitudes about privacy, consistently showing a high stated concern for privacy [26, 62, 78]. Direct observational studies of users have often contradicted these studies though, showing that users often violate their stated privacy concerns [6, 40, 69, 79]. Economists have attempted to resolve this "privacy paradox" by proposing models which describe why users' long-term preferences for privacy are ignored or forgotten when interacting with a website [7, 9, 58, 83]. This has been shown specifically to occur in the case of social networks, where individuals with high self-reported privacy awareness revealed significant amounts of data on their profiles [8]. Other research has focused on privacy policies, usually finding them to be far too difficult for ordinary users to understand [62, 86]. Computer scientists have proposed technological solutions to improve users' ability to make privacy choices [5, 16, 18, 33, 73].

8.3 Survey Methodology

8.3.1 Selection of Sites

We selected 45 social networking sites for our survey, the complete list is provided in Table 8.1. Our goal was both to conduct an exhaustive survey of the major, general-purpose social networking sites, and include several representatives of other common social-networking niches for comparison.

8.3.1.1 General-Purpose Sites

Our operational definition of a *general-purpose* social networking service is one which anybody is free to join, people commonly present their real-world identity, and the primary use of the site is interacting with others via profile pages on the Web. This excludes sites whose primary purpose is sharing content (e.g. YouTube, Flickr), sites which enforce limited membership (invitation-only networks such as A Small World), or sites where few users reveal any real-world information about themselves (such as online poker websites). While some of these services contain all or almost all features of general-purpose social networking sites, they can be separated by their different patterns of typical use. For example, a web crawl revealed that average users of YouTube and Flickr make less than 10% as many connections as those using Orkut [63].

Our definition is mostly functional, and does not exclude sites which are mainly populated by specific demographics of users. Several of the sites we regarded as

general-purpose target a specific demographic niche. For example, BlackPlanet is targeted to African Americans living in the United States, Eons is targeted at the older generation, and MyYearbook and Bahu are targeted specifically at teenagers. MocoSpace is a general-purpose social network on the web which additionally aims specifically to be usable via mobile devices. However, we still regard these sites as general-purpose as their feature set is similar to other general-purpose sites, they simply cater to a specific group of people with their marketing and graphic design.

An important omission from our study is sites which are not available in English. This includes several very large general-purpose sites, such as the Russian site VKontakte, the Japanese site Mixi, and the Spanish site Tuenti. This decision was necessary to ensure fair comparison between sites, particularly for privacy policies where word choice is critical. Our focus on the Web also excludes communication services such as Instant Messaging, online role-playing games such as World of Warcraft, and 3D virtual worlds such as SecondLife.

Within this definition, though, we were able to include 29 popular general-purpose sites from around the world, listed in full in Table 8.1. We enforced a minimum size of 500,000 users for general-purpose sites to keep the study tractable.

8.3.1.2 Niche Sites

In addition to general-purpose social networks, we examined 16 *niche* social networking services, also listed in full in Table 8.1. These sites either have a subset of general-purpose sites' functionality or are used in significantly different ways.

— **Business-networking sites** differ from general-purpose in that they specialize in maintaining professional contacts and searching for new jobs. Users typically share much less personal information, yet more professional information on these sites. They often implement specific features for specifying and managing business relationships and are frequently used for job-searching. We included LinkedIn, XING, and Viadeo, the most popular business-networking sites.
— **Media recommendation sites** specialize in allowing users to recommend and share films and music. While they have many features of general-purpose sites, users often interact with others based on similar tastes in music or movies, rather than real-world social connections. We included Last.fm, Imeem, Flickster, and Buzznet in this category.
— **Reunion sites** specialize in allowing people to search for old acquaintances from school or the military rather than actively maintaining profiles. They often aggregate contact information only and are designed to facilitate off-line connection rather than on-line interaction. We included Classmates.com and myLife (formerly Reunion.com) as representatives of this genre.
— **Activity-focused sites** center around allowing users to perform a specific activity. Habbo and Gaia Online are two pre-eminent gaming-centric social networks. CouchSurfing is designed for students and youth to share accommoda-

tion while traveling.[1] Finally, we included the surging micro-blogging service Twitter in this category, though arguably it is in a niche by itself.

— **Privacy-specific sites** have specific privacy-enabling features. Experience Project is designed as a pseudonymous social network for users to share intimate stories with strangers who have had similar life experiences. Imbee is a fledgling social network aimed to be safe for younger children, with strong administrative oversight and parental controls. Kaioo is a non-profit social network designed to be community-owned and governed, for those uncomfortable trusting their social details to a private company. We included Imbee and Kaioo in our survey due to their unique privacy goals, though neither site has an established user base yet.

8.3.2 Evaluation Methodology

We conducted a standardized, scripted evaluation for each website. The evaluations were conducted in February 2009, and all data is accurate as of the time of evaluation. Due to the rapid evolution of social networking, several data points had already changed by the time of analysis, but we kept all values as a consistent snapshot of the time of collection, recorded alongside the data itself.

8.3.2.1 Data Collection

First, we collected general information about the site, such as its launch date, estimated user count and traffic ranks, country of operation, and ownership status (presented in Section 8.4.1). Next, we examined the publicly viewable sections of the webpage which are presented to non-members who visit the site (typically after receiving an invitation by friends who are already members of the site). These offer the most valuable insight into the marketing strategies used by social networks, since very few rely on traditional advertisements. We recorded the selling points used to encourage visitors to sign up (Section 8.4.2).

Next, we signed up for each site, recording the amount of personal information required in order to register an account (Section 8.4.4). We also recorded the means by which users are presented with the sites' Terms of Use and/or Privacy Policy during sign-up (Section 8.4.3). We then evaluated the extent of privacy controls available to users of the site, and the default values provided with a new account (Section 8.4.5). In addition to privacy controls, we recorded general security features like the use of encryption, the existence of help pages for controlling privacy, and the existence of infrastructure for reporting abuse (Section 8.4.6).

Finally, we evaluated the formal privacy policy provided by each site (Section 8.4.7). Evaluation criteria for the privacy policies included accessibility, length,

[1] Because its intended use is connecting strangers, CouchSurfing is notable for having a complicated reputation system built into the site to encourage safety.

Table 8.1 Evaluated Social Networks, $N = 45$. User count in millions, rounded.

Site	Traffic Rank	Users (M)	Country	Category
Windows Live Spaces	4	120	USA	General-purpose
Facebook	5	175	USA	General-purpose
MySpace	7	250	USA	General-purpose
hi5	17	60	USA	General-purpose
SkyRock	43	13	France	General-purpose
Friendster	45	95	USA	General-purpose
NetLog	71	35	Belgium	General-purpose
Tagged	75	70	USA	General-purpose
Orkut	83	67	USA	General-purpose
LiveJournal	85	18	Russia	General-purpose
Bebo	119	40	USA	General-purpose
PerfSpot	124	20	USA	General-purpose
meinVZ	156	12	Germany	General-purpose
Multiply	161	12	USA	General-purpose
Badoo	168	19	UK	General-purpose
Sonico	183	33	Argentina	General-purpose
Ning	187	1	USA	General-purpose
CyWorld	315	20	South Korea	General-purpose
Xanga	346	40	USA	General-purpose
MyYearbook	406	15	USA	General-purpose
BlackPlanet	1021	18	USA	General-purpose
Plaxo	1486	20	USA	General-purpose
MocoSpace	2582	2	USA	General-purpose
Hyves	4166	8	Netherlands	General-purpose
Impulse	4782	1	Bulgaria	General-purpose
Yonja	5142	4	USA	General-purpose
Bahu	9977	1	France	General-purpose
Nexopia	12109	1	Canada	General-purpose
Eons	17872	1	USA	General-purpose
LinkedIn	149	35	USA	Business-networking
Imeem	186	30	USA	Media recommendation
Last.fm	317	21	USA	Media recommendation
Twitter	338	6	USA	Micro-blogging
Classmates.com	519	40	USA	Reunion
Gaia Online	628	7	USA	Gaming
MyLife	796	58	USA	Reunion
BuzzNet	954	10	USA	Media recommendation
Flixster	975	62	USA	Media recommendation
XING	1023	7	Germany	Business-networking
Viadeo	3280	7	France	Business-networking
Habbo	3349	124	Finland	Gaming
CouchSurfing	4326	1	USA	Travel
Experience Project	8878	2	USA	Privacy-specific
Kaioo	120679	n/a	Germany	Privacy-specific
Imbee	248170	n/a	USA	Privacy-specific

collection and retention of user data, the role of third-party advertisers, and compliance with privacy laws.

In addition to the raw data points, we computed and included in our dataset aggregate metrics per site. In particular, we define scores for data collection, privacy control, privacy policies, and functionality, presented in Table 8.7.

8.3.2.2 Data Provided During Signup

To ensure fair comparison, we supplied consistent data when asked to the fullest extent possible, and consistently withheld any information which was not mandatory. We signed up for an account with each site using the name "Upton Sinclair,"[2] a birth date of September 20, 1978, the Cambridge postcode `CB30DS`, and other standardized personal information consistent in all created accounts. We provided the same Yahoo! email account with a *ymail.com* suffix to each site. We only varied this information in a few necessary cases, such as Bahu, which prohibits users over the age of 25, or for US-targeted sites which required US postal codes.

8.3.2.3 Technical Set-up

Recognizing that websites may tailor interaction based on any observable data about the user, we were careful to keep the interaction conditions constant. All browsing was performed using IP addresses from the Cambridge Computer Laboratory's address space `128.232.*.*`. During sign-up and interaction with the studied websites, we used Mozilla Firefox v 3.0.6 running on OpenSUSE 11.1 Linux, configured to accept all cookies. We made use of the Screen Grab! v 0.95 plugin to capture images of web pages, as well as the CipherFox v 1.76 plugin to examine TLS connection details.

Examination of sites' Terms of Use and Privacy Policies was performed using a separate machine, running Internet Explorer 7.0 on Windows Vista. This was done to ensure that these documents would be presented as would be shown to a non-member of the site who is considering signing up.

8.4 Data

This section summarizes our major observations from the data we collected. In addition to the figures presented in this section, we have made our entire dataset available online for public analysis.[3]

[2] In honor of the pioneering investigatory journalist.

[3] *http://preibusch.de/publ/privacy_jungle/*

8.4.1 Market Dynamics

8.4.1.1 Network Size

The number of large social networks is impressive, though it is difficult to fairly assess their relative size. It is impossible to externally determine the number of members of a site, so we have relied on the sites' own claims, where available, and the most recent external estimates in other cases, giving us a rough estimates of network size in Table 8.1.

Member counts mean different things on different sites, for example, many users of Habbo control multiple accounts, inflating the reported number of users, while operating multiple accounts is uncommon and/or banned on other sites. Ning provides a particularly challenging case, as the service allows users to create "their own social network" from a template. Statistics are only published on the number of social networks created (500,000+) which surely underestimates the total number of users.

There are also problems due to large numbers of inactive or rarely-accessed accounts. Windows Live Spaces is particularly problematic because it automatically creates a profile page for every Hotmail user, leading to a huge number of reported users, despite many not actively maintaining their profile. This points to the larger problem of user account statistics including inactive or rarely-accessed accounts. Finally, we were unable to locate any reliable estimates for Imbee and Kaioo, both still too small report user numbers.

8.4.1.2 Site Popularity: Traffic Data

Due to the problems with network size, we feel that traffic data is a fairer indicator of a site's popularity, though this has complexities as well. We relied on the publicly available Alexa traffic rankings [1]. While these are commonly used as a general indicator of the amount of traffic a site is receiving, the algorithm to compute them is not publicly available so it is impossible to scientifically evaluate their accuracy.

Furthermore, because traffic rankings are produced at the second-level domain granularity, there are several difficulties for social networks which either share a domain with other services, or are spread across several domains. Windows Live Spaces again appears far more popular than it actually is, because *spaces.live.com* shares its traffic rank with *search.live.com* and other more popular services. Collectively, the *live.com* domain has the #4 traffic rank, although the social networking service accounts for just 1.9% of this traffic. On the other hand, MeinVZ operates under both the *meinvz.net* and *studivz.net* domains, which rank 380 and 156, respectively. In these cases, we simply took the rank of the highest-ranking domain, since there is no way to combine opaque or sub-divide opaque rank data.

8.4.1.3 Geographical Distribution: American Dominance

With two thirds of our sites head-quartered in the USA, we were initially concerned that our study appeared heavily biased towards American-operated sites, especially given our decision to exclude non-English language sites. However, after analyzing usage data we now believe that this mostly reflects the concentration of global web companies in the Silicon Valley area, as indeed most of the American-operated sites are based in the San Francisco Bay Area. We identified an interesting trend in that a number of large sites are based in the United States or at least nominally owned by American parent companies, despite being far more popular in foreign markets [1].

Orkut was famously designed for the US market by Google but has caught on primarily in Brazil and India, where it is now the most popular service. Hi5 is probably the best example, being founded in 2003 in San Francisco, and maintaining a traffic rank of just 96 in the USA, but being the most highly trafficked social networking site in countries as diverse as Honduras, Romania, Thailand and Angola. LiveJournal was founded and run in the USA for almost a decade despite being most popular in Russia, until finally being purchased by a Russian media conglomerate last year. Friendster is an interesting example: it was once the most popular service in the US market, but usage there has drastically fallen off [27], though it remains very popular in Asia, where it is the most popular service in Indonesia, Malaysia, and the Philippines. While these sites have caught on elsewhere despite being designed for the American market, Yonja was founded in the US by Turkish immigrants, and is almost exclusively visited by users from Turkey, though it is still operated in the US. Bebo has followed the opposite path to American ownership, starting in London and being recently purchased by US-based AOL, Inc., despite the majority if its users living in the UK and Ireland.

8.4.1.4 Site Evolution

Another interesting trend we observed by studying site histories is that many of the sites studied were not originally launched with the goal of becoming large social-networking services, but have evolved into them over the years. Facebook began as a service only for US university students, and MeinVZ similarly began as a directory service for German university students called StudiVZ. Both are now multi-lingual services open to the general public.

Other sites began with simple functionality, and gradually added social features to the point that they now look like general-purpose sites in many respects. LiveJournal, Xanga, and SkyRock (formerly SkyBlog) all began as blogging services, Classmates and MyLife both began with the goal of finding old classmates, and the media-sharing sites began only with anonymous media-ranking functionality. Similar to Zawinski's Law which predicts that all software expands until it can send mail, we propose a new law that all websites expand until users can add each other as friends.

The average age of the networks in our study is just 5.2 years, 5.07 for the general-purpose sites and 5.46 for the others. Impulse, Bahu, Kaioo, and Sonico were the only sites studied which launched within the past 2 years. Classmates, launched in 1995, is by far the oldest, with the next oldest being LiveJournal, CyWorld, and BlackPlanet, all launched in 1999. All of these sites had substantially different purposes when they launched.

8.4.1.5 Multilingualism

The degree of multilingualism in the sites surveyed was high, indicating that the online social networking paradigm is popular across many different cultures. The average site was offered in 9.1 languages, although the median was just 2, and the standard deviation was 11.1. There is a bimodal distribution between a number of sites offered in just 1 or a small handful of languages, and some very well internationalized sites. 7 sites (NetLog, hi5, Orkut, LiveJournal, Facebook, Windows Live Spaces, and PerfSpot) were offered in at least 25 languages. PerfSpot took the lead with an impressive 46 languages, including Cebuano, Estonian, and Tamil.

8.4.1.6 Competition

In addition to the variety of languages offered, we analyzed country-specific traffic rankings provided by Alexa to approximate the national markets in which sites are competing for new users. As a rough heuristic, we considered two sites to be "competing" within one national market if their traffic ranks are within a factor of two of each other. Using this metric we found significant competition is occurring; every single site surveyed is competing to catch another social network in at least one market. In the English-speaking world, Facebook, MySpace, and Bebo are fighting in different orders as the top 3 in the UK, Ireland, Australia, and New Zealand, with Facebook and MySpace alone competing at the top in the USA and Canada.

There is a common market dynamic throughout Europe, with most countries having a home-grown service competing against a larger, international challenger (usually Facebook but also MySpace, Hi5, and others). Facebook is currently one spot in the rankings behind local competitors SkyRock and Bebo in France and Ireland, respectively, and has recently overtaken local competitors Hyves, meinVZ, and Impulse in the Netherlands, Germany, and Bulgaria, respectively. Even CyWorld, which has dominated the South Korean market for a decade, is now seeing competition from Friendster and Facebook which have slipped into the top 20 sites for the country.

8.4.1.7 Business Model

Most sites rely on advertisements for revenue, with only the non-profit sites CouchSurfing and Kaioo, and the children's site Imbee not displaying advertisements. We also observed that 7 of the 29 general-purpose sites (24%), but 10 of the other 16 (63%) offered paid premium memberships. These premium memberships typically allow more space for uploading pictures, more control of one's profile, and the removal of advertisements. Premium memberships were offered on all of the business-networking sites and reunion-focused sites, and seem to be a major revenue stream: XING, for instance, generates 80% of its revenue from the 8% of users who are premium members [89]. Many other sites offered the ability for users to buy each other virtual gifts, though these typically sell for only $1 or €1.

Overall, there is a lack of reliable data on the financial situation of social networks, with almost all of them still privately held and operating as start-ups reliant on outside financing. The global market for social networking advertisements is estimated to be US$2.66 billion in 2009 [88], but some market analysis has questioned the profitability of sites given the slow growth of advertising revenue and sites' large operating costs [47].

8.4.2 Promotional Methods

Most social networks rely on word-of-mouth promotion and there is very little external advertising. However, most sites promote themselves aggressively to non-members who visit in the hope of converting visitors into new users. We compared this promotional process across networks, grouping the most common promotional tactics used into several categories displayed in Fig. 8.1.

8.4.2.1 Promotion of Social Interaction

Unsurprisingly, a very common marketing strategy is promotion of social interaction on the site. This was observed in sites promoting the ability to send messages using the site (20 / 69%), and extending the possibility of meeting new friends (17 / 59%). These approaches seem to loosely capitalize on the network effects of the site to indicate that one should join based on the ability to interact with other users already on the site.

8.4.2.2 Promotion via Network Effects

Capitalizing on network effects was an explicit strategy for 23 general-purpose sites (79%) which showed a sample of user photos from the site and/or explicitly listed the number of user accounts. This was in fact the most common promotion observed

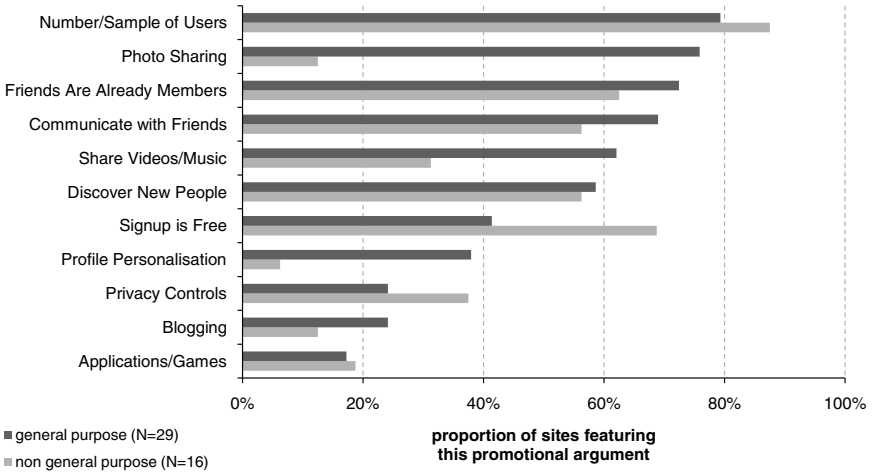


Fig. 8.1 Promotional technique prevalence.

in the sites studied. In addition to listing the total number of users, often as a live counter, many sites listed the number of users who were either currently logged in or came from the same area as the visitor in a further attempt to make the site appear actively used. 21 sites (72%) employed some variation of the argument that "Your friends are already using the site." Network effects were in fact even more commonplace in the niche sites surveyed, being used by all of the media-sharing sites, business-networking sites, gaming sites, and reunion sites.

Of the sites showing sample user profiles, some designated user profiles or content as "featured" while others purported to be showing a random sample. In no case did we actually see a random sample that changed when visiting from a separate IP address, indicating that most "sample" users were in fact selected specifically to make the network seem appealing.[4]

The heavy use of network effects is no surprise. User surveys have usually found that the most common reason given for joining a site is because users felt that the majority of their friends were already using it [23, 42].

8.4.2.3 Promotion of Functionality

Where general-purpose social networks choose to promote their actual functionality, the ability to share photos was by far the most common feature mentioned, advertised by 22 sites. Sharing videos and music was almost as common, mentioned by 18 sites. This suggests the interesting possibility that photo-sharing may be the real

[4] We certainly noticed a preponderance of attractive and young individuals featured in these photos.

killer-application which is driving social-networking growth.[5] Every single general-purpose site we surveyed implements the ability to upload multiple photos, whereas only 5 of the 16 other sites implemented photo-sharing, making the ability to share and tag photos seem to be a requirement for a general-purpose social network. This difference is indeed highly significant at $p = 0.04$.

In contrast, the ability to install applications or play games was relatively rarely promoted given the huge amount of attention received by the Facebook development platform and the claim that is decisive factor in the site's popularity [41]. Facebook itself did not mention its application platform. 14 of the surveyed sites implement some platform for third-party applications which users can add to their profiles, but only 5 mention this promotionally, indicating this is not yet considered a major selling point. Other functionality, such as the ability to blog (promoted by 7 sites) and the ability to customize one's profile (11 sites) were similarly much less common in marketing than photo and media-sharing.

The fact that account sign-up is free was promoted by 21 sites, although all the general-purpose sites we surveyed offered free accounts. The freeness of the accounts may be relatively unsurprising today as consumers are conditioned to expect web services to be free. However, 7 of surveyed general-purpose sites do offer premium accounts, usually removing advertising and offering more storage for a monthly fee. 4 of these 7 optionally paid-for sites still promoted free sign-up, a higher percentage than sites without paid accounts. Similarly, there was an increase in promotion based on sign-up being free among the niche sites, despite a higher proportion of them offering paid memberships. This is possibly an indication that consumers are less likely to expect sites in the areas of music, gaming, and business to be free.

8.4.2.4 Promotion of Privacy

Finally, privacy was used as a selling point in 7 out of 29 general-purpose sites, and when it was mentioned it was typically in a vague and general fashion. 4 sites explicitly mentioned privacy: PerfSpot claimed "unmatched privacy controls," meinVZ offered "a wide range of settings that concern your privacy," Eons mentioned the ability to "control your privacy," and Sonico to "share photos, videos, and your interests privately." 3 other sites made vague reference to access control: Windows Live Spaces stated that you decide "who sees your space, and who doesn't," Multiply claimed it was easy to share photos with "more people you want and fewer people you don't," and Hyves stated "you decide which information is available to whom." Hyves also deserves commendation for promising "we'll never sell your information,"—the only site we observed making such a guarantee. None of these promotions made any reference to or linked to the site's privacy policy, no site attempted to use the contents of its privacy policy as a promotional tool. 2 of the 3

[5] Indeed, Facebook hosts more user photos than any other website, with over 40 billion.

number of promotional arguments	promotion on privacy no	yes
$\leq$ avg	23	2
$>$ avg	9	11
significance	$p = 0.0008$	

Table 8.2 Privacy as a promotional argument is found significantly more often when many other arguments are also deployed.

Fig. 8.2 Weak privacy promotion in a long feature list (Eons).

business-networking sites mentioned privacy, as did 2 privacy-specific sites, but just 2 of the 11 other niche sites mentioned privacy.

In addition to the relative rarity with which privacy was mentioned promotionally, we found strong evidence that it is not used as a primary argument, but as one of many items in a long feature list. For general-purpose sites, sites mentioning privacy used an average of 8.0 promotional categories, whereas sites not mentioning privacy used an average of 5.74. Privacy was never mentioned by a site which used fewer than 5 other promotional arguments. Fisher's exact test reveals strong statistical significance in that privacy only emerges as a "yet another" argument (Table 8.2). The promotional page from Eons (Figure 8.2) provides a typical example of privacy being mentioned in a nondescript way, lost among other features.

8.4.3 Presentation of Terms of Use and Privacy Policy

We recorded the means in which users where presented with the site's Terms of Use and Privacy Policy, as signing up is effectively a legal agreement governed by the these documents. Typically, there is a disclaimer placed near the submission button during signup which contains a reference to the Terms of Use, and sometimes the Privacy Policy as well. A particularly clearly-stated example from MySpace is shown in Figure 8.3. Unfortunately, most sites made scant mention of their privacy policies during sign up.

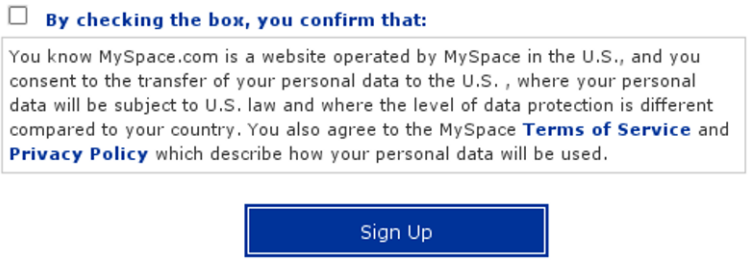


Fig. 8.3 Terms of Use and Privacy Policy acknowledgment (MySpace).

8.4.3.1 Privacy Policy Acknowledgment

Despite signing up being considered legal acceptance of the Privacy Policy in every site studied, only 5 of the 29 general-purpose sites required actively checking a box to indicate acknowledgment of the privacy policy, whereas 12 require checking a box to acknowledge the terms of service. 17 sites mentioned the privacy policy on the signup page, although only 11 of these placed the privacy policy reminder on the user's critical path, with 3 placing it in the page's margin and 3 placing it below the submission button. Results were even worse for the other sites surveyed, with 10 sites of 16 mentioning the privacy policy, but only 4 placing the reminder above the submission button.

8.4.3.2 Privacy Policy Review

In addition to not forcing users to actively acknowledge the privacy policy, very few sites encouraged users to read it. MeinVZ was a commendable exception, displaying a condensed version of the site's privacy policy on a separate page during the signup process. MySpace and Viadeo both displayed shorter extracted paragraphs from their privacy policies, and Imbee gave users a strong nudge to "Read our PRIVACY POLICY!" However, the remaining sites, including 27 of the 29 general-purpose sites, included essentially no pertinent information about the privacy policy on the user's path to creating an account.

Ten general-purpose sites made no reference to the privacy policy at all. Of the 17 general-purpose sites which did mention the privacy policy, 4 of them forgot to include a link to actually read it. 11 sites failed to mention the policy but provided a link in a standardized page footer, and 5 offered no link at all. On the sites linking to the privacy policy from a footer, typically it was grouped with many other links including help info, contact info, and information for advertisers. A glaring example is shown in Figure 8.4, as Friendster buried its privacy policy link along with 7 other links and a list of patents held. An additional 2 sites made the mistake of including links which did not open in a new window or tab, meaning that clicking on them would interrupt the signup process. Of the non-general-purpose sites, 4 failed to

provide links to their privacy policies during signup, and 2 more included links not opening in a new window.

Fig. 8.4 Privacy Policy link hidden in bloated page footer (Friendster).

8.4.4 Data Collected During Sign-up

While signing up for each of the networks in our study, we recorded the amount of data which must be reported create a new account. We also recorded the amount of data which is requested but not required, though we consistently chose to withhold such data. We found remarkable variation between the general-purpose sites as to what data was collected, summarized in Figure 8.5.

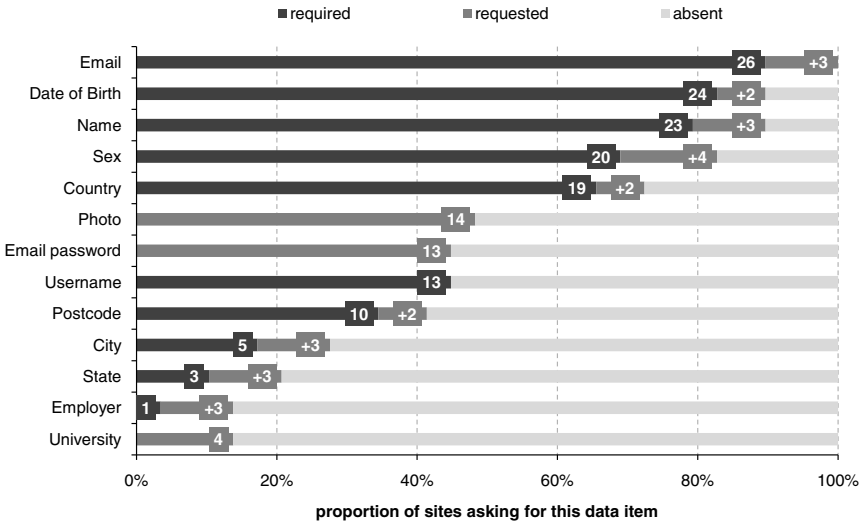


Fig. 8.5 Visibility of profile data, general-purpose networks, $N = 29$.

8.4.4.1 Over-Collection of Demographic Data

In general, far more personal data is collected than is needed for a user to interact with a social networking service, particularly gender and birth date information. Gender was required by 20 sites and requested by 4 others. A full date of birth was required by 24 sites and requested by 2 others.[6]

These two pieces of data are both useful to personalize the site but should not be mandatory. We did observe several sites promoting reminders of friends' birthdays as a reason to use the site; the huge popularity of this feature could be a reason that this data is often required [30]. Similarly, the majority of the sites offer demographic search capabilities: 22 out of 29 sites allow finding fellow members based on location, gender, interests, and other data, instead of just name.

Photographs and information on employment and university affiliations are similarly unnecessary, but were not required except in the case of BlackPlanet, which requires a user's "Job Type." BlackPlanet was an outlier as it also requested a user's race, ancestry, income level, and sexual orientation during sign-up. Yonja went a step further in actually requiring users to report their sexual orientation.

8.4.4.2 Requirement of Real Names

The widespread requirement of reporting names is similarly troubling, as 23 of the 29 sites require entering one's full name to join the site.[7] Only 3 sites were purely pseudonymous (Nexopia, Xanga, MocoSpace), with 3 other sites (LiveJournal, SkyRock, BlackPlanet) requesting a name but not requiring it. Of the sites which do not require a name, Xanga, LiveJournal and Skyrock all began as blogging services and have since transformed into social networking services, indicating that pseudonymity may be more desirable for blogging services than for general-purpose social networks.

In addition to the 6 sites for which a name is optional and a pseudonym is the main identifier on the site, 7 more sites require a pseudonym or username for the site. This does not provide much privacy however as names are still displayed on all of these sites. From the non-general-purpose sites, the gaming websites, 2 media-sharing sites and ExperienceProject were strongly pseudonymous, not collecting names at all.

The utility of pseudonyms on social networks is controversial. One study reported that an excess of fake profiles was a contributing factor in Friendster losing popularity [27], while others found that many youth desire the ability to sign up under pseudonyms [30, 85].

---

[6] Six of the sites requesting a user's data of birth provided a check-box to hide the visibility of the date of birth on the form in which it was requested.

[7] Of course, this is never strongly verified, and there is anecdotal evidence of fake names commonly being provided [23].

8.4.4.3 Requirement of Email Addresses

It is also notable that every site required an email address to join, including the privacy-specific sites. Most of the general-purpose sites (26 out of 29) further require email verification in order to use the site, with only Hyves, meinVZ, and MyYearbook not verifying email addresses. Requiring a valid email address could be seen as an anti-spam technique, although 25 of the general-purpose sites already require their own CAPTCHA to sign up. Although it is easy to obtain free and disposable email addresses online, most users will enter their real email-address, making the insistence on email addresses a needless privacy violation since they are not necessary for interaction with a social networking site. [8]

Almost half of the sites requested the password to one's email address as well, in order to automatically retrieve a person's friends from their email provider. A typical interface is shown in Figure 8.6. In addition to this feature, 4 sites offer an "invite friends" feature which will send invitations to join the network to every email address listed in the user's webmail account. On top of generating spam, these features are poor user training for phishing, as they reinforce the habit of entering passwords into third-party websites.

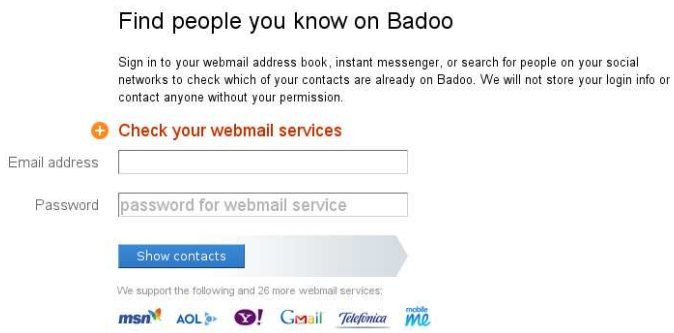


Fig. 8.6 Interface to import address book from external webmail account (Badoo).

8.4.5 Privacy Controls

After signing up, we examined the privacy controls offered by each site. While almost every site has invented its own unique terminology to describe access control, we were generally able to map these into categories which were common across sites. One limitation of our approach is that we did not verify the correct functioning

[8] Email addresses are used as login names on most sites, but this could be easily changed.

of the privacy controls, which would require creating multiple test accounts with each site and examining the visibility of one profile from another under different settings.

8.4.5.1 Profile Visibility Options

The fundamental privacy setting is the visibility of one's profile page, which is the main display of one's personal information on the site. Options provided were profiles accessible to the public internet, profiles only accessible to site members, limitations by geographical or other sub-networks, and limits to friends only or friends of friends. The availability of these levels of control is displayed in Table 8.3. Only 3 sites provided no choice on profile visibility, with Skyrock making all profiles internet-public by default, and Yonja and Multiply making profiles viewable by all members.

It is important to point out that limiting profile views to only members of the site provides very little privacy, since membership is free and easy to obtain for every site surveyed. This distinction is really only useful for privacy in that search engines will not crawl the pages from sites with members-only privacy settings. Sites probably choose to limit visibility to members only in order to force visitors to sign up to be able to view people's profiles. Facebook takes an interesting hybrid strategy, showing a limited "public listing" of profiles to visitors and search engines, using this to encourage membership.

Table 8.3 Visibility of profile data amongst general-purpose sites, $N = 29$: Most sites make profiles publicly visible by default.

visibility level	default	optional	unavailable
public Internet	41%	-	59%
all site users	48%	28%	24%
sub-networks only	7%	17%	76%
friends of friends	-	24%	76%
friends only	3%	79%	17%

8.4.5.2 Fine-Grained Controls

Many sites offer more fine-grained control of profile visibility, with 13 general-purpose sites offering a line-item setting where individual data items may have different visibility, argued to be a crucial feature for privacy management [77]. An average of 10 different profile items were configurable, with Windows Live Spaces offering the most at 27. Of these, only Facebook and LinkedIn offered the useful "audience view" feature, allowing users to see how their profile looks to different groups of users [57].

We found 8 sites which implemented a version of role-based access control by giving users the ability to segregate their friends into abstract, self-defined groups and make access control decisions at the group level. Of these 8, only 2, PerfSpot and Plaxo, made friend grouping mandatory, as has been shown to greatly enhance users' ability to control their privacy [67]

Table 8.4 Access controls for additional features, general-purpose networks, $N = 29$.

feature	separate ACL	profile ACL	no ACL
profile commenting	62%	21%	17%
messaging	52%	28%	21%
photo viewing	52%	41%	7%

Other common privacy controls regulate photo access and the ability to send messages and post public "comments" to other users on the site. Access control offerings for these features are shown in Table 8.4. Most sites offered the ability to restrict these features separately, only Skyrock and Badoo, which operate with all profiles being completely open, did not provide the ability to limit visibility of photos.

8.4.5.3 Permissive Defaults

The main problem observed, however, was not lack of options but the almost universality of open defaults. Estimates have varied in previous literature, and depend on the site in question, but between 80 and 99% of users are typically found to never change their privacy settings [8,52,54]. For more obscure privacy-violating features such as those described in Table 8.5, fewer than 1% of users are thought to opt-out [21,22]. A significant number of users are not even aware that privacy controls exist in social networks, estimated in two different studies at 26% [48] and 30% [8].

As seen in Table 8.3, all but 3 of the general-purpose sites (90%) leave new profiles initialised to be completely visible to at least all other members of the site by default. Of these, Friendster's default limitation to a user's continent and Facebook's limitation to a user's sub-networks provide relatively little privacy gain, since anybody can create a profile in these networks. Only Bebo defaulted users to a friends-only view among the general-purpose sites. Similarly poor default privacy was found in the niche sites, with only the child-specific site Imbee using friends-only privacy by default (and in fact as the only option).

Often, default privacy settings left unnecessarily detailed data traces available to other users. The publication of a stream of user events, such as "Upton uploaded a new photo" or "Upton changed his relationship status," and of the user's online status can be aggregated into temporal usage patterns with serious privacy implications. A network user determined to monitor other users' behavior may often benefit from demographic search capabilities to spot interesting surveillance targets, since most sites enable user discoverability beyond the name. Search was implemented

on all of the sites; only two general-purpose sites (Eons and Badoo) forced users to manually opt-in for the profiles to be indexed. Finally, bilateral profile viewing notifications constitute a privacy dilemma in that enabling them per default unveils the casual stalker but constitutes a hurdle for inconspicuous network browsing. Table 8.5 summarizes the proportion of sites requiring opt-out instead of opt-in for these privacy-invasive services.

Table 8.5 Most general-purpose sites have privacy-invasive discoverability features enabled by default and require manual opt-out from the user, $N = 29$.

feature	implemented	opt-out	% opt-out
user event stream	14	11	79%
online status visibility	25	22	88%
profile viewing notification	16	12	75%
profile searchability	29	27	93%

8.4.5.4 User Interface Problems

In addition to the problem of permissive default settings, we observed many possible user interface problems which could limit the ability of users to effectively use the available privacy controls. This was reflected by a survey which found that 24% of Facebook users did not understand the implications of their own privacy settings [8]. There is also anecdotal evidence from the web that users are confused about privacy settings, such as a guide to configuring one's privacy settings for social networks which was downloaded over 500,000 times [66]

Many sites presented controls in an excessively complex way, although academic studies have found that providing users too much information and too many configuration options can harm usability [87]. Facebook had the most complex settings, with 61 options to select spread across 7 different privacy settings pages. LinkedIn also stood out with 52 settings and 18 pages. Windows Live Spaces suffered from particularly poor usability. For each of its 27 settings, the user most load a new page to examine the value of the setting, load a second page to edit the setting, and then click "SAVE" and submit a form to record the new setting. The average general-purpose site offered 19.2 privacy settings on 3.7 separate pages (median 16 / 2). Users also face an often overwhelming array of choices for controlling the amount of email received from the site, with an average of 13.0 email settings available, with only Nexopia, SkyRock, and Yonja not allowing users to control the amount of email received.

In addition to the complexity observed, we found many cases of confusing settings, ambiguous wording, and inconsistent use of terminology between sections of the same site's privacy settings. Orkut provides a telling example in Figure 8.7. The check box marked "enable photo tagging" actually relates only to the ability of others to tag photos, and also controls the ability to view a list of a user's tagged photos

even if that user tagged the photos personally. The first sentence also includes a confusing dangling modifier; it is not clear if the phrase "with their friends" refers to who is being tagged or who is doing the tagging. Badoo provided another confusing example, offering the choice between making one's profile visible to "any users" or "only members." It is assumed that "any users" can include non-registered-members, though after selecting the "only members" setting it was displayed as "any members." Only 6 sites offered written help in managing privacy settings, exacerbating the problem of confusing terminology and labeling.

Fig. 8.7 Coarse-grained privacy setting with potentially confusing wording and non-standard input controls ("☑ yes")(Orkut).

Fig. 8.8 Pre-selected combinations of privacy settings (Sonico).

A very nice but rare feature was pre-set combinations of privacy settings which could be selected with one click. This was offered by Sonico, offering basic "Public," "Private," and "Custom" settings (Figure 8.8), and NetLog which offered "Meet new people" and "Keep in touch with my friends" settings, each with an additional "high privacy" option. MySpace also offered pre-set combinations of settings, but only to control who on the site is allowed to message a user.

8.4.6 Security Measures

8.4.6.1 Use of TLS Encryption and Authentication

We found an appallingly low adoption of the standard Transport Layer Security (TLS, formerly SSL) protocol. 17 of the 29 general-purpose sites failed to use TLS during log-in, and of the 12 using it, only 3 wrapped the entire log-in page in a TLS connection. The other 9 only used TLS for the HTTP `POST` action, which is undesirable because it prevents browsers' TLS indicators from being displayed, making users more susceptible to phishing. TLS adoption was slightly better in the other sites surveyed, with 6 of the 16 using TLS for the entire login page, and 2 for the `POST` action only.

A common error observed even among sites using TLS for login was forgetting to use TLS during the signup process, when passwords are also entered. 6 sites which used TLS during login did not use it at all during signup, with 2 sites making the opposite mistake. Both mistakes are a sign of careless implementation, as the

sites clearly have the ability to deploy TLS but forget that there are two common situations where passwords are entered. Plaxo provided a particularly bizarre example of TLS inconsistency, using TLS to protect the requested email password for its "retrieve friends" feature but failing to protect the password entered as part of the signup data itself. Overall, 21 of the general-purpose sites and 9 other sites used no TLS during signup.

Disappointingly, only one website surveyed, the business-network XING, provided TLS for all interaction with the site. Curiously, despite this strong security practice, XING was not one of the sites which promoted itself on the basis of privacy. In fact, of the 13 sites which did promote themselves based on privacy, 7 employed no TLS whatsoever, and only 2 provided TLS for their complete log-in pages. [9]

8.4.6.2 Phishing Prevention

There was a glaring lack of attention paid to phishing in the sites surveyed. Not a single site used any anti-phishing mechanisms during login, such as personalized images displayed next to password prompts. Only two websites surveyed (MySpace and BlackPlanet) made any mention of phishing in warning users only to enter their password at their site. Every single site sent us emails containing links requesting us to log-in to the site, easy for phishers to replicate fraudulently.

Coupled with the poor use of full-page TLS for log-in described in Section 8.4.6.1 and the common practice of requesting passwords for external email accounts described in Section 8.4.4, this represents an industry-wide disregard for the problem, though it has been made a point of government policy emphasis [17]. Academic research demonstrated years ago the power of "social phishing" using compromised account due to the social trust inherent in communication on social networks [46]. There is also empirical evidence that phishing is commonplace in large social networks [13, 17], and that phishers are now using stolen social network accounts to request money from unsuspecting online "friends" [39].

8.4.6.3 Online Safety Guidance & Abuse Reporting

Preventing abuse is another important challenge for social networks, as research has suggested cyber-bullying by peers is a significant threat [17], and the majority of young users report being harassed by another user to the extent that they blocked them [30]. Encouragingly, we observed widespread deployment of three mechanisms for preventing cyber-bullying: the ability to block access by specific users, the ability to report specific user profiles directly from the profile page, and web forms for reporting abuse. Every site implemented at least one of the three options, including at least one interface for reporting abuse, with the exception of Plaxo.

[9] We suggest that comprehensive TLS encryption might be used as a promotion technique for evading traffic logging schemes deployed in the European Union.

However, in many cases the abuse reporting web form required clicking on several links to reach. Habbo made the bizarre choice to require completing a CAPTCHA before submitting an abuse report.[10] 10 general-purpose sites failed to implement the much more user friendly "Report User" ability on each profile page. Only one site, PerfSpot, provided a telephone hotline to speak with a representative.

11 general-purpose sites also provide help pages for maintaining online safety, with 9 providing specific help pages for parents. More sites could easily provide such pages, since many of the pages had very little unique content and mostly contained links to the plethora of non-profit online safety pages available on the web [2]. Only 6 general-purpose sites provided help pages for managing privacy. Again, there was a lack of correlation with sites promoting their privacy and providing privacy settings help, with only 1 site, Multiply, doing both.

8.4.7 Privacy Policies

Besides being a legally binding contract between the social network operator and its users, the privacy policy is the only primary source that a prospective user can rely on to give informed consent for data collection, as is required in the EU. Therefore, it is critical that sites post documents which are accessible both technically and linguistically. The results of our inspection of the privacy policies are summarized in Table 8.6. Two sites, SkyRock and Impulse, failed to provide a privacy policy separate from their Terms of Use. We analyzed SkyRock's Terms of Use section on data protection practices since it was clearly labeled "Protection of Users' Privacy and Personal Data". We were unable to count Impulse's one-line statement on users' privacy[11] as an actual privacy policy. For completeness, we still report the analysis results of this statement as "Impulse (T&C)" in Table 8.6.

The quality of a privacy policy is not to be confused with the quality of data protection standards the site implements. Rather, as an enabler for informed consent, a policy should give a good account of the practices regardless of whether these are beneficial or detrimental for a user. As such, a site that honestly and clearly states horrific data collection, usage, and sharing has a better policy than a site with nebulously-phrased text that does not mention data sharing with third parties.

8.4.7.1 Technical Accessibility

It is critical for privacy policies to be accessible to a variety of web browsers and devices to avoid disenfranchising any users. As social networks grow, adherence to good accessibility principles is increasingly important to enable use from mobile

[10] In fact, Habbo utilized a more difficult CAPTCHA for reporting abuse than for signing up.

[11] Impulse's complete statement on privacy: "guarantees not tot [sic] share users' personal information with third parties (except for the cases provided by the law) and not to use it for any other purposes except those of the site;"

Table 8.6 Privacy Policy evaluation results. Fields are left blank where an evaluation criterion was inapplicable or if the site did not specify the information required for evaluation. Cells marked 'u' indicate implementation, but with errors, for the P3P policies and only partial data erasability for the criterion "PP user can delete data".

site	PP present	PP new window	PP requires JavaScript	P3P full	P3P compact	PP dated	PP word count	PP mobileOK	PP zoomable	PP durable Url	PP printable	PP savable	PP textually structured	PP contains operator email address	PP contains operator postal address	PP contains seal	PP external dispute resolution	PP Safe Harbor participant	PP specifies national laws	PP specifies data locations	PP specifies data retention period	PP IP address collected	PP browser data collected	PP external data collected	PP shares with third parties	PP data anonymised for third parties	PP shares with search engines	PP shares with law enforcement	PP has third-party advertisers	PP user can delete data	PP user notified of changes	PP changes take effect delayed	TC minimum age
Badoo	y	n	n	y	n	n	1713	100	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	y		y		y	y		18
Bahu	y	n	n	n	n	n	266	91	y	y	y	y	y	y	n	n	n	n	n	n	n				n					y	n	n	13
Bebo	y	n	n	n	n	y	2842	42	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y	y		y	y	y	y	n	13
BlackPlanet	y	y	n	n	n	y	2888	100	y	y	y	y	y	y	y	n	n	n	y	n	n	y	y	y	y	n		y	y	u	y	n	14
BuzzNet	y	n	n	n	n	y	1781	86	n	y	y	n	y	y	y	n	n	n	n	y	n	y	y		y	y	y	y	y	y	n	n	13
Classmates.com	y	y	y	y	y	y	4934	48	y	y	y	y	y	y	y	y	y	y	n	y	y	n	y	y	y	n			y	y	u	y	y	13
CouchSurfing	y	n	n	n	n	n	1211	53	y	y	y	y	y	n	n	n	n	n	n	n	n	n	y		y	n				y	y	n	18
CyWorld	y	y	n	n	n	y	1870	54	y	y	y	n	y	y	n	n	n	n	n	y	n	y	y		y	y		y	y	u	n	y	13
Eons	y	y	n	n	y	y	1814	80	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	y		y	y		y	n	13
Experience Project	y	n	n	n	n	y	1502	91	y	y	y	y	y	y	n	n	n	n	n	n	n	n	y	y		y	n		y		y	y	n	13
Facebook	y	y	n	n	u	y	3765	56	y	y	y	y	y	n	y	y	y	n	y	y	n	y	y	y	y	n	y	y	y	y	y	y		
Flixster	y	n	n	y	n	y	2254	65	y	y	y	y	y	n	y	n	n	n	n	y	n	y	y		y	n		y	y	y	y	n	13
Friendster	y	y	y	n	n	y	1973	83	y	y	y	y	y	y	y	n	n	y	y	y	n	y	y		y	y		y	y		n	n	
Gaia Online	y	n	n	n	n	y	2249	63	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	y		y	y	u	y	n	13
Habbo	y	y	n	n	y	y	4166	47	n	y	y	y	y	y	y	n	n	n	y	y	n	y	y	y	y	n		y	y	y	n	n	13
hi5	y	n	n	n	n	y	2193	87	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	n		y	y	y	y	n	
Hyves	y	y	n	u	n	y	1706	38	y	y	y	y	y	n	n	n	n	n	n	y	n	y	y		y	n		y			n	n	0
Imbee	y	n	n	n	n	y	2240		y	y	y	y	y	y	y	n	n	n	n	y	n	n	y					y		y	n	n	
Imeem	y	n	n	n	n	y	1887	46	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y		y	y		y	y	y	y	n	13
Impulse	n			n	n																												13
Kaioo	y	n	y	n	n	n	1418	57	y	n	y	n	y	y	n	n	n	n	y	n	n				n			y	n	u	n	n	14
Last.fm	y	y	n	n	n	y	4374	37	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y	n	y	n		y	y		n	n	13
LinkedIn	y	y	n	n	y	y	4957	58	y	y	y	y	y	y	y	y	y	y	y	y	n	y	y			y	y	y	y	u	y	n	18
LiveJournal	y	n	n	n	n	y	2655	69	y	y	y	y	y	y	n	n	n	n	y	n	n	y	y	y	y	n		y	y	u	y		13
meinVZ	y	y	n	n	n	y	8455	52	y	y	y	y	y	y	y	n	n	n	y	y	y	y	y		n			y	y	y			18
MocoSpace	y	n	n	n	n	n	1344	71	y	y	y	y	y	y	n	n	n	n	n	n	n	y	y		y	y		y	y	y	y	n	13
Multiply	y	n	n	n	n	n	2142	65	y	y	y	y	y	n	n	n	n	n	y	n	n	y	y		y	y		y	y	y	n	n	13
MyLife	y	n	n	y	y	n	4083	57	y	y	y	y	y	n	y	y	y	n	n	n	n	y	y	y	y	n	y	y	y	y	y	n	13
MySpace	y	y	n	n	n	y	2738	45	y	y	y	n	y	y	y	n	n	n	n	y	n	y	y		y	n		y	y		n	n	13
MyYearbook	y	n	n	n	n	y	955	57	y	y	y	y	y	n	y	n	n	n	n	n	n	y	y	y	y	n		y	y			n	13
NetLog	y	y	n	n	n	n	311	94	y	y	y	y	n	y	y	n	n	n	y	y	n							y		u			
Nexopia	y	n	n	n	n	y	2752	39	n	y	y	y	y	y	y	n	n	n	n	y	n	y	y		y	y		y	n	u	y	n	13
Ning	y	n	n	n	u	y	4135	79	y	y	y	y	y	y	y	n	n	n	n	y	n	y	y	y	n			y	n	u	n	n	13
Orkut	y	n	n	u	n	y	3073	80	y	y	y	y	y	n	y	n	n	y	y	y	n	y	y	y	y	y		y	y	y	y	n	
PerfSpot	y	n	n	n	n	y	2108	20	y	y	y	y	y	y	y	n	n	n	y	n	n	y	y		y	n		y	y	n	n	n	13
Plaxo	y	n	n	n	n	y	4271	51	y	y	y	y	y	y	y	y	y	n	n	y	y	y	y		y	y		y		y	y	y	13
SkyRock	n	n	n	n	y	n	641	69	y	y	n	y	y	n	y	n	n	n	y	y	n	y	y					y		y			
Sonico	y	n	n	n	n	n	923	73	y	y	y	y	y	y	n	n	n	n	n	n	n				n					u			
Tagged	y	y	y	n	n	y	2799	68	y	y	y	y	y	y	y	n	n	n	y	y	n	y	y		y	n		y	y	u	n	n	13
Twitter	y	n	n	n	n	y	1535	67	y	y	y	y	y	y	n	n	n	n	y	y	n	y	y		y	n		y		y	y	n	13
Viadeo	y	n	n	y	y	n	2848	18	n	y	y	y	y	n	y	n	n	n	y	y	n	n	y	y	n					y	n	n	18
Windows Live Spaces	y	y	n	n	u	y	4361		y	y	y	y	y	n	y	y	n	y	y	y	n	y	y	y	y	y		y	y		y	n	
Xanga	y	n	n	n	n	n	4948	55	y	y	y	y	y	n	n	n	n	n	n	y	n	y	y		y	y	y	y	y	u	y	n	13
XING	y	n	n	n	n	y	3237	58	y	y	y	y	y	n	n	n	n	n	y	n	n	y	y		n		y				y	n	17
Yonja	y	n	n	n	n	n	1557	56	y	y	y	y	y	n	n	n	n	n	n	n	n	n	y	y	n			y	y	u	y	y	18
Impulse (T&C)	u	n	n	n	n	n	32	67	y	y	y	y	y	n	n	n	n	n	n	n	n	n			n			y			n	n	13

devices and by elderly and disabled individuals who may have special accessibility needs [81].

Despite this, we noticed numerous accessibility problems. 15 sites opened their privacy policies in a new window, which can be blocked by pop-up blocking browsers or unsupported by mobile devices. 4 sites required JavaScript to display the privacy policy, which is incompatible with older browsers or some mobile devices. 4 sites deployed privacy policies which did not allow zooming, 4 sites deployed policies which could not be saved, and 1 site (SkyRock) had a policy which could not

be printed. These errors were not committed by the same few sites, 21 sites made at least one such accessibility error.

We also verified accessibility for mobile devices using the W3C mobileOK Checker [84], which checks a Web page against a defined set of recommended guidelines derived from best practices for the mobile Web and issues scores between 0 and 100. This is a rigorous test which is also a good indicator of accessibility in general. Only 2 sites, Badoo and BlackPlanet, received a perfect score. Even MocoSpace, targeted specifically at mobile devices, had numerous problems and received a score of just 71.

8.4.7.2 Length

Given the diversity of written privacy policies and the lack of a standardized vocabulary, we recorded the textual length only in place of a subjective measure of readability.[12] Only 10% of users claim in surveys to have read the privacy policy of their social-networking site [48], and examinations of server logs indicate the actual rate may be far less than 1% [86].

Privacy policies in general were too long to be expected to be read by most users, although the length varied greatly. The mean length was 2,633 words, with a median of 2,245, and a very large standard deviation of 1,546 words. The three shortest policies were all translated from originally French-language sites, the 266, 311, and 641 word policies of Bahu, NetLog, and SkyRock, respectively. The longest policy was the 8,455 word epic from meinVZ, nearly 3,500 words longer than the next longest, that of LinkedIn. There were 12 policies longer than 3,000 words, which are all far too long to provide usable privacy information.

8.4.7.3 Legal Issues

Due to the nature of privacy policies as legal contracts, it is critical for them to provide some basic contractual information. Nevertheless, 13 sites failed to provide a date on their privacy policies, 15 sites didn't list a physical contact address, 17 sites didn't provide an official email address, and 7 sites provided no contact info at all.

21 sites reserved the right to change the terms without notice, making them of questionable contractual value. Only 5 sites guaranteed a minimum notice period before changes could take effect.

Finally, there were problems with specifying legal jurisdiction, especially pressing given the noted discrepancy between the geographic location of operators' headquarters and their targeted regional markets (Section 8.4.1.3). 20 sites did not specify which nation's data protection laws they followed, and 20 sites did not specify in which nations data would be stored and processed in. Only 17 sites specified both,

[12] Subjectively, we generally found readability to be poor.

which would be required information in the case of a dispute. The EU Safe Harbour Agreement, designed to enable compliance with the EU Data Protection Directive for foreign companies with EU customers, was only acknowledged by 6 sites, despite the prevalence of this geographic pattern. 6 sites specifically named an external party to arbitrate disputes arising from the privacy policy.

8.4.7.4 Data Claims

Regarding the actual claims that were made in the policies, there was significant variation, but a pattern emerged of few meaningful rights being assigned to users and operators reserving many data collection and sharing rights for themselves. In addition to user-uploaded profile data, 40 sites specifically reserved the right to record IP addresses and/or browser data. No sites promised not to collect such data, the other 5 sites left the issue unspecified. 14 sites also reserved the collect user data from external sources. Most sites were unclear on this point, with only Last.fm promising not to do so. Few data retention guarantees were made, with only Bebo, meinVZ, and Plaxo providing specific limits on how long they could retain user data. 21 sites did explicitly grant users the right to have their data deleted upon request, as is legally required in the EU, with 24 sites either providing an incomplete guarantee or leaving the point in question.

Operators also often reserved many rights to share user data. 32 explicitly reserved the right to share with third parties, while only 8 promised not to. Of the 32, 17 promised to anonymize user data (although academic research has proven this is impossible for realistic social graph data [15, 65]). 39 sites indicated they would share data with law enforcement when required to do so, with 6 failing to mention this.

8.4.7.5 Availability of P3P Policies

We evaluated the adoption of policies conforming to the W3C's P3P format [4], designed to enable users to quickly determine if a site's privacy practices are acceptable given the user's privacy preferences [73]. P3P has been argued to be a critical element in enabling privacy protection in the future [11], and has been shown to strongly influence user decision-making when its display is mandatory [40].

We saw low adoption of P3P among sites surveyed, with only 7 sites implementing a full P3P policy, 5 of which parsed correctly. Badoo and Hyves were the only general-purpose sites with correctly implemented policies. 10 sites implemented a compact policy, 7 correctly, including just SkyRock and Eons among the general-purpose sites. The lack of P3P adoption and the existence of incorrectly written policies indicates a negative attitude toward the P3P project by some site operators. As shown in Fig. 8.9, Facebook's P3P compact policy provided a vivid example,

consisting of an incorrect policy element name "HONK." This seems crafted specifically to mock users with P3P-displaying browsers.[13]

Fig. 8.9 P3P compact policy file validation errors (Facebook).

8.4.7.6 Self-Promotion within Privacy Policies

Despite the poor observed quality of privacy policies, an interesting trend was that many sites included promotional claims about their privacy practices within the privacy policies themselves. Some typical examples are shown in Figure 8.10. We recorded the use of such reassuring but legally meaningless phrases in privacy policies. Typically, these are written in simple English and make strong claims that privacy is an important consideration within the site. Overall, we observed this tactic in 34 of the 45 sites studied, with 21 of 29 general-purpose sites and 13 of 16 other sites making such claims. We also observed 7 sites displaying a graphical privacy seal next to their privacy policy, despite none of them using the seal on their main signup page to convey the quality of the privacy policy, as the seals are intended. In Section 8.5.5, we report the lack of correlation between these privacy claims and good privacy practices.

> *At Badoo your privacy is of paramount importance. As the custodians of your personal information, we have developed this policy to ensure that your privacy is always protected while you are using the Badoo network.* —Badoo
>
> *Hyves consists of a network of friends. We deal with your information as you would expect from friends. So Hyves takes your privacy very seriously and will deal with your information with due care.* —Hyves
>
> *We have a pretty simple privacy policy. We are reasonably sure this won't annoy anyone.* —Last.fm
>
> *It is Buzznet's policy to respect the privacy of Members.* —Buzznet

Fig. 8.10 Examples of self-promotion within privacy policies.

[13] Indeed, this is a vulgar word in German, making it particularly insulting to a substantial portion of Facebook's users.

8.5 Data Analysis

Viewing our data as a whole, we wish to infer which factors are correlated with good privacy practices in social networking sites. This is a complicated question because it is difficult to exactly answer what constitutes "good practice." For example, an increase in privacy controls available may be seen as good to a certain point, but usability problems may arise from overly complex privacy setting pages [87].

Despite these difficulties, we defined and computed a synthetic privacy score. The formulae are explained in full on our project website [14]. We will use this privacy score to make broad inferences about a site's privacy practices. This privacy score included three subscores summarizing a site's data collection practices, privacy control interface, and privacy policy. We deducted points for unnecessary data collection, awarded points for privacy-enabling features and also for accessibility and usability of the privacy policy.

To compare the privacy practices of a site with the site's overall functionality, we defined an additional functionality score which awarded points for the number of non-privacy features implemented by a site. This score awarded points for providing features such as photo uploading and tagging, profile commenting, event streaming, and support for third-party applications.

The privacy and functionality scores for each site are shown in Table 8.7. Examining the overall privacy score, we found Bebo, LinkedIn, and GaiaOnline to have the overall best privacy practices, while Badoo, CouchSurfing, and myLife scored the lowest. Using our functionality score, we found Facebook, MySpace, and Windows Live Spaces to be the most feature-rich sites, while Twitter implemented the fewest features.

8.5.1 Privacy vs. Functionality

We found only a non-significant positive relationship between the functionality score and privacy score (Fig. 8.11, left). However, there is a pronounced relationship between a site's general functionality and its privacy-specific functionality. A correlation between the functionality score and the privacy control score yields a positive regression coefficient of $r = 0.50$ at $p = 0.0003, N = 45$ as determined by a t-test. Sites that provide more functionality in general also offer more advanced features and support for configuring data sharing. Yet, this is in fact an inherited effect since general-purpose sites, which provide better privacy controls (Fig. 8.11, right) have a significantly higher functionality score than niche sites ($p = 0.01$).

[14] *http://preibusch.de/publ/privacy_jungle/*

Table 8.7 Privacy and Functionality Scores. In this table, the Data Collection Score is inverted and normalized to span $[0, 1]$.

Site	1 – Data Collection Score	Privacy Control Score	Privacy Policy Score	Privacy Score	Functionality Score
Badoo	.33	.07	.33	**.23**	.40
Bahu	.24	.22	.43	**.35**	.50
Bebo	.62	.44	.57	**.70**	.60
BlackPlanet	.29	.26	.54	**.46**	.50
BuzzNet	.29	.22	.43	**.37**	.60
Classmates.com	.33	.22	.63	**.51**	.30
CouchSurfing	.14	.30	.26	**.26**	.30
CyWorld	.14	.47	.50	**.51**	.50
Eons	.24	.36	.48	**.46**	.50
Experience Project	.81	.19	.30	**.44**	.30
Facebook	.10	.61	.41	**.53**	.90
Flixster	.33	.26	.48	**.44**	.40
Friendster	.29	.30	.48	**.44**	.60
Gaia Online	.81	.44	.46	**.69**	.30
Habbo	.81	.37	.48	**.66**	.50
hi5	.43	.32	.43	**.48**	.70
Hyves	.29	.41	.41	**.47**	.70
Imbee	.05	.37	.57	**.46**	.30
Imeem	.71	.15	.57	**.55**	.50
Impulse	.43	.34	.13	**.30**	.30
Kaioo	.57	.15	.46	**.43**	.20
Last.fm	1.00	.22	.48	**.64**	.40
LinkedIn	.52	.39	.67	**.70**	.50
LiveJournal	.48	.60	.37	**.62**	.50
meinVZ	.38	.41	.65	**.65**	.40
MocoSpace	.52	.30	.43	**.49**	.30
Multiply	.05	.36	.39	**.34**	.40
MyLife	.29	.07	.43	**.28**	.30
MySpace	.29	.41	.43	**.48**	.80
MyYearbook	.24	.44	.17	**.33**	.70
NetLog	.52	.30	.35	**.44**	.60
Nexopia	.33	.22	.46	**.40**	.30
Ning	.52	.41	.48	**.59**	.70
Orkut	.43	.35	.46	**.51**	.70
PerfSpot	.19	.63	.48	**.61**	.60
Plaxo	.29	.44	.57	**.58**	.40
SkyRock	.38	.11	.39	**.31**	.40
Sonico	.00	.33	.37	**.30**	.30
Tagged	.24	.22	.35	**.30**	.60
Twitter	.81	.26	.30	**.49**	.10
Viadeo	.43	.15	.50	**.41**	.20
Windows Live Spaces	.33	.47	.50	**.58**	.80
Xanga	.76	.48	.37	**.65**	.50
XING	.24	.37	.57	**.52**	.30
Yonja	.57	.33	.37	**.49**	.40

8.5.2 Privacy vs. Site Age

We find a positive relationship between the age of a site (the time elapsed since it went online) and its privacy score. Sites that have been in existence for a longer time also have a significantly longer privacy policy in terms of word count, which can be explained by a (reactive or pro-active) privacy policy engineering process (Fig. 8.12 right). The lack of (negative) relationship between functionality score and privacy

functionality score	privacy score ≤ avg	> avg
≤ avg	13	9
> avg	9	14
significance	$p = 0.24$	

category	privacy score ≤ avg	> avg	priv. control score ≤ avg	> avg
gen. purpose	14	15	10	19
niche	8	8	11	5
significance	$p = 1.00$		$p = 0.03$	

Fig. 8.11 There is a positive, yet not significant relationship between functionality and privacy as revealed by Fisher's exact test, 2-tailed on the contingency tables between a site's functionality score and its privacy score (left) (data z-transformed and dichotomized by above / below average partition). General-purpose and niche sites cannot be differentiated based on their privacy practices (middle), but general-purpose sites offer more complete privacy settings and better support for configuring them (right). $N = 45$.

score indicates that network operators fail to exploit their users' willingness to give up more privacy when they receive more benefits in return (discussed further in Section 8.6.4).

privacy score	Alexa rank ≤ med	> med	user count ≤ med	> med
≤ avg	15	7	15	7
> avg	7	16	8	15
significance	$p = 0.02$		$p = 0.04$	

age	privacy score ≤ avg	> avg	policy length ≤ avg	> avg
≤ avg	16	10	17	8
> avg	6	13	7	12
significance	$p = 0.07$		$p = 0.07$	

Fig. 8.12 Larger and more popular sites as well as more mature sites have significantly better overall privacy protection and they feature longer privacy policies, as revealed by Fisher's exact test, 2-tailed on the contingency tables, data z-transformed. (Note that a lower rank means more popularity.) The privacy score increasing with age cannot be attributed to one single privacy subscore: there is no significant relationship between a site's age and its data collection, privacy policy or privacy control subscores. $N = 44$ for the privacy policy length, $N = 45$ otherwise.

8.5.3 Privacy vs. Size

Similarly, the resource constraints of the social network operator give an economic explanation for our finding that P3P is implemented more often among larger sites (Fig. 8.13). One can expect that bigger companies can more easily devote resources to deploying P3P policies. Unlike the mere presence of written privacy policies, the implementation of P3P policies is not mandated by law. As such, an operator who has invested in deploying a P3P policy has undertaken measures towards privacy enhancement beyond the required minimum. Similarly, more popular sites (by traffic rank and by user count) have an overall higher privacy score (Fig. 8.12, left).

P3P deployed	user count		Alexa rank	
	≤ average	> average	≤ median	> median
yes	7	7	7	7
no	23	6	15	16
significance	$p = 0.08$		$p = 1.00$	

Fig. 8.13 P3P policies are deployed more often on sites with above average user count ($N = 43$). However, there is no relationship between a site's popularity in terms of Alexa count and its P3P deployment ($N = 45$). p-values by a two-tailed Fisher's exact test.

8.5.4 Privacy vs. Growth Rate

Our sample provides evidence that privacy-enhanced sites have grown ahead of the market lately. The privacy score is positively associated with both the three-month change in traffic rank and the three-month change in page views. Similarly, the privacy control score is positively associated with the change in page views but negatively with the change in traffic rank, with only the latter relationship being significant, though ($p = 0.08$). It is possible that both phenomena may have a common cause such as the market concentrating on big sites with extensive configuration possibilities.

It is noteworthy that sites which promote on privacy are falling behind with respect to those sites which do not promote on privacy. Sites promoting on privacy have a weakly significant below-average traffic rank increase ($p = 0.10$). Implications of this are discussed further in Section 8.6.1.1.

8.5.5 Privacy Promotion and Claims vs. Actual Privacy Practices

A site's privacy claims do not necessarily indicate good privacy practices. We tested for a relationship between the privacy score and its constituent subscores with a site promoting on privacy and vaunting its data protection in the privacy policy. No significant relationship could be found between embellished claims in the privacy policy and actually good practices as captured by the privacy scores. On the contrary, sites that promoted privacy on their signup pages have a below-average privacy score ($p = 0.11$). Still, there is a weak positive relationship between the quality of a privacy policy and the existence of promotional arguments related to data protection ($p = 0.19$).

We conclude that sites mentioning good privacy practice during the signup phase actually have less favorable privacy practices, but they are well communicated in the privacy policy. These results can be interpreted as being similar to the adverse selection effect of privacy seals for general websites [32], or perhaps as the supply side analogy to the discrepancy between stated and actual privacy preferences on the demand side of the social networking market [8].

8.6 Economic Models

The diversity we found in the privacy practices across the sites indicates there are no universal rules for privacy in social networking. The market continues to be fluid and experimental, with some of the variation in privacy practices surely due to irrational decisions by site implementers. However, we have analyzed the data and found it supports several compelling models for why poor privacy may be a rational choice for social network operators. In particular, we propose a novel model which explains our observed data, the privacy communication game. We will then compare this game-theoretic explanatory approach with other economic models traditionally applied to privacy design choices.

8.6.1 The Privacy Communication Game

We propose a novel model to explain the varying levels of privacy-related advertising within a single site, taking into account heterogeneous privacy preferences in the user population, and the temporal dynamics of privacy concerns. We call this model the *privacy communication game*.

In our model, different users have different privacy concerns and the social network's strategy can be seen as an attempt to optimize its interaction with each group. Previous research has provided evidence that Web users can be divided into three groups based on privacy concerns: the *marginally concerned*, the *pragmatic majority*, and the *privacy fundamentalists* [6], a taxonomy originally due to Westin. The predominant group of users, the *pragmatic majority* claims when asked to be interested in privacy but has been shown in previous studies to forget about privacy when given an attractive service [6] or monetary rewards such as discounts [79].

In parallel, it has also been shown that providing more assurance of privacy can actually make non-fundamentalists less comfortable than simply ignoring privacy [58]. However, privacy fundamentalists care deeply about privacy, and may actively investigate a site and complain to non-fundamentalists if they are dissatisfied with a site. A successful site will therefore play a game of minimizing the concerns of the fundamentalists while simultaneously minimizing the awareness of privacy for the non-fundamentalists.

Expressed slightly more formally, the action space for the social network operator in the privacy communication game is {communicate, hide}. There are two categories of user, namely {non-fundamentalist, fundamentalist}. All users must choose between {sign up, cancel}, while the fundamentalists will also choose between {complain, silence}. Non-fundamentalists are inclined towards "sign up" when seeing "hide"; fundamentalists are inclined towards "cancel" and "complain" when seeing "hide" and vice versa when seeing "communicate".

Because the operator is inclined towards opposite strategies for the two groups of users, it can improve its outcomes by filtering the two groups based on observed signals about users' privacy preferences and then discriminating its strategy based

on the user's type. This is in some sense similar to the practice of price discrimination, as the site operator aims to serve both groups of customers in a dedicated way.

A more complex model would account for third parties such as journalists who can more strongly influence the public [61]. Eventually, only privacy negotiations with individualized communication strategies based on non-cloneable signals will enable the service provider to choose individually optimal privacy strategies and to take the corresponding communication actions.

The following subsections derive the network operator's optimal strategy in the privacy communication game and relate it to our empirical evidence.

8.6.1.1 Reducing Privacy Salience

When facing non-fundamentalist users, the goal of the network operator is to encourage not just sign-up but also disclosure of information. Since social networks are more valuable to each user the more of their friends' data is available, operators may seek to create an environment where people feel free to disclose their data, which for non-fundamentalists is best achieved by making minimal reference to privacy.

Talking about privacy, even in the context of underlining the site's positive privacy features, may have negative consequences for the social networking operator because the mere mention of data protection raises concerns amongst the visitors. This phenomenon is known as *privacy salience*, or privacy-priming. Experiments have shown that providing strong privacy assurance can actually make people less likely to disclose personal information than if none were provided [58]. Similarly, a study on P3P browsers found that users exposed to explicit privacy information reported higher privacy concerns afterwards [40]. Many users taking part in a survey about privacy on social networks were found to have restricted their visibility settings after taking the survey [8].

Due to privacy salience effects, even promoting positive privacy practices might actually fan fears and drive customers away or reduce their willingness to reveal personal information. This would have a negative impact on the valuation of the network by its two most important customer groups: users and advertisers. Ceteris paribus, a user of the site will perceive a the network as less useful when the amount of social information for viewing is decreasing—for instance due to users not entering personal information due to privacy concerns. For advertisers, less complete profiles limit the ability for targeted advertising.

This may explain the behavior of not promoting on privacy (Section 8.4.2.4) and minimizing mention of a site's privacy policy during sign-up (Section 8.4.3). Social networks have another powerful tool to decrease privacy salience, which is to showcase other users who have posted photos and other personal behavior, making this behavior seem normal and safe (Section 8.4.2.2). This is corroborated by evidence from previous studies, which suggest that the majority of users can be enticed to

enter more personal data by an animated character requesting it, or by framing the data input as situationally acceptable [58, 79].

Additionally, surfacing privacy concerns can be mitigated proactively by establishing trust with users without mentioning privacy. User studies have found that the quality and professionalism of a site is more effective in establishing trust then the contents of a privacy policy or the existence of privacy indicators [18]. This may explain survey results in the case of MySpace and Facebook, two sites mainly differing by screen design at first site, which find that Facebook is strongly trusted by its users [8, 48] more so than MySpace [31]. In our study, Facebook reached a privacy score of 0.53 compared to MySpace's 0.48, only coming out slightly ahead. The extra trust in Facebook may represent Facebook's cleaner and more consistent layout rather than its privacy practices.

8.6.1.2 Discouraging Privacy Fundamentalists

Fundamentalists make up a small portion of the market (estimated between 17% [6, 26] and 30% [79]), thus their participation may not be crucial for a social network's success, in particular because they are the least likely customers to begin with. Initial growth of a networking site will be created by less privacy-concerned early adopters. Individuals with strong privacy beliefs are significantly less likely to use social networks, as indicated by surveys [8], after they feel compelled to because their friends have already joined [23].

Most importantly, though, they may be less valuable or even have negative value as customers because of their privacy-conscious actions on a site. This has opportunity costs in that fundamentalists will upload less personal information, which is correlated both to having fewer friends on the site and using it less frequently [48, 55]. This makes these users less valuable for targeted advertising (we conjecture they are also likely to click on advertising links). There may also be indirect costs, however, such as the existence of fundamentalists with limited profiles or strict privacy settings raising the privacy salience of non-fundamentalists. Direct costs accrue naturally to the network operator from providing a typically free service.

The undesirability of privacy fundamentalists as social networking users may explain several trends we noticed where sites seem to avoid simple privacy practices that seem relatively cheap. For example, the poor deployment of TLS authentication and encryption (Section 8.4.6.1), the failure to implement P3P (Section 8.4.7.5), and the requirement of real names and gender (Section 8.4.4.2) are all likely to deter privacy fundamentalists, despite these being relatively small changes to make to the site. Similarly, there are often features which are not supported for users with strict privacy settings. Two Facebook users who both make their profiles unsearchable are given no support to become friends on the network [22]. These observations may reflect a rational choice to discourage privacy fundamentalists from joining.

8.6.1.3 Reducing Privacy Criticism

While fundamentalists make up a small enough population that the network may not wish them to join, they may exert power beyond their numbers by complaining to non-fundamentalists, ruining the network's attempt to minimize privacy salience. Indeed, even small, advantageously placed groups can influence the opinion in networks: fundamentalists may in fact be bloggers or journalists who wield a disproportionate influence over other users' opinions of the site [61]. Thus, the network it is strongly inclined to reduce their criticism. Another important class of privacy fundamentalists may be parents, who may not use the service themselves but are afraid of their children's online activities. It has been shown, for example, that people are consistently more afraid of privacy threats to their own children than they are to themselves [8].

As a result, while access to the legally-required privacy policies is veiled from non-fundamentalists, it is in the service provider's own interest to address privacy concerns to fundamentalists who may actually reach the documents and incorporate it into their decision to establish trust with the site [16]. We recall that, in addition to the decision whether to join or not to join, the fundamentalists potentially complain. This could explain the frequency with which operators vaunt their good privacy practices within their privacy policies, while not making such claims elsewhere on the site (Section 8.4.7.6). A particularly interesting manifestation of this strategy are (paid) privacy seals that are embedded in the privacy policy but not posted on the main pages of the site.

Similarly, social networking sites frequently make strong claims about their privacy practices when confronted with critical media attention due to major scandals. For example, in February an American teenager was caught soliciting naked pictures of under-age male Facebook users for blackmail purposes. In a press release responding the story, Facebook's first sentence was "Facebook seeks to provide a safe and trusted environment by offering users industry-leading tools that control access to their information..." [51]. This can be seen as another consequence of the privacy communication game, as Facebook realizes it needs to strongly promote its privacy practices to concerned users reading news articles.

This quote also points to the deployment of overly-complicated privacy settings with open defaults as a rational strategy for reducing privacy complaints while still minimizing salience. We frequently observed open-by default settings (Section 8.4.5.3), which is a good choice because most users will not adjust their privacy settings [8,21,22,35,48,54]. We also observed many cases of privacy controls which we considered too numerous or confusing to be practical (Section 8.4.5.4). Deploying such settings may be optimal because it will prevent non-fundamentalists from managing their privacy, while still giving fundamentalists the control they desire given sufficient effort to understand the interface.

8.6.1.4 Evolution of Communication

Finally, we propose within our privacy discrimination model that a site's optimal strategy may evolve over time as its user base changes. It can be expected that non-fundamentalists will dominate the early adopters of a social network. This has been found by ethnographic studies, as more privacy-concerned individuals report that they only join a social network when they feel compelled to do so after many of their friends have joined [23, 42]. Similarly, privacy fundamentalists, particularly journalists, may be less inclined to complain about newer sites with lower membership, focusing on major players instead. Individual users also reported that their privacy concerns increased over time when using a network [23], suggesting that the user base may inherently drift towards privacy fundamentalism as time passes.

Speculatively, an optimal strategy for a network may, therefore, be to begin with no privacy controls to minimize privacy salience and encourage growth, while slowly implementing privacy features as it ages and the user base complains, or mass media criticizes unfavorable data protection mechanisms. This may explain the common perception of social networks as following a "functionality first" paradigm, which Facebook's CEO acknowledged by stating that "growth is primary" in the industry [92]. We found evidence for this in the strong correlation of improved privacy practices in older networks in our survey (Fig. 8.12).

8.6.2 The Effects of Lock-in

Lock-in is an entrenched feature of the market for social networks, with users facing high-switching costs to create accounts on competitive networks. In addition the cost of learning a new interface, users have been found to invest significant amounts of time in building up their profiles, which is lost if the user changes networks [23, 48]. Previously, it has been argued that lock-in is an endemic problem in security applications which harms the quality of products on the market [59]. The same model may apply to social networking accounts, as lacking data portability or data extraction prevention make it impossible for a user to move his data out and to a new network if it hypothetically offered better privacy.

This theory is supported by our survey, which found very little evidence of portability of profiles between sites. No site which we studied offered any interface for exporting one's profile data, friendship links, or photos in a simple way.

We also found strong evidence that sites attempt to erode their competitors' lock-in advantages by offering to automatically retrieve friends from a user's email inbox, making it easier to get a new account started (Section 8.4.4.3). Smaller social-networking sites could potentially request a user's old account from a competitive site to retrieve profile information, but this is against most sites' terms of use and has already led to at least two lawsuits: Facebook sued startup Power.com in January for allowing users to enter their Facebook login details and then fetching their account data, after similarly suing to shut down Google's FriendConnect service in

May 2008 [14]. This demonstrates that sites are aware of the lock-in they possess and are actively fighting to maintain it.

The OpenSocial project [3] has been started to promote interoperability between sites. Seven of the sites surveyed implement OpenSocial applications, yet only Ning made any mention of this fact and none of the sites implement the project's goal of allowing users to take their profile data between sites. It is telling that sites have embraced OpenSocial to prevent application developers from being locked-in to one site's platform and ensure a large number of applications are available, but have avoided using it to actually allow users to more freely move between sites.

Thus, most users are locked into their current social network, meaning sites are primarily competing for the sign-up of new users. This is particularly problematic for privacy advocates. First, most new users have little data uploaded and thus their privacy is less of a concern, making data protection less of a selling point for a new account. Second, it can be difficult to assess the full spectrum of privacy controls before a significant amount of data is uploaded, thus it is even more difficult for users to asses privacy controls when considering joining. Sociological evidence may support this, as teenagers are infatuated with sharing when they first join a network, before eventually growing frustrated with the "drama" generated by people viewing their social networking pages [23]. Thus, lock-in may explain a lack of motivation for promoting privacy practices or building privacy controls, as users may be significantly locked-in to the network by the time they are concerned about privacy.

The lock-in model may be complementary to the privacy communication game model. Whilst the lock-in model captures the temporal dynamics of privacy preferences of the social network usage life-cycle and thereby explains why offering few privacy controls do not present a hurdle for joining the network, unlike the privacy communication game, lock-in effects do not account for heterogeneous privacy preferences among the user population and cannot fully explain the existing privacy practices.

8.6.3 Privacy as a Lemons Market

The market for privacy in social networks also fits the model of a lemons market well, as has been shown to occur in general for privacy and websites [83]. Because users have so much trouble assessing a site's privacy, sites have less incentive to provide good functionality and the market is dominated by "lemons." As argued previously, the obfuscated language employed by privacy policies deliberately deprives consumers of adequate information about what privacy is truly being offered by a website, preventing sites from needing to compete on privacy. This is consistent with our findings for social-networking privacy policies, which suffered from many usability problems (Section 8.4.7).

It is made stronger by our findings that privacy is not mentioned promotionally (Section 8.4.2.4), P3P—a potential remedy against information asymmetry—is rarely enabled (Section 8.4.7.5), and privacy controls are excessively numerous and

confusing (Section 8.4.5.4). Moreover, we found that promotional privacy claims were inversely correlated with superior privacy practices (Section 8.5.5), meaning users are in fact receiving misinformation. For these reasons, it is difficult for end users to adequately assess the privacy functionality of a social networking site. Indeed, in our evaluations it typically took around one hour just to collect rote data on privacy features offered by each site.

This model integrates with a privacy communications game well. The inability of non-fundamentalist users to distinguish between good and bad privacy further lessens the incentive for sites to promote their privacy, when doing so may raise privacy salience and have adverse effects.

8.6.4 Privacy Negotiations

The paradigm of *privacy negotiations* views a user's choice to use social-networking services as a privacy trade-off, weighing the functional benefits they get from a social networking site against the privacy they have to give up in order to qualify for these benefits [70,82]. A similar optimization can be made to determine if and when to reveal specific data items once signed up or to determine if and when to delete information or leave the platform. There is some evidence that users may rationalize their social network use this way, some survey respondents stated that they consider giving up some personal information to be the price of a useful, free service [30].

Whilst such a utility maximization problem can easily be stated formally, the subjective valuations associated with benefits as well as with privacy costs make a computational solution unrealistic—not withstanding systematic psychological distortions in privacy-related decision-making. In particular, the valuations need to be formed for non-monetary benefits and costs, under limited information, over expected values with small probabilities, and subject to uncontrollable externalities. Even if a user possessed all required information, the cognitive burden and finite resource one is ready to spend would make her use simple heuristics.

Regarding the economics of privacy on social networks, this resort to heuristics has two major implications. First, network operators who assume fully rational behavior of their users may see their expectations over the users' actions unfulfilled. Optimization procedures over privacy designs that assume a *homo economicus* are unlikely to yield successful results in practice. Second, operators may gainfully exploit the users' inability to make fully informed choices. When heuristics are used as decision rules, these can be tricked. An example is hiding bad privacy practices in the fine-print and equipping a privacy policy with a seal instead (Section 8.4.7.6).

In the decision heuristics, users will contrast perceived advantages with perceived disadvantages. The higher the user perceives the functional benefit, the more she is willing to provide information. The entirety of the promotional arguments a site uses to induce sign-up can be interpreted as increasing the perceived benefits in a privacy negotiations settings.

Our data suggest that social network operators do not yet strategically exploit the tradeoff between functionality and data protection as two alternative sources for a user's utility as they compete for users. We found no evidence that sites with more funcionality are able to offer less privacy, our data instead showed a weak trend in the opposite direction (Section 8.5.1). Nor did we observe any evidence that price discrimination with different (privacy, functionality)-bundles is implemented within individual sites. It could be argued that in the context of social networks site functionality is less important than network effects, which grow with the number of relevant peers, i.e. the number of potential contacts. However, sites more attractive by popularity or user count also exhibit a higher privacy score (Fig. 8.12). These trends lead us to generally reject a privacy negotiations paradigm. Still, this observation does not preclude that some users may consciously perform a cost-benefit analysis before joining a site.

8.7 Limitations

In light of the scale and the scope of this study, some limitations should be kept in mind that apply to all phases of our study. First, the selection of sites and criteria to assess them might be improved. We have given account of our sampling criteria in Section 8.3. They are aimed at defining a tractable sample for which an exhaustive evaluation could be performed. While considerable effort has been made to identify all sites that fall into the operational definition, it might be possible that some sites were missed. The sample size is particularly sensitive to cut-off levels defined on user count. Due to the scarcity of resources, expanding our sample would have forced us to compromise on the depth of analysis. The authors have adopted the point of view that—at a later point in time—the sample could be expanded more efficiently in breadth than in depth, henceforth our selection of 45 sites evaluated at approximately 260 criteria each. It might be possible we missed an important evaluation criterion or metadata entry. Our choices were driven by our analysis needs, lessons from past field studies, our expectations regarding discriminatory metrics, and eagerness for conciseness and completeness. We did not attempt to evaluate some more qualitative elements, such as the usability of privacy controls or the readability of privacy policies, relying on very rough indicators like word count and number of settings instead.

Second, the evaluation process needed to be done manually which introduces inevitable human error. Fine-grained evaluation criteria with little room for interpretation and operational definitions including tool support for evaluation (for instance in determining the privacy policy word count) are intended to keep this error small. The evaluation apparatus, as described in Section 8.3 was kept constant as much as possible. The evaluation tasks were split among the authors on a per criteria basis rather than on a per site basis.

Third, the scores we define, the privacy score and its constituting subscores for data collection, the privacy policy, and the privacy controls, as well as the function-

ality score, can be debated. We consider the definitions sound by intuition and we provide statistical backup for the definitions (Cronbach's α). Other scores may be defined at the reader's discretion. In making available the calculation formula (Section 8.5), we enable the reader to assess the suitability of each of these scores for her or his own analyses. Equally, in making the dataset publicly available, we provide the necessary data to define any alternative score.

Fourth, the authors acknowledge that durability of the data is limited given the high mutability of the market in social networking. Even so, the value of the dataset does not only originate in being the most comprehensive snapshot. It can also be used as an historical data point in longitudinal analyses.

Fifth, our analyses and the economic models we advance as explanations for the empirical evidence might be scrutinized. By making our dataset publicly available, we encourage the community to challenge our interpretations and conclusions.

8.8 Conclusions

Online social networking has become an indispensable activity, and research must keep up with the phenomenon. With the mass adoption of social networking sites over the last eighteen months, a scholarly review of privacy practices "in the wild" was overdue. Given our data, we have serious concerns about the current state of affairs.

In particular, we have found strong evidence that the social networking market is failing to provide users with adequate privacy control. The market is still in an early stage of aggressive competition for users that may eventually yield to a more static and consolidated supply. Our results suggest that the naive application of utility maximization theory fails to capture all the intricacies of the market for privacy in social networking. Experimental economics has long suggested that users' privacy-related decision-making is systematically distorted from full rationality and subject to limited information. We have found compelling evidence that a major problem is the lack of accessible information for users, encouraged by sites' strong incentives to limit privacy salience as part of the privacy communication game: the data suggests that sites may have evolved specifically to communicate differently to users with different levels of privacy concern.

Assuming that better privacy awareness and protection would be beneficial for users, regulation may be necessary in order for a privacy market to function properly. Reducing information asymmetry is an important first step, through standardized "privacy nutrition labels" [50] which can communicate privacy practices in a non-textual format to help users make more informed privacy choices. Increasing privacy salience is of critical importance. This could be achieved by requiring sites to provide clear, Web-integrated interfaces for users to see exactly what personal data of theirs is held, and exactly which parties have access to it. User access to data is a core principle of the EU Data Protection Directive, but we argue it must be far easier and more integrated into the user experience to be effective. Finally, reducing

lock-in effects through mandated data portability may be necessary to increase consumer choice in social networks. In this area, regulation seems most promising and may pay off in the short run.

We also think that much more research is necessary on the dynamics of privacy in social networks. Our results hint at many promising areas for further inquiry. The privacy salience phenomenon and its role in social networking in particular needs further analysis. We are planning a user experiment to study privacy-related decisions on social networks, focusing on the role of communication and privacy-functionality trade-offs each user has to solve. Research is also needed on methods to make privacy information more understandable, and better user interfaces for configuring social network access controls. We hope that our study, along with our published dataset, will be an important starting point.

Acknowledgments

The authors would like to thank Alastair Beresford, Jonathan Anderson, and Ross Anderson for their support and feedback.

References

1. Alexa: The Web Information Company (2009)
2. OnGuard Online. www.onguardonline.gov/ (2009)
3. OpenSocial Project. www.opensocial.org (2009)
4. Platform for Privacy Preferences (P3P) Project. http://www.w3.org/P3P/ (2009)
5. Ackerman, M.S.: Privacy in pervasive environments: next generation labeling protocols. Personal Ubiquitous Comput. **8**(6), 430–439 (2004). DOI http://dx.doi.org/10.1007/s00779-004-0305-8
6. Ackerman, M.S., Cranor, L.F., Reagle, J.: Privacy in e-commerce: examining user scenarios and privacy preferences. In: EC '99: Proceedings of the 1st ACM conference on Electronic commerce, pp. 1–8. ACM, New York, NY, USA (1999). DOI http://doi.acm.org/10.1145/336992.336995
7. Acquisti, A.: Privacy in electronic commerce and the economics of immediate gratification. In: EC '04: Proceedings of the 5th ACM conference on Electronic commerce, pp. 21–29. ACM, New York, NY, USA (2004). DOI http://doi.acm.org/10.1145/988772.988777
8. Acquisti, A., Gross, R.: Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. In: Privacy Enhancing Technologies – LNCS 4258, pp. 36–58. Springer Berlin / Heildelberg (2006). DOI {10.1007/11957454_3}
9. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security and Privacy **3**(1), 26–33 (2005). DOI http://dx.doi.org/10.1109/MSP.2005.22
10. Anderson, J., Diaz, C., Bonneau, J., Stajano, F.: Privacy preserving social networking over untrusted networks. Second ACM SIGCOMM Workshop on Online Social Networks (2009)
11. Antón, A.I., Bertino, E., Li, N., Yu, T.: A roadmap for comprehensive online privacy policy management. Commun. ACM **50**(7), 109–116 (2007). DOI http://doi.acm.org/10.1145/1272516.1272522
12. Arrington, M.: Elaborate Facebook Worm Spreading. TechCrunch (2008)
13. Arrington, M.: Phishing For Facebook. TechCrunch (2008)
14. Arrington, M.: Facebook Defends Its Turf, Sues Power.com. TechCrunch (2009). eMarketer

15. Backstrom, L., Dwork, C., Kleinberg, J.: Wherefore Art Thou R3579x?: Anonymized Social networks, Hidden Patterns, and Structural Steganography. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, pp. 181–190. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1242572.1242598
16. Bansal, G., Zahedi, F., Gefen, D.: The moderating influence of privacy concern on the efficacy of privacy assurance mechanisms fo building trust: A multiple context investigation. In: ICIS 2008: International Conference on Information Systems (2008)
17. Barroso, D., Barle, R., Chazerand, P., de Zwart, M., Doumen, J., Gorniak, S., Kaźmierczak, M., Kaskenmaa, M., López, D.B., Martin, A., Naumann, I., Reynolds, R., Richardson, J., Rossow, C., Rywczyoska, A., Thumann, M.: Security and Privacy in Massively-Multiplayer Online Games and Social and Corporate Virtual Worlds. Tech. rep., ENISA - European Network and Information Security Agency (2008)
18. Belanger, F., Hiller, J.S., Smith, W.J.: Trustworthiness in electronic commerce: the role of privacy, security, and site attributes. The Journal of Strategic Information Systems **11**(3-4), 245 – 270 (2002). DOI DOI:10.1016/S0963-8687(02)00018-5. URL `http://www.sciencedirect.com/science/article/B6VG3-475RJF6-1/2/1b644a64d596b015dfdbcb4e32b881ce`
19. Bennett, R.: Plea to ban employers trawling Facebook. The Times (2008). The Times
20. Bonneau, Joseph: New Facebook Photo Hacks (2009). URL `http://www.lightbluetouchpaper.org/2009/02/11/new-facebook-photo-hacks/`
21. Bonneau, Joseph and Anderson, Jonathan and Danezis, George: Prying data out of a social network. In: ASONAM 2009 : Advances in Social Networks Analysis and Mining (2009)
22. Bonneau, Joseph and Anderson, Jonathan and Stajano, Frank and Anderson, Ross: Eight Friends Are Enough: Social Graph Approximation via Public Listings. In: SNS '09: Proceeding of the 2nd ACM Workshop on Social Network Systems (2009)
23. danah boyd: Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life. Youth, Identity, and Digital Media pp. 119–142 (2008)
24. Buchegger, S., Datta, A.: A case for P2P infrastructure for social networks - opportunities and challenges. In: Proceedings of WONS 2009, The Sixth International Conference on Wireless On-demand Network Systems and Services. Snowbird, Utah, USA (2009)
25. Chau, D.H., Pandit, S., Wang, S., Faloutsos, C.: Parallel Crawling for Online Social Networks. In: WWW '07: Proceedings of the 16th international conference on World Wide Web, pp. 1283–1284 (2007)
26. Cranor, Lorrie F., Joseph Reagle, and Mark S. Ackerman: Beyond concern: Understanding net users' attitudes about online privacy. Tech. Rep. TR 99.4.3, AT&T Labs (1999)
27. danah boyd and Nicole Ellison: Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication (2007)
28. Danezis, G., Wittneben, B.: The Economics of Mass Surveillance and the Questionable Value of Anonymous Communications. WEIS: Workshop on the Economics of Information Security (2006)
29. Donath, J. and boyd, d.: Public displays of connection. BT Technology Journal **22**(4), 71–82 (2004). DOI http://dx.doi.org/10.1023/B:BTTJ.0000047585.06264.cc
30. Dwyer, C.: Digital relationships in the "myspace" generation: Results from a qualitative study. In: HICSS '07: Proceedings of the 40th Annual Hawaii International Conference on System Sciences, p. 19. IEEE Computer Society, Washington, DC, USA (2007). DOI http://dx.doi.org/10.1109/HICSS.2007.176
31. Dwyer, C., Hiltz, S.R., Passerini, K.: Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace. In: Proceedings of the Thirteenth Americas Conference on Information Systems (2007)
32. Edelman, B.: Adverse Selection in Online "Trust" Certifications. WEIS: Workshop on the Economics of Information Security (2006)
33. Egelman, S., Tsai, J., Cranor, L.F., Acquisti, A.: Timing is everything?: the effects of timing and placement of online privacy indicators. In: CHI '09: Proceedings of the 27th international conference on Human factors in computing systems, pp. 319–328. ACM, New York, NY, USA (2009). DOI http://doi.acm.org/10.1145/1518701.1518752
34. Felt, A.: Defacing Facebook: A Security Case Study. www.cs.virginia.edu/felt/fbook/facebook-xss.pdf (2007)

35. Felt, A., Evans, D.: Privacy Protection for Social Networking Platforms. Workshop on Web 2.0 Security and Privacy (2008)
36. Felt, A., Hooimeijer, P., Evans, D., Weimer, W.: Talking to strangers without taking their candy: isolating proxied content. In: SocialNets '08: Proceedings of the 1st workshop on Social network systems, pp. 25–30. ACM, New York, NY, USA (2008). DOI http://doi.acm.org/10.1145/1435497.1435502
37. Finder, A.: For Some, Online Persona Undermines a Resume. The New York Times (2006)
38. Frankowski, Dan and Cosley, Dan and Sen, Shilad and Terveen, Loren and Riedl, John: You are what you say: privacy risks of public mentions. In: SIGIR '06: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, pp. 565–572. ACM, New York, NY, USA (2006). DOI http://doi.acm.org/10.1145/1148170.1148267
39. Frommer, D.: What a Nigerian Facebook Scam Looks Like. The Business Insider (2009). URL `http://www.businessinsider.com/2009/1/nigerian-scammers-still-roosting-on-facebook`
40. Gideon, J., Cranor, L., Egelman, S., Acquisti, A.: Power strips, prophylactics, and privacy, oh my! In: SOUPS '06: Proceedings of the second symposium on Usable privacy and security, pp. 133–144. ACM, New York, NY, USA (2006). DOI http://doi.acm.org/10.1145/1143120.1143137
41. Gjoka, M., Sirivianos, M., Markopoulou, A., Yang, X.: Poking facebook: characterization of osn applications. In: WOSP '08: Proceedings of the first workshop on Online social networks, pp. 31–36. ACM, New York, NY, USA (2008). DOI http://doi.acm.org/10.1145/1397735.1397743
42. Govani, T., Pashley, H.: Student awareness of the privacy implications when using facebook (2005). URL `http://lorrie.cranor.org/courses/fa05/tubzhlp.pdf`
43. Guha, S., Tang, K., Francis, P.: NOYB: Privacy in Online Social Networks. In: Workshop on Online Social Networks – WOSN 2008, pp. 49 – 54 (2008)
44. Gürses, S., Rizk, R., Günther, O.: Privacy design in online social networks: Learning from privacy breaches and community feedback. In: ICIS 2008: Proceedings Twenty Ninth International Conference on Information Systems. ACM (2008)
45. Il-Horn Hann and Kai-Lung Hui and Tom S. Lee and I. P. L. Png: Online Information Privacy: Measuring the Cost-Benefit Trade-off. 23rd International Conference on Information Systems (2002)
46. Jagatic, T., Johnson, N., Jakobsoon, M., Menczer, F.: Social Phishing. Communications of the ACM **50**(10), 94 (2007). DOI {10.1145/1290958.1290968}
47. Jessi Hempel: Is Facebook Losing Its Glow? Fortune Magazine (2009)
48. Jones, H., Soltren, J.H.: Facebook: Threats to privacy. http://web.mit.edu/jsoltren/www/facebook.pdf (2005)
49. Jones, K.: Facebook Admits Sexual Assault Suspect Used Site. Information Week (2009)
50. Kelley, P.G., Bresee, J., Cranor, L.F., , Reeder, R.W.: A "nutrition label" for privacy. Symposium On Usable Privacy and Security (SOUPS) 2009 (2009)
51. Kincaid, Jason: Wakeup Call: Facebook Isn't a Safe Haven. TechCrunch (2009)
52. Kolek, E., Saunders, D.: Online disclosure: An empirical examination of undergraduate facebook profiles. National Association of Student Personnel Administrators journal (2008)
53. Korolova, A., Motwani, R., Nabar, S.U., Xu, Y.: Link Privacy in Social Networks. In: CIKM '08: Proceeding of the 17th ACM conference on Information and knowledge management, pp. 289–298 (2008)
54. Krishnamurthy, B., Wills, C.E.: Characterizing Privacy in Online Social Networks. In: WOSN: Workshop on Online Social Networks, pp. 37 – 42 (2008)
55. Lampe, C.A., Ellison, N., Steinfield, C.: A familiar face(book): profile elements as signals in an online social network. In: CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems, pp. 435–444. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1240624.1240695
56. Lindamood, J., Kantarcioglu, M.: Inferring Private Information Using Social Network Data. WOSN: Workshop on Online Social Networks (2008)
57. Lipford, H.R., Besmer, A., Watson, J.: Understanding Privacy Settings in Facebook with an Audience View. In: 1st Conference on Usability, Psychology, and Security. USENIX Association (2008)

58. Loewenstein, G.: Keynote Speech: Searching for Privacy in all the Wrong Places: A behavioral economics perspective on individual concern for privacy. WEIS 07: The Seventh Workshop on the Economics of Information Security (2007)
59. Lookabaugha, T., Sicker, D.: Security and Lock-in. WEIS '03: Proceedings of the Third Workshop on the Economics of Information Security (2003)
60. Lucas, M.M., Borisov, N.: FlyByNight: Mitigating the Privacy Risks of Social Networking. In: WPES 08 - Workshop on Privacy in the Electronic Society, p. 1 (2008). DOI {10.1145/1456403.1456405}
61. McCombs, M., Shaw, D.: The Agenda-Setting Function Of Mass Media. Public Opinion Quarterly **36**(2), 176–187 (1972)
62. Milne, G., Culnan, M.: Information privacy: measuring individuals' concerns about organizational practices. Journal of Interactive Marketing **18**(3) (2004)
63. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and Analysis of Online Social Networks. In: IMC '07: Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pp. 29–42 (2007)
64. Nagaraja, S.: The economics of covert community detection and hiding. WEIS: Workshop on the Economics of Information Security (2008)
65. Narayanan, A., Shmatikov, V.: De-anonymizing social networks. 30th IEEE Symposium on Security & Privacy (2009)
66. O'Neill, N.: 10 Privacy Settings Every Facebook User Should Know. http://www.allfacebook.com/2009/02/facebook-privacy (2009)
67. Onwuasoanya, A., Skornyakov, M., Post, J.: Enhancing privacy on social networks by segregating different social spheres. Rutgers Governor's School of Engineering and TechnologyResearch journal (2008)
68. Pilkington, E.: Blackmail claim stirs fears over Facebook. The Guardian (2007). The Guardian
69. Poindexter, J.C., Earp, J.B., Baumer, D.L.: An experimental economics approach toward quantifying online privacy choices. Information Systems Frontiers **8**(5), 363–374 (2006). DOI http://dx.doi.org/10.1007/s10796-006-9013-4
70. Preibusch, S.: Implementing privacy negotiations in e-commerce. Lecture Notes in Computer Science **3841**, 604–615 (2006)
71. Preibusch, S., Beresford, A.R.: Privacy-preserving friendship relations for mobile social networking. W3C Workshop on the Future of Social Networking (2009). URL `http://www.w3.org/2008/09/msnws/papers/Preibusch-Beresford_Privacy-Preserving-Friendship-Relations.pdf`
72. Randall, D., Richards, V.: Facebook can ruin your life. And so can MySpace, Bebo... The Independent (2008). The Independent
73. Reagle, J., Cranor, L.F.: The platform for privacy preferences. Commun. ACM **42**(2), 48–55 (1999). DOI http://doi.acm.org/10.1145/293411.293455
74. Rosenblum, D.: What Anyone Can Know: The Privacy Risks of Social Networking Sites. IEEE Security & Privacy Magazine **5**(3), 40 (2007). DOI {10.1109/MSP.2007.75}
75. Schmidt, T.S.: Inside the Backlash Against Facebook. Time Magazine (2006)
76. Shepherd, J., Shariatmadari, D.: Would-be students checked on Facebook. The Guardian (2008). The Guardian
77. Simpson, A.: On the need for user-defined fine-grained access control policies for social networking applications. In: SOSOC '08: Proceedings of the workshop on Security in Opportunistic and SOCial networks, pp. 1–8. ACM, New York, NY, USA (2008). DOI http://doi.acm.org/10.1145/1461469.1461470
78. Smith, H.J., Milberg, S.J.: Information privacy: measuring individuals' concerns about organizational practices. MIS Q. **20**(2), 167–196 (1996). DOI http://dx.doi.org/10.2307/249477
79. Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In: EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce, pp. 38–47. ACM, New York, NY, USA (2001). DOI http://doi.acm.org/10.1145/501158.501163
80. Story, L., Stone, B.: Facebook Retreats on Online Tracking. The New York Times (2007)
81. Swan, H.: Social networking across devices: opportunity and risk for the disabled and older community. W3C Workshop on the Future of Social Networking (2009)
82. Varian, H.R.: Economic aspects of personal privacy. Topics in Regulatory Economics and Policy (2002)

83. Vila, T., Greenstadt, R., Molnar, D.: Why We Can't Be Bothered to Read Privacy Policies: Models of Privacy Economics as a Lemons Market. In: ICEC '03: Proceedings of the 5th International Conference on Electronic commerce, pp. 403–407. ACM, New York, NY, USA (2003). DOI http://doi.acm.org/10.1145/948005.948057
84. W3C, Mobile Web Best Practices Working Group, Checker Task Force: W3C mobileOK Checker (2009). URL http://validator.w3.org/mobile
85. Westlake, E.: Friend me if you facebook: Generation y and performative surveillance. TDR: The Drama Review **52**(4), 21–40 (2008). DOI 10.1162/dram.2008.52.4.21. URL http://www.mitpressjournals.org/doi/abs/10.1162/dram.2008.52.4.21
86. Wham, T.: Transcript of the FTC Workshop on Information Privacy: Measuring Individuals' Concerns about Organizational Practices. http://www.ftc.gov/bcp/workshops/infomktplace/transcript.htm (2001)
87. Whitten, A., Tygar, J.D.: Why Johnny can't encrypt: A usability evaluation of PGP 5.0. In: 8th USENIX Security Symposium (1999)
88. Williamson, D.A.: Social Networking Ad Spending. eMarketer (2008). eMarketer
89. XING AG: Press release: XING AG increases revenues by 80 percent and continues to grow profitably (2009). URL http://corporate.xing.com/english/press/press-releases/details/article/pm-de/7/3f79db5dea/?tx_ttnews[pointer]=2
90. Xu, W., Zhou, X., Li, L.: Inferring Privacy Information via Social Relations. International Conference on Data Engineering (2008)
91. Zheleva, E., Getoor, L.: To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. WWW: The International World Wide Web Conference (2009)
92. Zuckerberg, M., Schmidt, H.: Facebook CEO Mark Zuckerberg: Our focus is growth, not revenue. Frankfurter Allgemeine Zeitung / FAZ.NET (2008)

Chapter 9
The Policy Maker's Anguish: Regulating Personal Data Behavior Between Paradoxes and Dilemmas

Ramón Compañó, Wainer Lusoli

Abstract Regulators in Europe and elsewhere are paying great attention to identity, privacy and trust in online and converging environments. Appropriate regulation of identity in a ubiquitous information environment is seen as one of the major drivers of the future Internet economy. Regulation of personal identity data has come to the fore including mapping conducted on digital personhood by the OECD; work on human rights and profiling by the Council of Europe and major studies by the European Commission with regard to self-regulation in the privacy market, electronic identity technical interoperability and enhanced safety for young people. These domains overlap onto an increasingly complex model of regulation of individuals' identity management, online and offline. This chapter argues that policy makers struggle to deal with issues concerning electronic identity, due to the apparently irrational and unpredictable behavior of users when engaging in online interactions involving identity management. Building on empirical survey evidence from four EU countries, we examine the first aspect in detail – citizens' management of identity in a digital environment. We build on data from a large scale $(n = 5,265)$ online survey of attitudes to electronic identity among young Europeans (France, Germany, Spain, UK) conducted in August 2008. The survey asked questions about perceptions and acceptance of risks, general motivations, attitudes and behaviors concerning electronic identity. Four behavioral paradoxes are identified in the analysis: a privacy paradox (to date well known), but also a control paradox, a responsibility paradox and an awareness paradox. The chapter then examines the paradoxes in relation of three main policy dilemmas framing the debate on digital identity. The paper concludes by arguing for an expanded identity debate spanning policy circles and the engineering community.

Ramón Compañó
European Commission - Directorate General Joint Research Centre (JRC), Institute for Prospective Technological Studies (IPTS), e-mail: Ramon.compano@ec.europa.eu

Wainer Lusoli
European Commission - Directorate General Joint Research Centre (JRC), Visiting Research Fellow, University of Chester,e-mail: Wainer.lusoli@ec.europa.eu

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_9,

9.1 Introduction

Regulators in Europe and elsewhere are paying great attention to identity, privacy and trust in online and converging environments. Understanding and regulating identity in a ubiquitous information environment is seen as one of the major drivers of the future Internet economy [15]. Regulation of personal identity data has come to the fore including mapping conducted on digital personhood by the OECD [18]; work on human rights and profiling by the Council of Europe [4] and major studies by the European Commission with regard to self-regulation in the privacy market, electronic identity technical interoperability and enhanced safety for young people [13].

These domains overlap onto an increasingly complex model of regulation of individuals' identity management, online and offline. This model comprises consumer policy, where priorities are set based on the critical assessment of location and service fruition and trust and privacy as prerequisites for the future common digital market [11]; human rights agenda, in line with the consequences of advanced profiling techniques [4] and with surveillance concerns in relation to information society security [10]; online safety policy, especially in relation with younger users [5] policies concerning the right of access to advanced, interoperable EU services in the sphere of justice [6]; and a set of policies regarding the economic impact of future networks [15]. This implies a regulatory infrastructure of identity which, if fully sketched, is way grander than one that to date tackles identity-theft and ensures smooth services fruition across the EU (interoperability).

The paper claims that policy makers struggle to deal with issues concerning electronic identity. This has two main reasons: the apparently irrational and unpredictable behavior of users when engaging in online interactions involving identity management and a seemingly intractable set of dilemmas. The former problem, verily a set of behavioral paradoxes, is compounded by the lack of multi-country, systematic, comprehensive data on users' attitudes and behaviors: trust, privacy, behavioral intentions and confidence in relation to personal identity data. In addition, debate is mainly limited to the so-called privacy paradox and people's willingness to disclose personal data.

Building on empirical survey evidence from four EU countries, this paper examines the last aspect in detail – citizens' management of identity in a digital environment. We build on data from of a large scale (n = 5,265) online survey of attitudes to electronic identity among young Europeans' (France, Germany, Spain, UK) conducted in August 2008. The survey asked questions about perceptions and acceptance of risks, general motivations, attitudes and behaviors concerning electronic identity.

This paper is unusual as it defies the established practice of hypothesis testing, corroboration or rejection. Rather, data and results follow a logical argument to support the main thrust of the paper that identity-related policy making is hampered by multiple aims, behavioral idiosyncrasies and systemic dilemmas. While this may be seen as less than 'scientific' in traditional hard science milieus (physical security of identity systems), it contributes to articulate a debate that is sometimes overlooked

in such circles. In the conclusion, the paper argues for the extension of the identity debate to span policy circles, the engineering community and a growing section of multi-disciplinary approaches to identity.

9.2 Existing Work on the Privacy Paradox

The so-called 'privacy paradox' is one of the central topics in the debate on privacy and identity. The privacy paradox states that users are concerned about privacy but they disclose a significant amount of personal data and take no action to protect themselves. Several studies confirmed the paradox. It has been found in experimental settings, with specific reference to the role of risk as a discriminant predictor of attitudes (positive association) vs. behavior (no association) [14]. The paradox has been found in relation to social networking behaviors among US college students [9]. A study of Dutch students confirms the paradox across a range of possible defensive behaviors such as behavioral measures and common and more advanced privacy enhancing technologies [16]. For specific services, such as instant messaging, the relation between privacy concerns and protective action may be stronger. People who remain unprotected do so because of lack of skills [17]. Again in relation to social networking, young people were found to adopt copings tactics rather than adapting strategically to the new information environment [22]. This may be a way of reconciling actual behaviors with attitudes and social desirability. Finally, privacy concerns have a negative effect on information disclosure but a positive effect on protection intention; transaction intention, however, remains unaffected. Furthermore, information sensitivity has a negative effect on information disclosure and transaction intention [20]. To summarize, people do disclose online despite privacy risks, but go to some length to mitigate the effects of disclosure, especially in relation to sensitive information.

However, work on the privacy paradox struggles to cast a net wider than a single country (e.g. the Netherlands), a target group (e.g. students), a limited theoretical focus (e.g. the paradox itself). This is in some way understandable; most of the studies reviewed are small scale experiments; official, multi-country data that would help casting a wider net are lacking; work is often uni- rather than multi-disciplinary. To your knowledge, five studies come close to an encompassing definition of possible, relevant variables:

- European Commission's Eurobarometer Flash study on 27 Member States on confidence in the Information Society, with questions on security risk awareness / knowledge, damage and protective behaviors [7];
- European Commission's Eurobarometer Flash study on 27 Member States with questions in relation to data protection in own country, plus one question on privacy-enhancing technologies and one on internet trust [8];
- OCLC survey of six countries, focusing on social networking and social media in relation to privacy and trust [3];

Table 9.1 Survey totals.

	France	UK	Germany	Spain	Total
Emails sent	129,828	143,476	101,086	157,053	531,443
Invalid email addresses	1,580	3,000	3,015	559	8,154
Invalid email rate	**1.2%**	**2.1%**	**3%**	**0.4%**	**1.5%**
Valid email addresses	128,248	140,476	98,071	156,494	523,289
Emails opened	47,724	20,209	12,009	30,149	110,091
Open rate	**37%**	**14%**	**12%**	**19%**	**21%**
Emails clicked on	9,155	3,020	2,672	4,240	18,087
Click rate	**7.1%**	**2.1%**	**1.7%**	**2.7%**	**3.5%**
Respondents to the first question	4,485	2,631	1,709	3,318	12,143
Respondents to the last question	2,014	1,258	819	1,174	5,265
Full answer rate	**45%**	**48%**	**48%**	**35%**	**43%**

- OECD review of safety and security official statistics focusing mainly on security, with limited if no focus on other aspects such as privacy trust and confidence [19];
- FIDIS (Future of ID in the Information Society Network of Excellence) web survey in 19 EU countries on perceptions of institution-based trust in the handling of personal data [1].

9.3 Methodology

To examine citizens' seemingly irrational behavior concerning the management of the identity in a digital environment, we build on data from of a large-scale online survey of attitudes to electronic identity among young Europeans' in France, Germany, Spain and UK conducted in August 2008[1]. The survey examines the attitudes and behaviors of young people because they are the next generation of internet users, citizens and consumers; arguably, they also differ from previous generating in their proximity to and confidence with new digital technologies [2].

Preliminary research steps included two focus groups in each country on a topic guide consonant with the final questionnaire; a two-day expert workshop to validate the questionnaire; a pre-test conducted with 100 young people in the UK in June 2008. Once the questionnaire was finalized and pre-tested, invitations to the online survey were sent to 531,443 young people in France, UK, Spain and Germany, in July and August 2008. The survey obtained 12,143 responses to the first question and 5,265 responses to the whole questionnaire [which we use for the analysis reported here]. The survey obtained at least 1000 respondents per country except in Germany, where the number of completed questionnaires was n = 819. Table 9.1 reports the details of the recruitment process.

In terms of representativeness,

[1] More details on the methodology of the study can be found in [12].

- Of all respondents (partial and complete), 37% from France French, 27% from Spain, 22% from the UK and 14% from Germany.
- Overall 56% are male and 44% female, this proportion being different in some countries, notably in Spain (78% male) and in the UK (65 % male).
- The majority are 15-18 years old (46%), 29% are between 19 and 21 and 26% are 22 years old or older. There are less 'younger' people from the UK and Germany.
- Nearly 50% are students (more students in UK and less in Spain). Around 30% of young people are 'blue collar' workers (but only 2.6% in England and 50% in Spain).
- Considering education, only 2% have a Doctorate and 18% a Master (less in UK and Germany). The most common degree is 'licence' with 41% (30% in UK and Spain).

Overall, therefore, there is considerable variance in terms of socio-demographic factors across the four countries. In future studies, steps need to be taken to standardize the parameter estimates of the sample on those of the population. Conversely, however, the sample represents very closely the internet access and use of young people 15-25 years olds in the respective countries (data not reported here, please refer to [12]).

The survey asked questions about perceptions and acceptance of risks, general motivations, attitudes and behaviors concerning electronic identity. Dimensional analysis and factor analysis were used to extract latent indicators. Below, we provide a list of indicators and variables relevant to this paper. We report below the overall theme of the question/s, the question formulation, the factor/s extracted via dimensional analysis and other items that are used in the discussion. Question wording, options, level of measurement and values are provided in the Appendix.

1. *Enablers of identifications systems*
   - Q21 Which of the following elements could encourage you to use identification systems?
     - 2 factors: guarantees and control devices
2. *Online personal data disclosure*
   - Q22 Indicate what information you provide on Internet
     - 4 factors: low disclosure [information that gets rarely disclosed], basic social networking [SNS], advanced SNS and high disclosure
3. *Internet confidence*
   - Q24 More generally, concerning the Internet, you would say that ...
     - 1 factor: Internet confidence; 1 single item used in analysis: self-confidence in ability to protect oneself online
4. *Privacy risk perceptions*
   - Q26 How concerned are you about the following risks in relation to your personal information

  - 2 factors: identity damage, data tracking

5. *Responsibility*
   - Q27 Who is responsible to protect personal data on line?
6. *Data protection strategies*
   - Q28 On the Internet, how often do you . . .
   - Q29 On the Internet, I usually protect my personal data in the following ways
     - 5 factors: offline strategies [hardware based], online strategies [software based], shielding strategies, minimization strategies and avoidance strategies
7. *Data protection knowledge*
   - Q30 Do you know your rights in terms of data protection?
     - 1 scale of data protection knowledge
8. *Data protection attitudes*
   - Q31 For each of the following statements, please state if you tend to agree or not
     - 1 factor: attitude towards data protection
9. *Remedies*
   - Q32 What do you think are efficient ways to protect your identity, online and offline?
     - 2 factors: awareness raising and direct intervention; 1 single item used in analysis: give users more control on their personal data

The survey also included standard socio-demographic questions and a range of questions on internet access and use, and knowledge and use of identification systems. The latter are used to argue the point in relation to policy makers' dilemmas, discussed in section 5, and are reported in the Appendix. Socio-demographic questions and other questions included in the survey are not reported for reasons of space and relevance to the argument proposed here.

9.4 Paradoxes

Overall, survey results are in line with previous findings from the literature, particularly those on young people's perception of technologies and public policies, privacy, trust and enablers. However, results point to a number of unexpected attitudes of young people that appear irrational.

9.4.1 The Privacy Paradox

The survey confirms the prevalence of the privacy paradox [Table 2, marked in yellow], whereby young people disclose a range of personal information despite high perception of privacy risks. In general, the public is primarily concerned about loss of privacy that lead to security problems but few everyday activities are considered extremely or very private. Our results confirm as much, as disclosure of 'basic' biographic information is unrelated to privacy concern; on the other hand, there is a very weak negative correlation (Pearson's R2 -.04) between these and disclosure of potentially more sensitive data (medical history, etc). The survey confirms that social networkers, particularly younger users, may well be ill informed about the detail they are making publicly available, as it is often unrelated to their privacy concerns. But the need to appear seems to justify disclosure in young people's eyes. Online social networking, for instance, is more about enhanced and increased personal disclosure than about the maintenance of wider social networks [21].

9.4.2 The Control Paradox

People desire full control on their personal data, but avoid the hassle to keep it up to date. People know that there are technology tools to protect them and think they may be efficient, but they do not use them [Table 2, marked in red]. More than 70% of respondents think that there are efficient solutions to identity-related problems online. Technical solutions are favoured, alongside other supply-side solutions. While 73% claim that it is efficient to 'give users more direct control on their own identity data', a minority employs strategies such as data minimization, avoidance or active management of own personal data. In detail, there is no correlation between shielding and minimization user practices and the call for more user control; there are weak correlations between data avoidance and hardware-based strategies and the perception that user should have more control; and there are conflicting (positive and negative) correlation between employment of Internet-based tactics and user control perception.

9.4.3 The Responsibility Paradox

Overall, young people consider that the responsibility to manage personal data is shared. They do not attribute responsibility for the protection of personal data to governments or police and courts. Most young people believe that it is either their own responsibility to protect their data online or the responsibility of the companies they are transacting with. They are asking for tools that give them more direct control on their own identity data. But at the same time, they are not confident in their own ability to keep their data protected. Overall, while only half of the re-

Table 9.2 Correlations between main variables and indicators.

Variables and indicators	Self efficacy	Low disclose	Advanced SNS	High disclose	Basic SNS	Risk: data tracking	Risk: identity damage	DP tactics: offline	DP tactics: online	DP tactics: shield	DP tactics : minimise	DP tactics: avoid	Remedies: awareness	Remedies: intervene	Enabler: guarantee	Enabler: control	Remedies: user control	DP attitudes
Low disclosure	.06	1			-.04	-.17	-.11	.05	-.09	.16	-.14	-.12	.08	.05	.04	-.15	-.05	.37
Advanced SNS	.06		1		-.04	-.11	-.06	.05	-.04	.04	.06	.06	.08	.06	.07	.04	-.06	.37
High disclosure	.04			1					-.11	-.19	.06	-.23	.07	.07	.04	.15	.07	.04	.09	.09	Privacy paradox
Basic SNS	.07				1				-.07			-.14		.04	.05	.04		.09	Control paradox
Risk: data tracking	-.17	-.04				1		.13	.06	-.08	.08		.18	.16	.16	.04	-.12	-.22	Awareness paradox
Risk: identity damage	-.11						1		.08	-.05		.05	.15	.06	.04	.06	-.04	-.11		
DP tactics: offline	.05	-.09				.13		1					.10	.08	.16	.05	.06	.04			
DP tactics: online			-.04	-.11	-.07	.06	.08		1				.07		-.05	.07	-.06	.20			
DP tactics: shielding		.16	.04	-.19		-.08	-.05			1			-.15	-.06	-.21			.04			
DP tactics: minimisation		-.14		.06		.08					1					.15		.05			
DP tactics: avoidance	-.08	-.12		-.23	-.14		.05					1		-.06				.08			
Remedies: awareness				.07		.18	.15	.10	.07	-.15			1			.14	.09	.05			
Remedies: intervention	.05		.06	.07	.04	.16	.06	.08		-.06	.06	-.06		1		.14	.09	.06			
Enabler: guarantees		-.15		.15	.05	.16	.04	.16	-.05	-.21	.15				1						
Enabler: control	.05	.04	.07	.04		.04	.06	.05	.07				.09			1					
Remedies: user control [1 item]	.04	-.05	-.06			-.12	-.04		-.06			.04	-.10	-.72		.06	1				
DP knowledge [1 tem]						.09	.04	.22	.20			.05		.05	.10	.06	-.05	1			
DP attitudes	.37	0.04	0.06	0.09	0.09	-.22	-.11	.04		.04	.05	.08		.06	.06	.04	.07	.07			
Shading codes	Responsibility paradox																				

NOTE: All correlations shown are significant at the 0.01 level (2-tailed).

spondents said they are confident they can protect their own privacy online, only 21% claim that it is very efficient to 'give users more direct control on their own identity data'. While most people believe that it is either their own responsibility, they seem to admit that many users do not have the knowledge to do this effectively [Table 2, marked in blue]. Furthermore, young people tend to neglect trust seals and do not appreciate privacy enhancing tools. Overall, there is a negative correlation between perceived efficacy of user control on their own data and perception of actual measures that would enable this control (such as receipts, information on systems and counter-profiling information). The awareness paradox Data protection (DP) legislation is unknown and unloved [Table 2, marked in green]. Young EU citizens' knowledge level about DP laws is low. Even lower is their appreciation of the current DP framework. Paradoxically, more knowledge only breeds slightly more positive attitudes (Pearson's R2 .07). People knowing a lot or nothing about DP (24%), are significantly different in their attitudes. However, for the majority of the people in the middle (76 % knowing a bit or not much) there is practically no correlation with attitudes. Moreover, more knowledge on DP rights does not to influence the behavioral intention to adopt digital services based on personal data disclosure (weak negative correlation). Finally, there is a strong correlation (.37) of self-efficacy with DP attitudes, but not with knowledge. But it is knowledge that gets people to stay protected (correlation .20), rather than attitudes, positive or negative (no correlation). These findings suggest that personal experience may matter more than understanding of the legal system. It is not surprising that young people should ask for 'hands-on' regulation. Young people desire reassurance, via practical tools more than via awareness raising. Tools such as guarantees (labels and logos) appeal to young people, while they also appreciate tools that may assist control of personal data provided to public or private authorities.

9.5 Dilemmas

Alongside having to deal with a number of paradoxes, policy-makers also face a number of dilemmas when devising identity-related policies.

9.5.1 The Cultural Dilemma

As digital culture and behavioral attitudes vary across Member States, pass-partout policies are not available. There are significant differences between countries in terms of digital culture and markets. Countries vary in terms of mode of Internet connection. In France, 95% connect using home broadband, but 40% also connect at school or university and 20% through pay wi-fi network. In the UK, 34% connect at work but only 15% at school or university and very few in other ways. In Spain,

only 66% connect using home broadband, 24% using dial-up and 19% in an internet café.

In terms of Internet activities, discrepancies appear between countries. Managing profile on social networks is today prevalent (43%), although it is less widespread in Spain (30%). France has a blogging and instant messaging culture; French young people author more blogs (35%) than people in other countries (<15%), 85% of French youngsters use instant messaging (more any other country) youngsters are more skilled in Germany than elsewhere. Fewer youngsters from all countries design a web site or install plug-ins than in Germany (27%).

Internet access and activities are important for personal innovativeness, and, in turn, for the take up and regulation of digital services.

9.5.2 The Market Fragmentation Dilemma

The digital market that supports and profits from personal data disclosure is significantly fragmented. Young EU citizens are Web experts and connected mainly at home using broadband. They constitute a specific part of the population particularly Internet minded. However, they are not a homogeneous group. There are three distinct groups in terms of activities. A group (48%) of new Internet users doing classical activities (check emails; search engines); a group (34%) of older Internet users also having web 2.0 activities on social networks; a group (18%) using all the social possibilities of the Internet such as keeping a blog and participating in online discussion forums and chats. Young, innovative people who have being going online via broadband several times a day for more than 5 years are leaders in relation to managing their identity online. This behavior often requires significant online disclosure of personal data, which youngsters are mostly happy to provide

However, young people who engage in most advanced internet behavior have a more positive attitude concerning the Internet and lesser perceptions of risk. How to cater for these two different publics (lesser skilled, likely to disclose, lacking confidence; more skilled, very likely to disclose, having more confidence) is matter of great complexity. This segmentation is further propelled by cultural and economic differences across EU Member States. Difference in technical skills, cultural appreciations and market maturity may lead to different applications of personal data disclosure across the EU. From a policy maker's point of view, however, governments must strive in offering all citizens equal opportunities and this is more likely the lesser such fragmentation.

9.5.3 The Public-Private Dilemma

Governments, as active stakeholders to promote digital service take-up suffer from a triple dilemma. First, the survey evaluated the perceived benefits and risks towards

personal data disclosure. Contradictory perceptions exist. While systems are not always seen as risky, EU citizens demand more security and privacy, personalization of services and ease of use. People want to be safe online, but they are wary of governments. Young people do not trust governments but expect them to act.

Second, the public hand as one of the largest investors of ICTs would be in a key position to shape and promote the development of innovative services based on data disclosure. But the majority of digital services developed by governments are largely regarded as unattractive by young people, making them useless as platform for wider deployment in other domains like leisure, work or business.

Third, unlike business players, governments have little room for maneuver for negotiations. While some people would accept profiling in exchange of commercial benefits or personalized services, similar incentives are very limited for governments. It would be unacceptable, for instance, to award a tax discount only to those citizens submitting the tax declaration online, while asking the payment of full taxes all others submitting it in paper.

9.6 Conclusion

In their decisions, policy makers need to take into account that citizens do not always behave rationally. The paper highlights a number of behavioral paradoxes that became apparent from an online survey of young people. In spite of these apparently irrational patterns, governments are increasingly under pressure to design a viable framework to enable innovative services to the benefit for their citizens, largely based on personal data disclosure.

From many quarters, based on evidence beyond our own survey, there is a strong call for effective, fair and transparent data protection rules [7]. In our survey, trust in rules (fair play by service providers) emerged as an important factor in addition to traditional understandings of trust. Indeed, there are multiple enablers of identity disclosure. Guarantees, assurance of data protection law respect and precise information on systems are likely to encourage the adoption of services based on personal data disclosure. Solutions based on these principles need implementing, regulating and enforcing.

For this to happen, there is an urgent need to look at a wider picture. A complex equation involving internet skills, self-efficacy, privacy perception, global risks and disclosure needs to be constructed in relation to the efficacy of different regulatory alternatives in relation to eID. The survey confirmed the privacy paradox. It also showed that behavioral paradoxes concerning data control, responsibility and awareness compound the picture. Any solution tailored to tackle the former needs to factor in system effects in other domains. But this, it was argued, is not the full picture altogether. Policy action faces systemic constraints.

Governments have to struggle with a number of dilemmas that further limit the range of viable policy options. First, governments need to design policies that enhance the public good, in contrast to companies that can follow a market segmen-

tation approach. Second, the EU ICT markets are very different across the Member States. Finally, there is a cultural component to take into account. These may become serious issues, as there are considerable differences in attitudes with respect to the use and perception of digital services within society, our survey shows.

References

1. Backhouse, J., Halperin, R.: A Survey on citizen's trust in ID systems and authorities. Fidis Journal **1**(Online) (2007)
2. Buckingham, D. ed.: Youth, identity, and digital media. The John D. and Catherine T. MacArthur Foundation series on digital media and learning. MIT Press, Cambridge, MA (2008)
3. De Rosa, C., Cantrell, J., Havens, A., Hawk, J., Jenkins, L., Cellentani, D., Dalrymple, T., Olszewski, L., Smith, S., Storey, T.: Sharing, privacy and trust in our networked world. Online Computer Library Center, Dublin, OH (2008)
4. Dinant, J.-M., Lazaro, C., Poullet, Y., Lefever, N., Rouvroy, A.: Application of Convention 108 to the profiling mechanism: Some ideas for the future work of the consultative committee (T-PD). Council of Europe, T-PD, Strasbourg, France (2008)
5. EDPS: Opinion on the proposed multiannual Community programme on protecting children using the Internet and other communication technologies. EDPS, Brussels, Belgium (2008)
6. European Commission: Communication from the Commission – Towards a European e-justice Ssrategy. European Commission – DG JLS, Brussels, Belgium (2007)
7. Gallup: Confidence in information society. EC DG Information Society and Media: Brussels, Belgium (2009)
8. Gallup: Data protection in the European Union – Citizens' perceptions. EC DG Justice, Liberty and Security: Brussels, Belgium (2008)
9. Gross, R., Acquisti, A.: Information revelation and privacy in online social networks. In: Proceedings of the ACM Workshop on Privacy in the Electronic Society, pp. 71–80. ACM Press, New York (2005)
10. Hammarberg, T.: Strong data protection rules are needed to prevent the emergence of a surveillance society. Council of Europe, Commissioner for Human Rights, Strasbourg, France (2008)
11. Kuneva, M.: Key Challenges for Consumer Policy in the Digital Age. Roundtable on Digital Issues, London, 20 June (2008)
12. Lusoli, W., Miltgen, C.: Young People and Emerging Digital Services. An Exploratory Survey on Motivations, Perceptions and Acceptance of Risks, W. Lusoli, R. Compañó, and I. Maghiros, Editors. EC JRC Institute for Prospective Technological Studies, Sevilla (2009)
13. Marcus, J. S., Carter, K., Robinson, N., Klautzer, L., Marsden, C., Reidenberg, J., Abder, C., Burton, C., Cooms, L., Kover, E., Poullet, Y., De Villenfagne, F., Dumortier F., Peake, A., Kamimura, K., Tanaka ,T.: Comparison of Privacy and Trust Policies in the Area of Electronic Communications. wik-Consult/RAND Europe, CLIP/CRID/GLOCOM, Bad Honnef, Germany (2007)
14. Norberg, P.A., Horne D.R., Horne D.A.: The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of Consumer Affairs **41**(1), 100–126 (2007)
15. OECD: The Seoul declaration for the future of the Internet economy. Seoul, OECD Ministerial meeting, 18 June (2008)
16. Oomen, I., Leenes, R.: Privacy risk perceptions and privacy protection strategies. In: Proceedings of the First IFIP WW 11.6 Working Conference on Policies and Research in Identity Management. Springer, Berlin Heidelberg, Germany (2007)
17. Paine, C. Reips, U.-D., Stieger, S., Joinson, A., Buchanan, T.: Internet users' perceptions of 'privacy concerns' and 'privacy actions'. International Journal of Human-Computer Studies **65**(6), 526-536 (2007)
18. Rundle, M. C., Blakley, B., Broberg, J., Nadalin, A., Olds, D., Ruddy, M., Marcelo Thompson, M., Mello Guimarães, M., Trevithick, P.: At a crossroads: "Personhood" and the digital identity in the information society. OECD, Paris, France (2007)

19. Schaaper, M.: Measuring security and trust in the online environment: a view using official data. EAS, DSTI, OECD, Paris (2008)
20. Shu, Y., Kanliang, W.: The influence of information sensitivity compensation on privacy concern and behavioral intention. ACM SIGMIS Database **4**0(1), 38–51 (2009)
21. The Economist: Primates on Facebook: even online, the neocortex is the limit. The Economist (2009)
22. Tufekci, Z.: Can you see me now? Audience and disclosure regulation in online social network sites. Bulletin of Science Technology and Society **28**(1), 20–36 (2008)

9.7 Appendix

[enablers]
Q21 Which of the following elements could encourage you to use identification systems?
Tick all that apply
2101 A receipt after you have provided the information
2102 Information on the identification system
2103 Information on the use of the data you provide
2104 Testimonials of persons having experimented the identification system
2105 The assurance that law on personal data protection is respected
2106 A label or logo proving that the system is secure
2107 Guarantees that data are not resold or reused by another organization
2108 A single record with all my transactions, interactions, traces, so I know what is around about me
2109 Others (specify)
2110 None

[online personal data disclosure]
Q22 Indicate what information you provide on Internet
Yes No Don't know
2201 Name/surname
2202 Age
2203 Nationality
2204 ID number
2205 Postal address
2206 Bodily appearance
2207 Things I do
2208 Tastes / Opinions
2209 People I meet regularly, my friends / Membership of associations
2210 Places where I usually go
2211 Information you give on social networks such as Facebook or Study VZ
2212 Photos of me
2213 Financial information (revenues, credits, . . .)
2214 Medical information (social security number, . . .)
2215 Bank information (bank card number, account number, . . .)
2216 Judicial information (criminal record, . . .)
2217 Biometric information (fingerprint, iris. . .)

[Internet confidence]
Q24 More generally, concerning the Internet, you would say that. . .
7-point scale, Strongly disagree To Strongly agree
2401 The internet has enough safeguards to make me feel comfortable giving my personal details online
2402 The internet is now a robust and safe environment in which to transact.

2403 The internet provides a trusted environment in which to make transactions for leisure, work and business

2404 The internet is safe enough to preserve my privacy as I carry out business and personal activities

2405 I am confident that I can protect my privacy online

[privacy risk perceptions]

Q26 How concerned are you about the following risks in relation to your personal information

5-point scale, Very concerned To Not at all concerned

2601 Companies possess information about me that I consider private

2602 My personal information is used without my knowledge

2603 My personal data is shared with third parties without my agreement

2604 My behavior and activities can be monitored online

2605 My online personal data is used to send me commercial offers

2606 My identity is reconstructed using personal data from various sources

2607 My views and behaviors may be misrepresented based on my online personal information

2608 My reputation may be damaged by online personal information

2609 My identity is at risk of theft online

2610 My personal safety may be at risk due to online personal information

2611 I may be victim of financial fraud online

[responsibility]

Q27 Who is responsible to protect personal data on line?

Tick one

2701 On the Internet, it is my responsibility to protect my personal data

2702 It is the government responsibility to protect my personal data online

2703 It is everybody's responsibility to make sure personal data are safe online

2704 It is the responsibility of the company I transact with to protect my personal data online

2705 It is the responsibility of the police and courts to ensure that personal data are protected online

[data protection strategies 1] **Q28 On Internet, how often do you …** Never Sometimes Often Always

2801 Give your real identity

2802 Use a pseudonym

2803 Give a minimum of information

2804 Give wrong information

2805 Do not answer personal questions

2806 Give the identity of another person

[data protection strategies 2]

Q29 On the Internet, I usually protect my personal data and identity in the following ways

Never Sometimes Often Always

2901 Read the privacy policy of web sites
2902 Use dummy email account to shield my identity
2903 Update virus protection
2904 Scan data with anti-spy ware
Q2905 Install operating system patches
2906 Erase cookies
2907 Use tools and strategies to limit unwanted email (spam)
2908 Check that the transaction is protected or the site has a safety badge before I enter personal data
2909 Adapt my personal data so that no linking between profiles is possible
2910 Change the security settings of my browser to increase privacy
2911 Use tools limiting the collection of personal data from my computer (e.g. Firewall, cookie filtering)

[data protection knowledge]

Q30 Do you know your rights in terms of data protection?
Tick one
I never heard about it
I heard about it but I do not know it really
I know a little bit about it
I know it very well

[data protection attitudes]
Q31 For each of the following statements, please state if you tend to agree or not
7-point scale, Strongly disagree To Strongly agree
3101 In [country], my personal data are properly protected
3102 [Nationality] legislation can cope with the growing number of people leaving personal information on the Internet
3103 I believe that the systems used by the public authorities to manage the citizens' personal data are technically secure.
3104 I believe that citizens will be able to keep a good level of control over their personal data
3105 I will always be able to rely on public authorities for help if problems arise with my personal data
3106 I believe that the authorities that manage my personal data are professional and competent

[remedies]
Q32 What do you think are efficient ways to protect your identity, online and offline?
Very efficient to Not at all efficient
3201 Give users more direct control on their own identity data
3202 Allocate more resources to monitoring and enforcing existing regulations
3203 Require that service providers take greater care of their customer's identity
3204 Find better technical solution that preserve users' privacy and safety

3205 Provide formal education on safe identity management
3206 Raise awareness of the implication of unsafe identity behavior
3207 Set up clear guidelines for safe identity management, online and offline
3208 Make greater use of warnings and signs to signal possible unsafe behaviors

[Internet access and activities]
Q2 How do you connect to the Internet ?
Tick all that apply
201 Where I usually live (home, parent's home, Uni) using broadband
202 Where I usually live (home, parent's home, Uni) using dial-up
203 At work
204 At school or university
205 Through pay wi-fi network (airport, train station. . .)
206 In an internet cafe
Q3 How often do you connect to the Internet?
Tick one
301 Several times a day
302 Once a day
303 A few times a week
304 Less than once a week
305 Less than once a month
306 Never
Q4 What devices do you use to connect to the Internet?
Tick all that apply
401 Personal Desktop PC
402 Shared Desktop PC
403 Laptop computer
404 WII, playstation or other gaming console
405 On mobile phone or PDA, using GPRS or 3G
Q5 Do you do the following activities on the internet?
Tick all that apply
501 Check email
502 Instant messaging
503 Participate in chat rooms, newsgroups or an online discussion forum
504 Use a search engine to find information
505 Use website (flicker, Youtube, etc) to share pictures, videos, movies etc.
506 Make or received phone calls over the Internet
507 Manage your profile on a social networking site such as Youtube, myspace or Facebook
508 Design or maintain a website (not just a blog)
509 Keep a web-log (or what is called a Blog)
510 Install plug-ins in browser to extend its capability
511 Use peer-to-peer software to exchange movies, music, etc.

Chapter 10
Valuating Privacy with Option Pricing Theory

Stefan Berthold and Rainer Böhme

Abstract One of the key challenges in the information society is responsible handling of personal data. An often-cited reason why people fail to make rational decisions regarding their own informational privacy is the high uncertainty about future consequences of information disclosures today. This chapter builds an analogy to financial options and draws on principles of option pricing to account for this uncertainty in the valuation of privacy. For this purpose, the development of a data subject's personal attributes over time and the development of the attribute distribution in the population are modeled as two stochastic processes, which fit into the Binomial Option Pricing Model (BOPM). Possible applications of such valuation methods to guide decision support in future privacy-enhancing technologies (PETs) are sketched.

10.1 Introduction

In certain jurisdictions, the right of informational self-determination implies active control of one's personal data. To exercise such control, it is crucial for people to understand the implications of data disclosure. While visions for privacy-enhanced identity management [22] seek to provide technical means for *securing the disclosure* of personal data under different threat models, it is still a challenging question how individuals can be supported in *assessing the value* of their personal data. However, the latter is a prerequisite for the former: making informed disclosure decisions depends on the ability to compare between the alternatives in the first

Stefan Berthold
Karlstads Universitet, Fakulteten för Ekonomi, Kommunikation och IT, Universitetsgatan 2, 651 88 Karlstad, Sweden, e-mail: stefan.berthold@kau.se
Rainer Böhme
International Computer Science Institute, 1947 Center Street, Ste 600, Berkeley, CA 94704, USA, e-mail: rainer.boehme@icsi.berkeley.edu

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_10,

place. Irrespective of the concrete supporting technology, a major obstacle that prevents people from making rational decision regarding their privacy is the uncertainty about possible future consequences of data disclosure at present [1]. Similarly, known information-theoretic privacy metrics at best reflect the present value of personal data. These metrics ignore that the value of personal data, for instance for re-identification, may change over time. However, the time between disclosure and exploitation of personal data is very relevant for the inter-temporal value of personal data: the more time passes between both events, the more uncertainty arises about the value. This is so because attribute values which apply to a data subject at the time of disclosure may not be applicable to the same data subject anymore when the data is exploited. Also the distribution of attribute values in the entire population changes over time. Attribute values which uniquely describe a single data subject at present may become common in the population in the future. Accordingly, their value for the purpose of re-identification would decline over time.

In this chapter, we present a framework to model this kind of uncertainty and account it in measures of the future value of attributes that are to be disclosed at present. Although novel to the field of privacy research, modeling uncertainty about future states has a long tradition in other disciplines, such as finance and accounting. So we will draw on concepts from option pricing theory and show how this theory translates to the problem of personal data disclosure. The core idea is to interpret data disclosure as writing a call option that allows the counterpart to use the data for identification later on.

To start with a simple case and focus on the core idea, we confine ourselves in this chapter to binomial stochastic processes, similar to the Binomial Option Pricing Model (BOPM) [11]. In principle, the theory generalizes so that any stochastic process with better fit to reality can be plugged into our framework. The choice of the most appropriate process for specific attributes in a certain context is an empirical question. It thus falls beyond the scope of this work. Again for the sake of simplicity, we limit our view to a single attribute with finite and discrete attribute values. Extensions to multiple attributes are possible, but increase the dimensionality of the problem substantially. We further rule out any ambiguity or measurement error and assume that exactly one attribute value can be assigned to each data subject.

Under the above-stated assumptions, the value of an attribute to re-identify a data subject after some time is determined by a combination of two factors:

1. by the chance that the attribute value still applies to the particular data subject. This factor is governed by the individual behavior of the data subject. So we will refer to it as the *micro* level.

   And, if this condition holds,

2. by the uniqueness of the attribute value, i. e., how many other data subjects in the population do meanwhile share the same attribute value and thus form an equivalence class? This factor is driven by the aggregate behavior of all, possibly heterogeneous, data subjects in the population. So we will refer to it as the *macro* level.

In our framework, each factor is a source of uncertainty and can be modeled by a stochastic process from the point of view of a transaction counterpart, who

1. learns the attribute value of a data subject at the time of disclosure, and
2. can observe the distribution of attribute values in the population at any time (e. g., through representative anonymous surveys or observation).

Hence, changes of individual attributes remain private information of each data subject. We deem this a reasonable and practical abstraction.

The remainder of this chapter is organized as follows. Section 10.2 recalls existing approaches to quantify anonymity and privacy in databases and communication systems as well as generalizations. Since none of these metrics is designed to consider value over time, inspiration is sought from financial mathematics. We briefly review existing adaptations of quantitative financial methods to information security before we present our notion of *privacy options* in Sect. 10.3, the 'heart' of this chapter. Section 10.4 implements the ideas in a concrete proposal to model the two relevant quantities as independent stochastic processes: a state-space model is suggested for individual attribute value transitions (Sect. 10.4.1), and a binomial random walk serves as proxy for the distribution of attribute values in the population (Sect. 10.4.2). We combine both components to a valuation method in Sect. 10.5 and interpret the results in Sect. 13.5. The concluding Sect. 10.7 sketches future directions.

10.2 Related Work

We have identified two areas of relevant prior art. First, measurement of privacy with information theory and probability calculus has some tradition as a sub-field of computer science [25]. Section 10.2.1 briefly reviews this string of research. Second, another set of relevant publications are prior attempts to adopt quantitative methods from finance to information security and privacy. These are summarized in Sect. 10.2.2.

10.2.1 Measurement of Anonymity and Unlinkability

Measuring *anonymity* with information theory was—to the best of our knowledge—first motivated in the 1980s after a public debate about the census in Germany.[1] Fischer-Hübner [16, 17] uses the entropy of attributes (columns) in a database, for instance demographic data in a census survey, to measure their average information. This way, it is possible to compute the average number of records in the database that would match a given set of attributes. The degree of anonymity (or the "risk of re-identification" in [17]) is the reciprocal of this number of records. Attempts to

[1] *Confidentiality* in statistical databases has a much longer research track, e. g., [13, 35].

measure anonymity in statistical databases [42] have led to a number of combinatorial metrics, most prominently *k*-anonymity [40].

Aside from statistical databases, benchmarking anonymous communication systems has stirred a need for research on privacy metrics. Díaz et al. [14] as well as Serjantov and Danezis [36] propose Shannon entropy [37] to measure the uncertainty of an outside observer about the assignment of users to roles (sender, recipient, uninvolved) in a communication system. Shannon entropy quantifies the amount of additional information an observer would need in order to unanimously identify the role of the user. From this metric, it is possible to calculate the average size of the *anonymity set* [33] an anonymous communication system can provide. The larger the entropy the more information is effectively concealed from the observer, and hence the more anonymous the users of a system are. By contrast, Tóth et al. [41] point out that even if a communication system provides a reasonable degree of anonymity *on average*, the probability for *a single user* of being identifiable can still be unacceptably high. Therefore Tóth et al. define an upper bound for the probability of identification as *degree of anonymity*, which no user must exceed [41].

Another modification is to relax the strict focus on communication systems and model *unlinkability* between two arbitrary items [33, 39]. This view has been taken up for example by Clauß [10], who approximates unlinkability measures in a model world where each data subject's *identity* is defined by a set of finite discrete attributes. Only part of their values may be known to an outside observer. So a data disclosure decision effectively deals with the problem of whether or not an additional attribute value (previously unknown to the observer) should be disclosed. Our model assumptions later in Sect. 10.4 are compatible with this stylized view of the world. Though not carried out in this chapter, our approach is extendable to joint unlinkability measures between more than two items. Obviously, there exist infinitely many projections that map the resulting probability space over the exponentially growing number of set partitions to a scalar. Specific instances of such projections with more [18] or less [15] clear information-theoretic interpretation have been proposed in the literature as concrete metrics of unlinkability.

Most of existing privacy metrics were conceived with the aim to compare between alternative technical systems. All methods have in common that the value of personal data is measured at a single point in time[2] and not account for its value in possible future states. When the area of application shifts from comparing systems to supporting individual disclosure decisions, this limitation prevails: existing metrics neglect the fairly accepted principle that so-called *adversaries* against one's informational privacy will never forget any information disclosed to them (see for instance [33]). As already outlined in the introduction, the inability of individuals to anticipate future states in disclosure decisions is named as the main reason to explain partly puzzling results from laboratory experiments that try to measure people's valuation of personal data empirically [1, 5, 24].

[2] We are aware about only one commendable exception: a metric targeted to location-privacy [44], which accounts for changing locations over time.

10.2.2 Financial Methods in Information Security

Option pricing has its roots in financial mathematics and deals with finding the 'fair' price for contracts that allow their holders to choose between a security and a fixed amount of money at a future point in time. The field has grown rapidly since the seminal work by the meanwhile Nobel laureates Black, Scholes and Merton [6, 28] was published in the 1970s. Financial options became a popular tool for risk managers because they allow portfolio managers to 'hedge' idiosyncratic risks on financial markets, that is to shape the distribution of possible outcomes in sophisticated ways and thereby adjust it to the investor's risk appetite. But the idea soon spread to other domains than marketable securities. So-called *real options* have been proposed to gauge investment decision, in particular in project management [3]. They are tools to model project risk and opportunities with sound financial valuation methods to compare between alternatives. One advantage of real options in project management is the possibility to anticipate midcourse strategy corrections to react to uncertain future states.

Several authors have proposed to apply real options to information security investment [12, 20, 23, 26] to complement other accounting metrics, such as return on information security investment (ROSI) and annual loss expectancy (ALE) [19, 34, 38]. Interestingly, Gordon et al. [20] use real options to criticize security overinvestment, whereas Daneva [12] makes the case for higher spending.

Other applications of financial methods on specific information security problems include Matsuura's [27] option pricing approach to model the value of what he calls *digital security token*. These tokens can be thought of as media objects with attached protection, as suggested in the context of digital rights management (DRM). In [8], we have adapted the idea of prediction markets [43] to fix incentives in software vulnerability disclosure with so-called *exploit derivatives*. Ozment [30] has tackled vulnerability disclosure with auction theory.

To the best of our knowledge, this work is the first to apply option pricing theory to informational privacy. Neither are we aware of any work in other domains that suggests financial derivatives written on information measures (in Shannon's sense [37]) as underlying.

10.3 From Financial to Privacy Options

The key idea of this work is that disclosing a single attribute value can be interpreted as writing an option for exploiting the attribute in the future. Here, 'to exploit' refers to the act of using the attribute to draw inference on the data subject's identity or preference, and to base decisions on this information that may affect the data subject. One prominent example brought forward by Odlyzko [29] and Acquisti and Varian [2] is price discrimination in buyer–seller relationships. Thus, the data subject who discloses an attribute value thereby writes an option, whereas the transaction counterpart buys an option to use the information for decision-making. We

follow the convention in the information security literature and further refer to the transaction counterpart as *adversary*. This term reflects a convention and should not be interpreted as an adoption of the normative view that collecting personal data is necessarily hostile or evil.

Most elements of financial option pricing theory have direct correspondences in our notion of *privacy options*.

The *currency* in which privacy options are denominated is *information* in Shannon's [37] sense. Knowing an attribute value (i. e., holding the option), if valid, helps to reduce the uncertainty of the adversary about the identity of the data subject. The means to express uncertainty in information theory is entropy and the contribution of the attribute value has information value. The *unit* of information is *bits*.

The *underlying asset* of privacy options is the disclosed attribute value and the *market price* corresponds to the information (in Shannon's sense) which the adversary gains from the attribute value by exploiting it.

The privacy option is a *call option*, in which the data subject takes a *short position*. The asset, that is the attribute value, is handed over to the adversary (*long position*) at the time of the option purchase. The action that may be performed by the adversary is *exploiting* the underlying attribute value rather than *buying* the underlying security.

The correspondence to the *premium* is the compensation the adversary has to pay in return for the attribute value. However, this compensation is not necessarily denominated in the currency 'information'. For example, a merchant could offer a small rebate to the sales price to incentivize the use of loyalty cards from which personal data can be collected. This way, empirical measurements of this monetary premium, such as in [21, 24], could be linked to information-theoretic quantities by calibrating information-utility functions.

The increasing uncertainty about the linkability of the attribute value to the data subject can be interpreted as *interest rate* of an alternative investment: the probability of a valid link between the disclosed attribute value and the data subject decreases with the time elapsed since the disclosure of an attribute value. The value of the option decreases proportionately to the probability of a valid link because this linkability determines whether the adversary can benefit from the option at all.

Analogies also exists for the distinction of the two vanilla option styles, i. e., the *American* option and the *European* option. The difference between both styles is the time period in which the option may be exercised. An American option may be exercised at any time starting from the purchase of the option until it expires. This applies to the situation where a service provider does not depend on the assistance of the data subject for exploiting the data after the data subject has once disclosed its attribute value. An European option may only be exercised at the date of expiry. This applies to situations where the benefit for the adversary depends on some action of the data subject. For example, a personalized purchase history is only valuable to a seller if (and when) the data subject decides to revisit his store [2, 9].

Other elements of financial options do not have direct correspondences in our notion of privacy options developed in this chapter. *Put options* are impractical since 'negative information' does not exist. They could, however, make sense in special

(and largely hypothetical) cases where deletion of previously disclosed personal data can be enforced [7]. Due to the non-rivalrous nature of information goods, we were also unable to conceive a correspondence to *dividends* of the attributes underlying our privacy options. Finally, the *strike price* (or *exercise price*) is the amount of money to be paid when the option is actually exercised. If exploiting the attribute value does not depend on other attributes, then there is no way to enforce a transfer of money or information, hence the strike price is always zero. One can conceive to change this by introducing a trusted third party who acts as an information broker, or by allowing for partial disclosure of multiple dependent attributes. Another interpretation for the strike price is the effort of the adversary to retrieve the personal data at the time of exploitation. It may vary with organizational and technical factors, but it is largely determined by the adversary and not—like for financial options—by the contract itself. All this highlights that there is room for further extension of the analogy, though they are clearly beyond the scope of this chapter.

10.4 Sources of Uncertainty

In this section, we specify models for each source of uncertainty. In order to keep the calculations tractable, we model the two sources of uncertainty as *independent* stochastic processes; more precisely, the *timed linkability process* for attribute value changes of a single data subject (microscopic view, Sect. 10.4.1), and another stochastic process that drives the *distribution of attribute values in the population* (macroscopic view, Sect. 10.4.2). The latter model has many similarities with simple models of asset value fluctuations in financial option pricing.

10.4.1 Micro Model: Timed Linkability Process

Attribute values that have just been disclosed by a data subject are linkable to the data subject by the adversary. Here, we do not consider misinformation and thus assume links to be valid as long as the data subject does not change—intentionally or unintentionally—to another attribute value. We further assume that it is generally possible to change attribute values, however, the actual change, particularly its time and the new value, is not observable by the adversary.

This suggests modeling the attribute values over time as a stochastic process. The process can be expressed in a (discrete time-invariant) state-space model without inputs nor outputs. The state vector $\mathbf{x}(t)$ contains the probability of a valid link in the first element and the probability of an invalid link in the second element. The next state $\mathbf{x}(t+1)$ of this state-space model is defined in a recursive manner depending on the current state $\mathbf{x}(t)$ and a state transition matrix $\mathbf{A}$,

$$\mathbf{x}(t+1) = \mathbf{A}\mathbf{x}(t)\,. \tag{10.1}$$

Elements $a_{i,j}$ of $\mathbf{A}$ hold the probability of a state change from state j to state i. The absence of inputs allows us to simplify the model and use matrix multiplication instead of recursion to calculate a particular $\mathbf{x}(t+1)$,

$$\mathbf{x}(t+1) = \mathbf{A}^{t+1}\mathbf{x}(0)\ . \tag{10.2}$$

In the simplest case, the state matrix $\mathbf{A}$ has dimension 2×2 and is defined by only two probabilities, p and $\bar{p}$. Let p be the probability that the data subject keeps its linkable attribute value and $\bar{p}$ be the probability that a data subject, who once changed the attribute value to something unlinkable, does not revert to the linkable attribute value. The state vector $\mathbf{x}(0)$ at the time of disclosure is

$$\mathbf{x}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}\ . \tag{10.3}$$

The first element of vector $\mathbf{x}(0)$ holds the initial probability of linkability, which equals 1 by definition: the attribute value is definitely linkable when it has just been disclosed. Accordingly, we define the state matrix $\mathbf{A}$ as

$$\mathbf{A} = \begin{pmatrix} p & 1-\bar{p} \\ 1-p & \bar{p} \end{pmatrix}\ . \tag{10.4}$$

This allows us to model time aspects of attribute value changes. If, for instance, the attribute describes the attribute haircut and its value is *ponytail*, then the attribute might change instantly to any other value that describes a shorter haircut, but, naturally, hair cannot grow as fast as it can be cut off. And thus the probability of reverting back to *ponytail* is limited by a natural upper bound. Assume that the probability of keeping that haircut would be fairly high. Then Fig. 10.1 illustrates a hypothetical development of the probability of linkability over time. The functional form of Eq. (10.2) imposes an exponential decay.

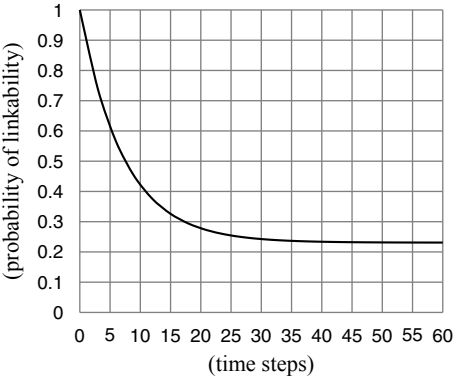


Fig. 10.1 Development of the probability of linkability if both the probability p of keeping the disclosed attribute value and the probability $\bar{p}$ of staying with anotnher attribute value are high. The diagram shows 60 time steps for $p = 0.9$ and $\bar{p} = 0.97$.

Other attributes might follow different processes, say, two attribute values and whenever the attribute has taken one value, the data subject tends to choose the other one with high probability. One can think of this as a model of fashions that alternate every couple of years. Thus, after the disclosure of the attribute value, it is possible to predict the values in the future, but with exponentially decreasing certainty. Fig. 10.2 depicts such a setting.

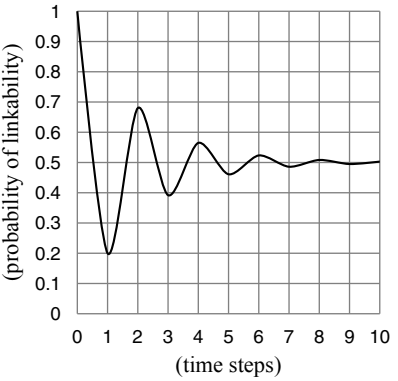


Fig. 10.2 Development of the probability of linkability if both the probability p of keeping the disclosed attribute value and the probability $\overline{p}$ of staying with another attribute value is small. The diagram shows 10 time steps for $p = 0.2$ and $\overline{p} = 0.2$.

Yet another situation emerges for attributes such as passport numbers: there is a vast number of different attribute values. The probability of requesting a new passport and therefore changing the attribute value might be small, depending on the travel habits of the data subject and on constraints imposed by the issuing country. But the probability of reverting back to exactly the same passport number is negligibly small. If we assume that this probability is in fact zero, then is it easy to see that the probability of linkability in the state-space model reduces to an exponential function of p, since for $\overline{p} = 1$, it holds that (after t time steps)

$$\mathbf{x}(t) = \mathbf{A}^t \mathbf{x}(0) = \begin{pmatrix} p & 0 \\ 1-p & 1 \end{pmatrix}^t \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} p^t & 0 \\ 1-p^t & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} p^t \\ 1-p^t \end{pmatrix} . \qquad (10.5)$$

In Fig. 10.3, we show an example for the development of linkability, if $\overline{p} = 1$.

Note that generalizations to higher-order state-space models are possible and can be useful to represent other than binary attributes. We defer discussion of and examples for this case to future work.

10.4.2 Macro Model: Population Development

In the population, individual data subjects can be distinguished by their attribute values. A metric for the average discernibility is the self-information of an attribute value in the population. In terms of Shannon's information theory, the attribute can

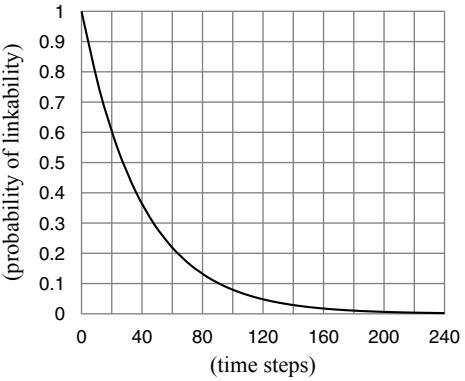


Fig. 10.3 Development of the probability of linkability if the probability $1 - \bar{p}$ of returning to the same attribute value after a change is zero. The diagram shows 240 time steps for $p = 0.975$ and $\bar{p} = 1$.

be understood as source of information, the attribute values as alphabet, and the (relative) frequency of each attribute value as the probability of the symbol. Let v be an attribute value and r_v be the relative frequency of this value in the population, then

$$H_v = -\log_2 r_v \tag{10.6}$$

is the self-information of the attribute value. It expresses the amount of information conveyed by the attribute value to the adversary, who may exploit it to re-identify the data subject.

The relative frequencies vary over time depending on how the individual data subjects change their attribute values. Thus, also the self-information of each attribute value fluctuates. The straight approach of modeling the behavior of *all* data subjects, their attribute values, and the changes over time by an aggregation of many micro-level models would be analytically intractable and computationally demanding. Moreover, generalizing one fixed state-space model of Sect. 10.4.1 to all data subjects neglects possible heterogeneity between them and is therefore debatable with theoretical arguments. Instead, we model the macroscopic changes of the distribution of attribute values in the population as a separate stochastic process.

A similar approach is taken in financial option pricing, where the market price of the underlying asset can be modeled in a similar way [11]. Both the market price and the relative frequency of the attribute value can *move up* or *down* in each single time step. This is in line with our notion that the attribute value corresponds to the underlying asset in option pricing, and the self-information, as a function of the relative frequency, can be understood as a price denoted in self-information as currency. Figure 10.4 shows a single time step of that process. The uncertainty about

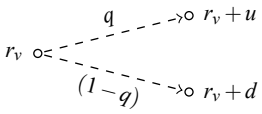


Fig. 10.4 Single time step in the development of self-information, analogous to the market price development in financial option pricing.

an increase or decrease of the relative frequency r_v is modeled by step size u (*upward move*, increase of the relative frequency) and by the probability q of an increase. Correspondingly, a *downward move* can be modeled by adding d. In line with [11], we assume u and d are chosen such that

$$d = -u \,. \tag{10.7}$$

Thus, all possible developments of the frequency for a fixed number of time steps form a lattice similar to the pricing lattice in Binomial Option Pricing. An example lattice is displayed in Fig. 10.5.

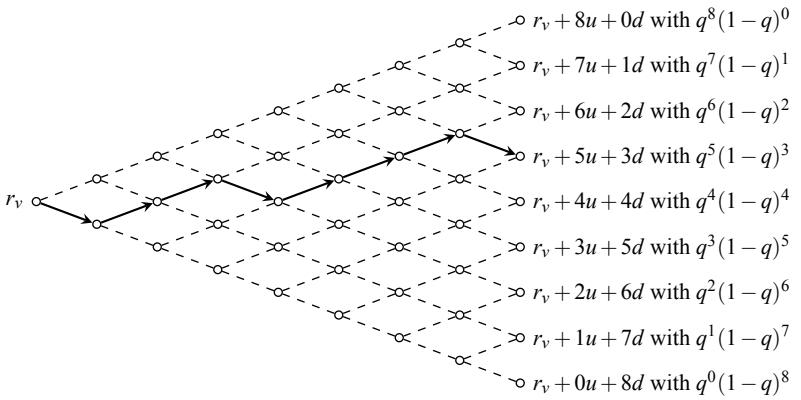


Fig. 10.5 Population model: discrete binomial stochastic process (random walk). The diagram shows all possible result states for the development of the relative frequency of an attribute value, starting with r_v, and their probabilities after eight time steps.

Even though this model is pretty simple, we can capture almost all developments of the frequency as long as we are able to choose the time steps small enough. A stagnation, for instance, can be represented by alternating up and down movements. Linear upward or downward trends of arbitrary strength can be modeled intuitively by combining upward or downward movements, respectively, with stagnation.

However, a direct analogy between market price and frequency development is not fully adequate, since market prices could increase without upper bound, but the relative frequency is defined only between zero and one. One way of dealing with the bounds would be forcing the next step of the random walk in a fixed direction, if the other direction led beyond a bound. This would simulate a stagnation at the margins of the domain. However, the approach has several drawbacks. For instance, once the upper bound is reached, any number of further upward movements would have exactly the same total effect as no further upward movements at all. In order to avoid that, we propose to transform the bounded domain of the relative frequency to an unbounded domain for the random walk. We have chosen the logit function for this transformation,

$$\operatorname{logit}(x) = \log \frac{x}{1-x} \,. \tag{10.8}$$

After running the random walk in the logit-transformed domain, we transform the value back to the frequency domain by means of the inverse logit function $\operatorname{logit}^{-1}$,

$$\operatorname{logit}^{-1}(x) = \frac{\mathrm{e}^x}{1+\mathrm{e}^x} \,. \tag{10.9}$$

The transformation to the unbounded domain allows us to rely on the same lattice process as known from Binomial Option Pricing. After the logit transformation, any number of movements in one direction is possible and exactly the same amount of movements in the other direction is necessary for compensation. Independent of the number of upward or downward moves, the outcome will remain within the bounds after the inverse transformation. Another nice property of the logit transformation is that the absolute changes in the relative frequency are the smaller the closer the level approaches the domain bounds. This captures a kind of base effect of very persistent individuals, who can be found in most heterogeneous populations.

The *information value* of an attribute value that will be exploited after $T > 0$ time steps can be computed by averaging the self-information over all possible relative frequencies, weighted with their respected probability of occurrence (right-hand side in Fig. 10.5). With $Q(n)$ being the probability of n upward moves,

$$Q(n) = \binom{T}{n} \cdot q^n (1-q)^{T-n} \,, \tag{10.10}$$

and $r_v^{(n)}$ being the relative frequency, taken from the result of the random walk with n upward moves,

$$\begin{aligned} r_v^{(n)} &= \operatorname{logit}^{-1}\Big[\operatorname{logit}(r_v) + nu + (T-n)d\Big] \\ &= \operatorname{logit}^{-1}\Big[\operatorname{logit}(r_v) + (2n-T)u\Big] \,, \end{aligned} \tag{10.11}$$

the expected self-information of the entire stochastic process after T steps is $\mathscr{H}_v(T)$:

$$\mathscr{H}_v(T) = -\sum_{n=0}^{T} Q(n) \log_2 r_v^{(n)} \,. \tag{10.12}$$

This measure of expected self-information accounts for fluctuations over time that are caused more generally by the society or the population, respectively, rather than by the individual data subject. Knowledge about an *attribute value* is the more valuable the higher the *self-information* of the attribute value becomes in the future and thus the smaller its relative frequency becomes in the population. Generalizing one step, knowledge about an *attribute* is less valuable the higher the *entropy* of the attribute is expected to grow (or remain) in the future.

Fig. 10.8 shows the development of a downward trend in a lattice diagram. Imagine an adversary who exploits technical attributes, such as browser or operating

system, of data subjects for re-identification. The parameters to be plugged into the process could be estimated from the dynamics of the market share of web browsers or operating systems in the population. The hypothetical development shows a clear downward trend in the market share of one particular browser, which had a dominant share before ($r_v = 0.7$). The downward trend might be due to data subjects switching to a competing alternative browser. The fewer data subjects use the formerly dominant browser, the higher is the value of the information that a specific data subject to be identified uses this particular browser. Thus, the expected self-information increases over time. Assuming that sufficiently accurate parameters can be estimated from historical observations, scaled down to a single time step, and predicted to remain valid for the next 100 steps, then we can continue the lattice shown in Fig. 10.8 in order to calculate the expected self-information after 100 steps. The development of the self-information for that time period is shown in Fig. 10.6.

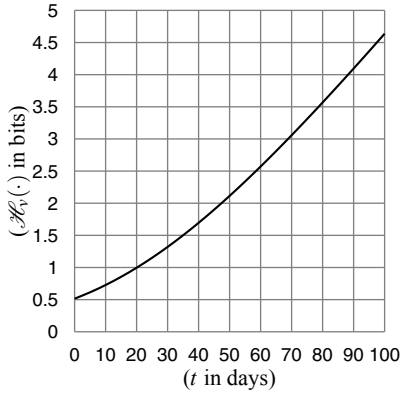


Fig. 10.6 Trend development of the expected self-information $\mathscr{H}_v(\cdot)$ for an attribute value with an initial relative frequency of the attribute value in the population $r_v = 0.7$, the probability of an increase $q = 0.3$, and the step size parameter $u = 0.1$. States that can be reached by a random walk in the first five steps are illustrated in Fig. 10.8.

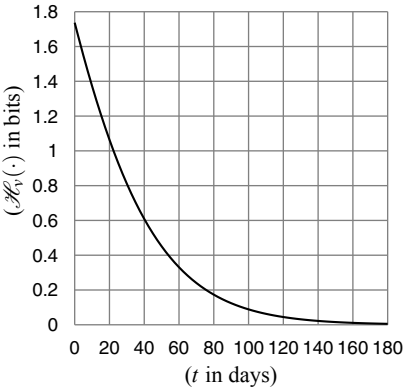


Fig. 10.7 Trend development of the expected self-information $\mathscr{H}_v(\cdot)$ for an attribute value. The parameters $r_v = 0.3$, $q = 0.7$, and $u = 0.1$ are chosen such that a positive trend can be observed for the attribute value. States that can be reached by a random walk in the first five steps are illustrated in Fig. 10.9.

Similarly, a browser or an operating system which has previously been used by a minority in the population ($r_v = 0.3$) may quickly become popular ($q = 0.7$). And therefore, the attribute value soon applies to a majority in the population. Thus, the expected self-information of that attribute value will decrease over time. We have outlined the first five steps in a lattice again, see Fig. 10.9, and continued the next 180 steps of the expected self-information in a diagram, see Fig. 10.7.

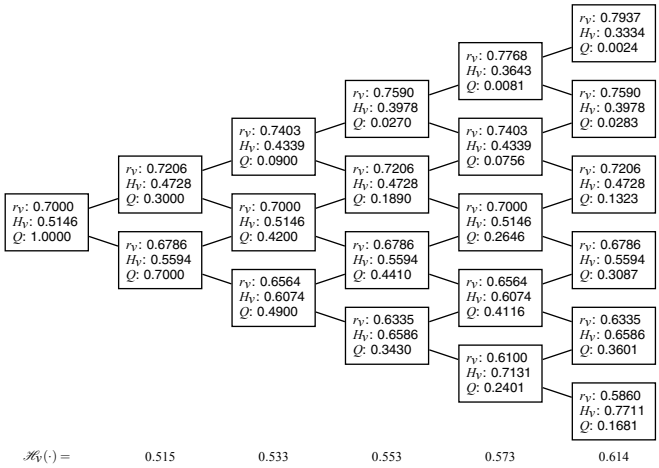


Fig. 10.8 First five steps of the development of the self-information with the parameters as described in Fig. 10.6. Each box represents a possible intermediate step of the random walk and for each step r_v denotes the relative frequency, H_v denotes the self-information, and Q denotes the probability. The expected self-information $\mathscr{H}_v(\cdot)$ after each time step is printed below each column of the lattice.

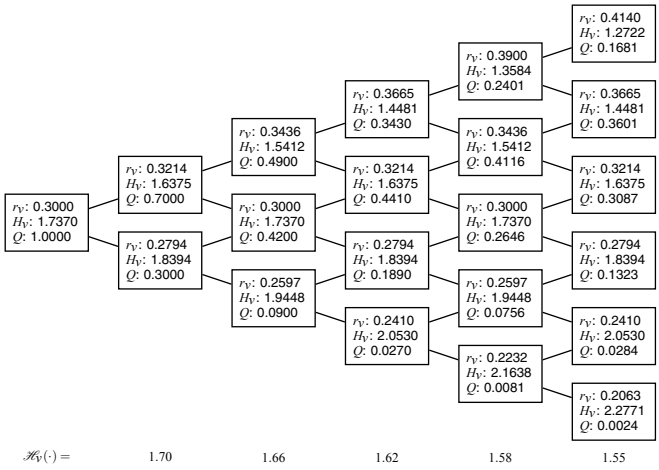


Fig. 10.9 First five steps of the development of the self-information for a positive trend development as described in Fig. 10.7. The notation is the same as in Fig. 10.8.

10.5 Valuation of Privacy Options

The main observation underlying our notion of privacy options is that disclosure of personal data and its exploitation does often not take place at the same point in time. In the previous section, we have argued that two sources of uncertainty drive the valuation of privacy options, and both can be modeled as independent stochastic processes. Now we will show how to combine the processes on the micro and macro level to obtain an inter-temporal measure of the value of personal data disclosure.

It is intuitively clear that the value of personal data (for re-identification of the corresponding data subject) depends on the self-information at the time of exploitation. The value is the lower, the lower the probability of a link between the data (i. e., attribute value) and the data subject is at that time. Thus, the probability of the link, modeled by Eq. (10.2) of the micro model, discounts the value of the (European) privacy option $\mathscr{V}_{\mathrm{Eu}}(T)$ at time T,

$$\mathscr{V}_{\mathrm{Eu}}(T) = x_1(T) \cdot \mathscr{H}_v(T) \,. \tag{10.13}$$

Recall from Eq. (10.3) that $x_1(T)$ denotes the first element of vector $\mathbf{x}(T)$, which holds the probability of a valid link.

The value of a privacy option depends on the parameters for the linkability model, namely the probabilities p and $\bar{p}$, and the parameters of the population development, namely the current relative frequency r_v of the disclosed attribute value, the probability of an upward movement in the random walk q, the step size u of an upward movement in the random walk, and the exercising time T of the option. For example, consider a privacy option with the parameters

$$\begin{aligned} p &= 0.95 \,, \quad & \bar{p} &= 1 \,, \\ r_v &= 0.5 \,, \quad & q &= 0.5 \,, \\ u &= 1.2 \,, \quad & T &= 100 \,. \end{aligned} \tag{10.14}$$

Observe in Fig. 14.4 that there is a substantial difference between the current value of personal data, i. e., the attribute value, and its information value for re-identification after several time steps. Further, it would be best to exercise the privacy option after seven time steps. Before reaching the seventh time step, the value of the option (solid line) is dominated by the increasing self-information of the attribute value (dotted line). Afterwards, the value diminishes due to the decreasing probability of a link between attribute value and data subject (dashed line).

$\mathscr{V}_{\mathrm{Eu}}(T)$ is the value of the privacy option, if it is exploited in the 'European' style, that is, the option can only be exercised when it expires. By contrast, American options can be exercised at any time between purchase and expiry. For privacy options, this means that personal data can be exploited at any time between disclosure and, for instance, the date of an obligation to erase the data. One can even think of the data being exploited more than once in the period of time. However, to allow for a better comparison, we normalize the valuation to exactly one exploitation. Thus, the value of an American privacy option $\mathscr{V}_{\mathrm{Am}}(T)$ is the average of the expected value at

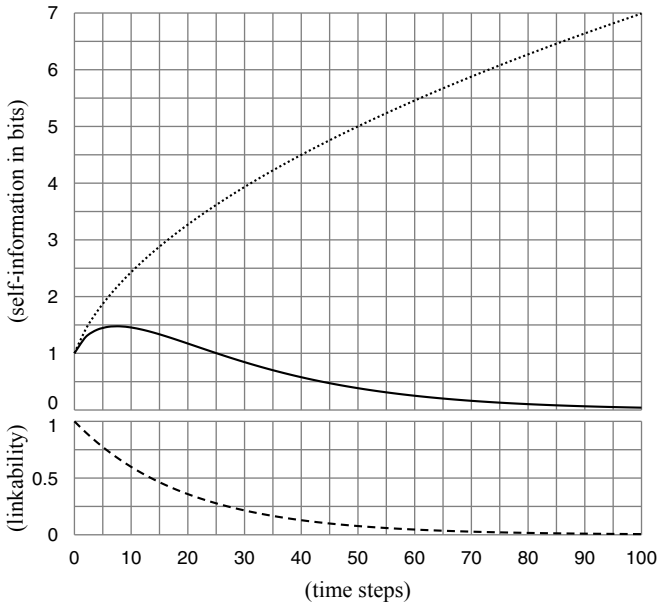


Fig. 10.10 Development of the privacy option value (solid line) with unit *bits*. The dashed line shows the corresponding probability of linkability with low probability of a change of the attribute value ($p = 0.95$), but once the attribute value has been changed, the previous linkable attribute value will never be recovered ($\bar{p} = 1$). The dotted line shows the development of the expected self-information of the attribute value (in *bits*). The distribution of the attribute value is assumed to remain the same ($r_v = 0.5$, $q = 0.5$), but the dispersion is high ($u = 1.2$).

each point in time between data disclosure and expiry of the option,

$$\mathscr{V}_{\mathrm{Am}}(T) = \frac{1}{T} \sum_{t=0}^{T} x_1(t) \cdot \mathscr{H}_v(t) \; . \tag{10.15}$$

If the attribute value is exploited several times between the attribute disclosure and the expiry of the privacy option, say, k denotes the number of exploits, then the value of the option is $\mathscr{V}_{\mathrm{Am}}(T)$, multiplied by k. One can also consider a weighted average to reflect a given prior on the possible time of exploitation.

10.6 Discussion of Results

A common assumption of privacy measures in the literature is that personal data, once disclosed, reduces the informational privacy of the data subject by its present self-information. Implicitly, this implies that the present self-information of disclosed data remain constant over time, at least until the data is exploited.

Table 10.1 Valuation of privacy options. Comparison of inter-temporal valuation with the self-information of the attribute value at present.

assumptions			valuation (in bits)			over/ under-valuation indicator[a]
stochastic process		option	point in time		time range	
linkability	population dev.	expiry date	present	future[b]	present–future[c]	
Fig. 10.1 (page 194)	Fig. 10.6 (page 199)	0	0.515			
		10		0.307	0.430	↘/↘
		25		0.293	0.342	↘/↘
		50		0.489	0.363	↘/↘
		100		1.070	0.569	↗/↗
Fig. 10.3 (page 196)	Fig. 10.6 (page 199)	0	0.515			
		10		0.565	0.595	↗/↗
		25		0.611	0.594	↗/↗
		50		0.596	0.603	↗/↗
		100		0.369	0.546	↘/↗
Fig. 10.1 (page 194)	Fig. 10.7 (page 199)	0	1.737			
		10		0.579	1.136	↘/↘
		25		0.237	0.664	↘/↘
		50		0.104	0.410	↘/↘
		100		0.020	0.231	↘/↘
Fig. 10.3 (page 196)	Fig. 10.7 (page 199)	0	1.737			
		10		1.066	1.517	↘/↘
		25		0.494	1.043	↘/↘
		50		0.127	0.654	↘/↘
		100		0.007	0.347	↘/↘

[a] Comparison between the present value, i. e., the present self-information of the attribute value, and the privacy option value, i. e., the expected self-information at a point in time, discounted by the probability of a link between attribute value and data subject. "↗" denotes that the present self-information underestimates the actual value, whereas "↘" denotes overestimation. The first arrow in this column refers to the "future" value and the other to the "present-to-future" value.

[b] This corresponds to a European option, which can be exercised at the expiry date.

[c] This is the value of the privacy option, if it is exercised at exactly *one* point in time between the date of disclosure and the expiry date. We assume that the point in time is randomly drawn from a uniform distribution.

Our examples show that the present self-information of personal data is only an appropriate measure for the information an adversary obtains when exploiting the data, if the disclosure and the exploit take place instantaneously. Otherwise, i. e., if time elapses between disclosure and exploit of personal data, the self-information at present can lead to both over- and underestimation of the 'true' value of the information passed over to the adversary.

Table 10.1 summarizes our findings by selected examples. It shows four privacy options derived from examples of the previous sections and their valuation with regard to expiry dates between 0 and 100. This corresponds to the situation where an attribute value is disclosed now and exploited either at the expiry date (column "future") or sometimes between now and the expiry date (column "present–future").

The resulting values of the privacy options, and thus the expected self-information of the underlying attribute value, is compared to the present self-information. Under- and over-valuations are indicated by arrows that point up or down, respectively.

In general, when the probability of a link between attribute value and data subject is uncertain, the value of personal data will be over-valuated by the present self-information, if the expected self-information is constant or decreasing over time. Undervaluations only occur, if an increasing expected self-information compensates the discount induced by the declining probability of linkability. In Table 10.1, this is the case for the first two privacy options, depending on the expiry date.

10.7 Conclusions and Outlook

In this chapter, we have motivated why and explained how option pricing theory can be useful for the valuation of informational privacy. In a first step towards this direction, we have proposed a very simple model that highlights the main features of our approach, namely the description of changes in each individual data subject's attribute values and the evolution of the distribution of attribute values in the population as two independent stochastic processes.

Once the realm of option pricing theory has been touched, possible extension and refinements are abundant. Most notably, it would be interesting to allow more than two attribute values in the state-space model, or to consider more than one attribute. This would not only allow to use the valuation results as guidance on which of a set of alternative attributes should be disclosed (if the data subject has a choice), but also to extract the self-information of combinations of attributes over time. Another obvious next step is to replace the binomial process with more appropriate processes. Ideally these processes should be validated with and calibrated to empirical data, e. g., from longitudinal population surveys. Replacing the discrete-time process with a continuous-time process could bring our model closer to (variants of) the Black–Scholes [6] formula, which promise closed-form solutions. This avoids computational effort when the number of time steps grows large (though at the price of additional assumptions). While the analysis in this chapter was strictly confined to expected values, one could also calculate and interpret other summary measures of the distribution functions over time. In particular small quantiles could be interesting to study (un)linkability with a security or risk management mindset by regarding the ϵ-worst case.

But there is more than just tweaks in the proposed framework: implementing true and conscious control of personal data in everyday social interactions is generally difficult. The fact that more and more social interactions happen in the digital sphere aggravates this problem substantially. Following in Baran's [4] footsteps, ideas of comprehensive privacy-enhancing technologies (PETs) have been conceived. Their vision is to cure the problems created by technology with more technology. So-called privacy-enhanced identity management is envisaged to assist people on deciding if, when, which, and at what price personal data should be disclosed. As

with every decision support system, this implies that several alternatives have to be evaluated and compared more or less automatically. And since most interactions do have consequences for the future, this evaluation would be incomplete if it does not consider time [22]. So privacy-enhancing technologies are an obvious field of application for our framework.

Existing blueprints for such PETs use so-called *privacy policies* to define how personal data should be handled (although enforcement of such policies against realistic adversaries is largely unsolved). Ideally, privacy policies are formulated in formal languages, which should support complex enough semantics to capture all relevant aspects of personal data disclosure—including time. Interestingly, a similar problem exists for modern financial contracts: nested derivatives quickly create a complexity in semantics that is manually intractable. The solution, again, lies in the intersection between finance and computer science. For example, Peyton Jones [31, 32] has proposed domain-specific languages to model complex financial constructs and enable their valuation over time. This can be seen as a generalization of classical option pricing theory. An interesting direction for future research is to adapt this to privacy policies and develop a formal language that can express aspects of time, and thereby generalize the valuation framework presented here.

Beyond direct applications in informational privacy protection through data avoidance, measuring the inter-temporal value of attribute values for linkability could also be useful in other contexts, even with opposite sign. It is conceivable that the data subject seeks to disclose as much information as possible to ensure clear identification in the future. This perspective will most likely matter when communicating bandwidth for attribute values is a scarce resource (e. g., through a hidden channel) and one must select those attributes which will be most informative later on. Moreover, although the exposition in this chapter was framed from the data subjects' perspective and targeted to protecting their personal data, the very same underlying ideas and valuation methods can also be useful for businesses to estimate the value of their customer databases. This is generally considered a hard task due to the intangible nature of personal data, so a new perspective might stimulate further advances in this area, too.

To conclude, although the idea of valuating privacy with option pricing theory sounds intriguing on paper, we have to recall that this framework is in no way a panacea. Many obstacles ignored in this exposition are likely to remain as serious limitations in practice: complexity, measurement problems, heterogeneous preferences, model mismatch, and bounded rationality, among others. So the confidence bands of our privacy metrics will most likely be loose in practice, but having a theoretically founded measurement method which can deliver some point estimates is certainly better than nothing at all.

Acknowledgments

This chapter incorporates some valuable comments by the anonymous reviewers for WEIS 2009. The first author was partially funded by the Research Council of Norway through the PETweb II project. The second author was supported by a post-doctoral fellowship of the German Academic Exchange Service (DAAD). Research leading to these results has also received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement No. 216483.

References

1. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security and Privacy **3**(1), 26–33 (2005)
2. Acquisti, A., Varian, H.R.: Conditioning prices on purchase history. Marketing Science **24**(3), 1–15 (2005)
3. Amram, M., Kulatilaka, N.: Real Options: Managing Strategic Investment in an Uncertain World. Harvard Business School Press (1999)
4. Baran, P.: Communications, computers and people. Tech. rep., RAND Corporation, Santa Monica, CA (1965)
5. Berendt, B., Günther, O., Spiekermann, S.: Privacy in e-commerce: Stated preferences vs. actual behavior. Communications of the ACM **48**(4), 101–106 (2005)
6. Black, F., Scholes, M.: The pricing of options and corporate liabilities. Journal of Political Economy **81**, 637–654 (1973)
7. Blanchette, J.F., Johnson, D.G.: Data retention and the panoptic society: The social benefits of forgetfulness. Information Society **18**(1), 33–45 (2002)
8. Böhme, R.: A comparison of market approaches to software vulnerability disclosure. In: G. Müller (ed.) Emerging Trends in Information and Communication Security (Proc. of ETRICS), *LNCS*, vol. 3995, pp. 298–311. Springer, Berlin Heidelberg (2006)
9. Böhme, R., Koble, S.: Pricing strategies in electronic marketplaces with privacy-enhancing technologies. Wirtschaftsinformatik **49**(1), 16–25 (2007)
10. Clauß, S.: A framework for quantification of linkability within a privacy-enhancing identity management system. In: G. Müller (ed.) Emerging Trends in Information and Communication Security (ETRICS), *LNCS*, vol. 3995, pp. 191–205. Springer, Berlin Heidelberg (2006)
11. Cox, J., Ross, S., Rubinstein, M.: Option pricing: A simplified approach. Journal of Financial Economics (1979)
12. Daneva, M.: Applying real options thinking to information security in networked organizations. Tech. Rep. TR-CTIT-06-11, Centre for Telematics and Information Technology, University of Twente, Enschede, NL (2006)
13. Denning, D.E., Denning, P.J., Schwart, M.D.: The tracker: A threat to statistical database security. ACM Trans. on Database Systems **4**(1), 76–96 (1979)
14. Díaz, C., Seys, S., Claessens, J., Preneel, B.: Towards measuring anonymity. In: P. Syverson, R. Dingledine (eds.) Workshop on Privacy Enhancing Technologies, *LNCS*, vol. 2482. Springer, Berlin Heidelberg (2002)
15. Fischer, L., Katzenbeisser, S., Eckert, C.: Measuring unlinkability revisited. In: Proc. of Workshop on Privacy in the Electronic Society (WPES), pp. 105–109. ACM Press, New York (2008)
16. Fischer-Hübner, S.: Zur reidentifikationssicheren statistischen Auswertung personenbezogener Daten in staatlichen Datenbanken *[Towards reidentification-secure statistical data analysis of personal data in governmental databases]*. Diploma thesis, Universität Hamburg (1987). In German
17. Fischer-Hübner, S.: IT-security and privacy: Design and use of privacy-enhancing security mechanisms, *LNCS*, vol. 1958. Springer, Berlin Heidelberg (2001)

18. Franz, M., Meyer, B., Pashalidis, A.: Attacking unlinkability: The importance of context. In: N. Borisov, P. Golle (eds.) Privacy Enhancing Technologies, *LNCS*, vol. 4776, pp. 1–16. Springer, Berlin Heidelberg (2007)
19. Gordon, L.A., Loeb, M.P.: The economics of information security investment. ACM Trans. on Information and System Security **5**(4), 438–457 (2002)
20. Gordon, L.A., Loeb, M.P., Lucyshyn, W.: Information security expenditures and real options: A wait-and-see approach. Computer Security Journal **14**(2), 1–7 (2003)
21. Grossklags, J., Acquisti, A.: When 25 cents is too much: An experiment on willingness-to-sell and willingness-to-protect personal information. In: Workshop of Economics and Information Security (WEIS). Carnegie Mellon University, Pittsburgh, PA (2007). `http://weis2007.econinfosec.org/papers/66.pdf`
22. Hansen, M., Pfitzmann, A., Steinbrecher, S.: Identity management throughout one's whole life. Information Security Technical Report **13**(2), 83–94 (2008)
23. Herath, H.S.B., Herath, T.C.: Investments in information security: A real options perspective with Bayesian postaudit. Journal of Management Information Systems **25**(3), 337–375 (2008)
24. Huberman, B.A., Adar, E., Fine, L.R.: Valuating privacy. IEEE Security and Privacy **3**(1), 22–25 (2005)
25. Kelly, D.J., Raines, R.A., Grimaila, M.R., Baldwin, R.O., Mullins, B.E.: A survey of state-of-the-art in anonymity metrics. In: Proc. of ACM Workshop on Network Data Anonymization (NDA), pp. 31–40. ACM Press, New York (2008)
26. Li, J., Su, X.: Making cost effective security decision with real option thinking. In: Proc. of International Conference on Software Engineering Advances (ICSEA 2007), pp. 14–22. IEEE Computer Society, Washington, DC, USA (2007)
27. Matsuura, K.: Security tokens and their derivatives. Tech. rep., Centre for Communications Systems Research (CCSR), University of Cambridge, UK (2001)
28. Merton, R.C.: Theory of rational option pricing. Bell Journal of Economics and Management Science **4**(1), 141–183 (1973)
29. Odlyzko, A.: Privacy, economics, and price discrimination on the Internet. In: N. Sadeh (ed.) ICEC2003: Fifth International Conference on Electronic Commerce, pp. 355–366 (2003)
30. Ozment, A.: Bug auctions: Vulnerability markets reconsidered. In: Workshop of Economics and Information Security (WEIS). University of Minnesota, Minneapolis, MN (2004). `http://www.dtc.umn.edu/weis2004/ozment.pdf`
31. Peyton Jones, S.: Composing contracts: An adventure in financial engineering. In: J.N. Oliveira, P. Zave (eds.) FME 2001: Formal Methods for Increasing Software Productivity, *LNCS*, vol. 2021. Springer, Berlin Heidelberg (2001)
32. Peyton Jones, S., Eber, J.M.: How to write a financial contract. In: J. Gibbons, O. de Moor (eds.) The Fun of Programming. Palgrave Macmillan (2003)
33. Pfitzmann, A., Hansen, M.: Anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management – A consolidated proposal for terminology. `http://dud.inf.tu-dresden.de/Anon_Terminology.shtml` (2008). (Version 0.31)
34. Purser, S.A.: Improving the ROI of the security management process. Computers & Security **23**, 542–546 (2004)
35. Schlörer, J.: Zum Problem der Anonymität der Befragten bei statistischen Datenbanken mit Dialogauswertung [On the problem of respondents' anonymity in statistical databases with dialogue analysis]. In: D. Siefkes (ed.) 4. GI-Jahrestagung, *LNCS*, vol. 26, pp. 502–511. Springer, Berlin Heidelberg (1975)
36. Serjantov, A., Danezis, G.: Towards an information theoretic metric for anonymity. In: P. Syverson, R. Dingledine (eds.) Workshop on Privacy Enhancing Technologies, *LNCS*, vol. 2482. Springer, Berlin Heidelberg (2002)
37. Shannon, C.E.: A mathematical theory of communications. Bell System Technical Journal **27**, 379–423, 623–656 (1948)
38. Soo Hoo, K.J.: How much is enough? A risk-management approach to computer security. In: Workshop on Economics and Information Security (WEIS). Berkeley, CA (2002). `http://www.sims.berkeley.edu/resources/affiliates/workshops/econsecurity/`
39. Steinbrecher, S., Köpsell, S.: Modelling unlinkability. In: R. Dingledine (ed.) Workshop on Privacy Enhancing Technologies, *LNCS*, vol. 2760, pp. 32–47. Springer, Berlin Heidelberg (2003)

40. Sweeney, L.: *k*-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **10**(5), 571–588 (2002)
41. Tóth, G., Hornák, Z., Vajda, F.: Measuring anonymity revisited. In: S. Liimatainen, T. Virtanen (eds.) Proc. of the Ninth Nordic Workshop on Secure IT Systems, pp. 85–90. Espoo, Finland (2004)
42. Willenborg, L., De Waal, T.: Statistical Disclosure Control in Practice. Springer, New York (1996)
43. Wolfers, J., Zitzewitz, E.: Prediction markets. Journal of Economic Perspectives **18**(2), 107–126 (2004)
44. Xiaoxin, W., Bertino, E.: Achieving *k*-anonymity in mobile and ad hoc networks. In: Proc. of IEEE ICNP Workshop on Secure Network Protocols, pp. 37–42. IEEE Press, New York (2005)

List of Symbols

General Symbols

v	attribute value
t	discrete point in time
T	total number of time steps
$\mathscr{V}_{\mathrm{Eu}}(T)$	value of a European privacy option at time T
$\mathscr{V}_{\mathrm{Am}}(T)$	value of a American privacy option at time T

Timed Linkability Process (Micro Model)

p	probability of an attribute which is currently linkable to a data subject to remain linkable in the next time step
$\bar{p}$	probability of an attribute which is currently not linkable to a data subject not to become linkable in the next time step
$\mathbf{x}(T)$	state vector in the state-space model at time T
$x_1(T)$	probability of a valid link between disclosed attribute value and data subject after T time steps
$\mathbf{A}$	state transition matrix of the state-space model
$a_{i,j}$	elements of $\mathbf{A}$

Population Development (Macro Model)

r_v	relative frequency of attribute value v in the population
H_v	self-information of attribute value v
q	probability of an upward move in the random walk
u	step size of an upward move in the random walk
d	step size of an downward move with $d = -u$
$Q(n)$	probability of n upward moves in T moves in total
$r_v^{(n)}$	relative frequency of attribute value v after randomly walking through T time steps of which n are upward moves
$\mathscr{H}_v(T)$	expected self-information of attribute value v after T time steps

Chapter 11
Optimal Timing of Information Security Investment: A Real Options Approach

Ken-ichi Tatsumi and Makoto Goto

Abstract This chapter applies real options analytic framework to firms' investment activity in information security technology and then a dynamic analysis of information security investment is explored by extending Gordon–Loeb (2002). The current research provides how firms have to respond to immediate or remote threat numerically. It shows that although positive drift of threat causes both larger and later investment expenditure, negative drift causes immediate investment and lower investment expenditure. The efficiency of vulnerability reduction technology encourages firms to invest earlier and induces cost reduction. To know the form of vulnerability is important because the effect of high vulnerability on timing and amount of the investment expenditure is mixed.

11.1 Introduction

Importance of information security has emerged very rapidly as information society has developed great deal. The information security investment has accordingly been considered by Gordon and Loeb in 2002. The highlight of their analysis is an introduction of vulnerability concept to formal optimization problem. Although Gordon–Loeb (2002) mentioned aspects of dynamics such as a first-mover advantage or the time value of money, their analysis is static and they did not consider any aspect of dynamic theory of information security at all. A dynamic analysis of information security investment is therefore explored in the following of this chapter, in terms of real options theory often used for the analytic tools of investment timing.

Ken-ichi Tatsumi
Faculty of Economics, Gakushuin University, Mejiro 1-5-1, Toshima-ku, Tokyo 171-8588, Japan; e-mail: Kenichi.Tatsumi@gakushuin.ac.jp

Makoto Goto
Graduate School of Economics and Business Administration, Hokkaido University, Kita 9, Nishi 7, Kita-ku, Sapporo 060-0809, Japan; e-mail: goto@econ.hokudai.ac.jp

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_11,

The chapter is organized as follows. First, Sect. 15.2 presents an outline of Gordon–Loeb model. Next in Sect. 15.3 we introduce real options theory that achieves the optimal timing of the investment level. In Sect. 15.4, we numerically calculate the optimal investment timing and level, and additionally some comparative statics. Then finally, Sect. 15.5 draws some conclusions and mentions directions for future works. We point out necessary extension of the model which captures an equilibrium nature of information security investments and needs to estimate the parameters of the dynamics.

11.2 Optimum Investment Size: The Model of Gordon and Loeb

In order to estimate the optimal level of information security investment for protecting some information system within a firm or an organization, Gordon–Loeb (2002) considers several variables and parameters of the system. We will utilize similar notation with a little change only for expositional purpose.

First, let L denote the potential loss associated with the threat against the information system, i.e. $L = T\lambda$, where T is a random variable of the threat occurring and λ is the (monetary) loss suffered on conditioned on the breach occurring. Further, let v denote vulnerability, i.e. the success probability of the attack once launched; vL is then the total expected loss associated with the threat against the information system.

If a firm invests z dollars in security, the *remaining vulnerability* will be denoted by $S(z,v)$. The expected benefit from the investment which is the reduction in the expected loss attributable to the investment can then be computed as $(v - S(z,v))L$, where $(v - S(z,v))$ is the reduction in the vulnerability of the information system. The expected net benefit can therefore be computed as $(v - S(z,v))L - z$. Under suitable differentiability assumptions (see the conditions A1–A3 below), we can see that the optimal level of investment can be found by computing the local optimum z^* of the expected net benefit, i.e. by solving the first order equation:

$$\frac{\partial[(v - S(z,v))L - z]}{\partial z} = 0, \tag{11.1}$$

and obtaining the following condition for $z^* = z^*(v)$:

$$-\frac{\partial S(z^*,v)L}{\partial z} = 1. \tag{11.2}$$

Of course, the remaining vulnerability function can not be arbitrary. Since $S(z,v)$ could be interpreted to be a probability, we must clearly have $0 \leq S(z,v) \leq 1$. Its first argument is an investment and the second one another probability, so that $0 \leq z$ and $0 \leq v \leq 1$. Besides that, the following restrictions are defined in Gordon–Loeb (2002):

A1 $\forall z, S(z,0) = 0$, i.e. if the attack success probability is 0, it stays so after every possible investment.

A2 $\forall v, S(0,v) = v$, i.e. if we spend no money for investment, there will be no change in the attack success probability.

A3 The function $S(z,v)$ is continuously twice differentiable and for $0 < v$, $\partial S(z,v)/\partial z < 0$ and $\partial^2 S(z,v)/\partial z^2 > 0$. Additionally, $\forall v, \lim_{z\to\infty} S(z,v) = 0$.

The condition A3 is asserting that with increasing investments it is possible to decrease the vulnerability level, but at a decreasing rate. Nevertheless, investing larger and larger amounts it is possible to make the attack probability arbitrarily small.

In their paper, Gordon and Loeb give two examples of function families that satisfy the conditions A1–A3,[1] namely:

$$S^{\mathrm{I}} = \frac{v}{(\alpha z+1)^{\gamma}},\ (\alpha > 0, \gamma \in \mathbb{R}) \quad \text{and} \quad S^{\mathrm{II}} = v^{\alpha z+1},\ (\alpha > 0). \tag{11.3}$$

There are several characteristics in Gordon–Loeb (2002). Applying the first order condition (11.2) we can find the optimal level of investment $z^*(v)$. It is a natural idea to compare the optimal investment level to the total expected loss vL. Although it is proved that $z^*(v) < vL$ for all functions $S(z,v)$ satisfying the conditions A1–A3 and even more that $z^*(v) < (1/e)vL$, where $(1/e)$ is a constant, security investment z may be or may not be greater than loss λ in Gordon–Loeb (2002).

It is another characteristic of Gordon–Loeb (2002) that the vulnerability v, the remaining vulnerability $S(z,v)$ and the loss λ are independent of the value of the information system defended against attack.

11.3 Optimal Timing of Information Security Investment

11.3.1 Dynamic Considerations

The analysis by Gordon–Loeb (2002) is often referenced, very important and very fundamental. However in their framework the effect of investment does not affect future security of the information system although they mention it as "investment." With the model we could not analyze the timing of the investment. After all we could understand they are not dealing with "investment." Optimal starting time problem which is one facet of investment, is therefore explored in the following, using real options theory in order to know dynamic aspect of information security investment.

In reality, information security management often has considerable flexibility on when to enter or exit an investment project, and on the scale of the initial and subsequent commitments to make to the project. The firm's option to abandon a project

[1] Willemson (2006) postulated A3 slight differently and obtained other functional form, which state that $S(z,v) = 0$ if z is greater than a certain amount.

during its life amounts to a put option on the remaining cash flows associated with the project. Ignoring the value of these options as done in standard discounted cash flow techniques can lead to incorrect investment evaluation decisions. The value of this flexibility is best captured by real options analysis, the application of option pricing techniques to capital budgeting. It has been established more than a decade ago (Pindyck, 1991; Dixit–Pindyck, 1994; Trigeorgis, 1996 and also Copeland–Antikarnov, 2001, for example) that real options are crucially important in project evaluation.

The value of the option to postpone or delay the new activity (or discontinuing the old activity) becomes a value added which the decision could create, although it must bear cost. The value added could be calculated as a function of the value of the information system.

Traditional cost minimization technique such as Gordon–Loeb (2002) systematically undervalues most investments. Real options analysis allows us to arrive at more accurate valuations of managerial flexibility or strategic value for information security that facilitate better investment decisions than we would arrive at using standard cost minimization analysis.

11.3.2 Literature Review

Using a real options model, this chapter addresses two fundamental questions in the economics of information security area: (i) "How much to invest in information security" and (ii) "When to invest." Although several articles which deal with real options and information security address the issue: for example, Gordon–Loeb–Lucyshyn (2003), roundtable discussion on options and security presented at the second WEIS 2003 and Herath–Herath (2009), this chapter represents one of the first attempts at analytically modeling continuous real options applied to information security.

As a capital budgeting technique for evaluating any projects, a real options approach is known to be promising. It is thus very natural to apply it to information security projects. Gordon–Loeb–Lucyshyn (2003) introduces a discrete tree model of real options into the security literature for manager/practitioner focus. However a formal model is not developed and neither the optimal solution is considered. Herath-Harath (2009) introduces also a discrete tree model of real options with an additional feature of Bayesian postaudit from the management point of view.

Continuous real options model is different from either financial options or discrete tree model of real options. Real options have more flexible feature than financial options as Trigeorgis (1996) and others emphasize. The discrete tree model of real options has such a definite advantage as visibly showing the underlying mechanism. It is also very easy to understand and calculate solutions in a simple example. Although it provides good exhibition or classroom materials, its complexity explodes and it becomes very hard to derive the solution once applied to the complicated real world.

Discrete tree model could not deal with infinite horizon optimization problem. Since firm is a typical infinite entity as a going concern (at least intends to be so), this defect is crucial when we are treating optimization problems which firm faces. Discrete tree model could not be solved analytically, only be solved by numerical analysis. Error by its approximation enters inevitably in numerical analysis and accumulates when we calculate solutions of long distant future. This causes troubles in risk management.

It is true that both continuous model and discrete tree model are needed, but there are actually no works on building continuous real options model applied to information security investment. These considerations make clear the contribution of this chapter beyond the preceding literatures.

It would be always very nice to see how well modeling fits real data. Our concern is not only purely theoretical that how it is formulated theoretically, but also to see how well this fits real data. We set realistic and plausible parameter values to see how the model works in the real world.

11.3.3 Formulation and Solution

In order to give an example of suitable dynamic decision by a firm with optimal starting time for information security investment, we extend the model of Gordon–Loeb (2002).

First of all we let the threat of attempted breach T_t follows geometric Brownian motion with drift:

$$\mathrm{d}T_t = \mu T_t \mathrm{d}t + \sigma T_t \mathrm{d}w, \tag{11.4}$$

where the subscript t is the time of calculation, $\mathrm{d}w$ is the increment of the Weiner process, μ is a drift parameter and σ is the volatility of the process. We denote the initial value of the threat $T_0 = T$ (unsubscripted capital letter).

The drift parameter μ could be negative although the volatility σ of the process has to be positive. Gordon–Loeb (2002) considers T_t as the probability rather than a random variate and confined it to $[0,1]$. We do not need to stick to this assumption. We assume further, letting the risk free interest rate r that

$$r - \mu > 0, \tag{11.5}$$

for the existence of the maximization, avoiding the explosion of the maximand.

The present value of the expected benefit from the investment for the life after at the time of τ security action will be taken is:

$$\int_{\tau}^{\infty} \mathrm{e}^{-rt} \left\{ (v - S(z,v)) \lambda T_t - z \right\} \mathrm{d}t. \tag{11.6}$$

The present value discounted at the risk free interest rate r for the whole life is just the value of the system. Since z is zero and $S(0,v)$ is v until the time of τ because of A2, the maximand until the time of τ is therefore zero. Thus the general formula

for the total expected benefit value of the system is given by Eq. (11.6), which firms try to maximize.[2]

We assume that v and λ are independent of time and security investment decision is made only once at the time of τ. $S(z,v)$ is therefore independent of time. The maximized value $V(T)$ then becomes:

$$\begin{aligned} V(T) &= \sup_{\tau\in\mathscr{T}} \mathbb{E}\left[\int_\tau^\infty \mathrm{e}^{-rt}\big\{(v-S(z,v))\lambda T_t - z\big\}\mathrm{d}t\right], \\ &= \sup_{\tau\in\mathscr{T}} \mathbb{E}\left[\mathrm{e}^{-r\tau}\int_\tau^\infty \mathrm{e}^{-r(t-\tau)}\big\{(v-S(z,v))\lambda T_t - z\big\}\mathrm{d}t\right], \\ &= \sup_{\tau\in\mathscr{T}} \mathbb{E}\left[\mathrm{e}^{-r\tau}\left(\frac{(v-S(z,v))\lambda T_\tau}{r-\mu} - \frac{z}{r}\right)\right]. \end{aligned} \tag{11.7}$$

The derivation of last equation in (11.7) can be done similarly to that in Pindyck (1991). Thus we obtain the following solution. The value of an infinite option must satisfy an ordinary differential equation (ODE)

$$\frac{1}{2}\sigma^2 T^2 V''(T) + \mu T V'(T) - rV(T) + (v-S(z,v))\lambda T - z = 0, \tag{11.8}$$

which can be solved analytically, and a solution to the second order ordinary differential equation can be found by testing a power solution of the form:

$$V(T) = \begin{cases} A_1 T^{\beta_1} + A_2 T^{\beta_2}, & \text{for } T < T^*, \\ \dfrac{(v-S(z,v))\lambda T}{r-\mu} - \dfrac{z}{r}, & \text{for } T \geq T^*, \end{cases} \tag{11.9}$$

where $\beta_1 > 1$ and $\beta_2 < 0$ are the roots of the characteristic equation:

$$\frac{1}{2}\sigma^2\beta^2 + \left(\mu - \frac{1}{2}\sigma^2\right)\beta - r = 0. \tag{11.10}$$

The following boundary conditions have to be satisfied at the optimal time of making the decision:

[2] A typical real options problem is the model where the value of a firm once the capital stock K is determined is just the present value added of S (like M&A synergy) to K in the sacrifice of paying a cost f at the time of τ, discounted at the risk free interest rate. The general formula for the optimal timing problem with the value-added S obtained in the sacrifice of paying cost of f is given by:

$$\int_0^\tau \mathrm{e}^{-rt} K x_t \mathrm{d}t + \int_\tau^\infty \mathrm{e}^{-rt}\big\{(K+S)x_t - f\big\}\mathrm{d}t,$$

where x_t is the return on capital. If we define $S(z,v)$ and v differently, the model in the text has very similar solutions to this problem.

$$\lim_{T\to 0} V(T) = 0, \tag{11.11}$$

$$A_1(T^*)^{\beta_1} = \frac{(v - S(z,v))\lambda T^*}{r-\mu} - \frac{z}{r}, \tag{11.12}$$

$$\beta_1 A_1 (T^*)^{\beta_1 - 1} = \frac{(v - S(z,v))\lambda}{r-\mu}. \tag{11.13}$$

Eq. (11.11) is called as "no-bubble condition" which prevents the divergence of the value function when $T = 0$, that is, there are no value without potential threats. Eq. (11.12) is the "value-matching condition" which states that two equations in (11.9) become equal at T^*. Eq. (11.13) is the "smooth-pasting condition" that states tangencies of both equations are equal. The above three conditions define the parameter A_1, A_2 and T^*:

$$A_1 = \left(\frac{(v - S(z,v))\lambda T^*}{r-\mu} - \frac{z}{r}\right)\left(\frac{1}{T^*}\right)^{\beta_1}, \tag{11.14}$$

$$A_2 = 0, \tag{11.15}$$

$$T^* = \frac{\beta_1}{\beta_1 - 1}\frac{r-\mu}{(v - S(z,v))\lambda}\frac{z}{r}. \tag{11.16}$$

Eq. (11.14) follows from Eq. (11.12) directly, that is, by solving Eq. (11.12) for A_1. Eq. (11.15) is immediately derived from Eq. (11.11). Eq. (11.13) together with Eq. (11.14) yields Eq. (11.16). Then the value function becomes:

$$V(T) = \begin{cases} \left(\dfrac{(v - S(z,v))\lambda T^*}{r-\mu} - \dfrac{z}{r}\right)\left(\dfrac{T}{T^*}\right)^{\beta_1}, & \text{for } T < T^*, \\ \dfrac{(v - S(z,v))\lambda T}{r-\mu} - \dfrac{z}{r}, & \text{for } T \geq T^*, \end{cases} \tag{11.17}$$

which is dependent on the initial value of potential threat T.

This model is based on the real options theory (Pindyck, 1991; Dixit–Pindyck, 1994; Trigeorgis, 1996). Formally speaking it is rather orthodox. The increment $\mathrm{d}V(T)$ increases as T increases where T is smaller than T^* as seen from Eq. (11.17). Then it stays constant as T becomes larger than T^* because $\partial V(T)/\partial T = (v - S(z,v))\lambda/(r-\mu)$. The maximization of $V(T)$ is therefore attained at T^*. In order to further detect the behavior of the value function $V(T)$, we define NPV (net present value) as the present value of the expected benefit from immediate investment, which is given by substituting $\tau = 0$ to Eq. (11.6). Consequently, the formula of NPV is given by the second equation in (11.9) or (11.17). The difference of $(V(T) - \mathrm{NPV})$ is the value of waiting to invest.

Next, we find the optimal level of investment z^*. It is attained by maximizing the expected benefit from the investment at T^*:

$$z(T^*) = \arg\max_{z\in\mathbb{R}} V(T^*; z). \tag{11.18}$$

Note that the optimal level of investment depends on T^*. Because T^* also depends on z, the realized optimal level of investment must satisfy

$$z^* = z\left(\frac{\beta_1}{\beta_1 - 1}\frac{r-\mu}{(v - S(z^*, v))\lambda}\frac{z^*}{r}\right). \tag{11.19}$$

This expression, from Eq. (11.16), might not cause confusion. Finally, Eq. (11.18) means maximization of the second equation in (11.17), so we have the first order condition for z^*:

$$-\frac{\partial S(z^*, v)\lambda T^*}{\partial z} = \frac{r-\mu}{r}, \tag{11.20}$$

which is the same as Gordon–Loeb's deterministic case if $\mu = 0$.

11.3.4 Interpretation

It is an economic problem whether firm should start information security investment today or later. The decision depends on the functional form of the *remaining vulnerability S* and also the properties of Brownian motion of the treat T_t. For example, facing negative trend of the threat (negative drift parameter μ) the firm may have inclined to postpone the investment. The value of the firm is furthermore considered dependent on the volatility of the process σ.

It is a natural interpretation in real options literature that if T_t becomes greater than T^* while watching the process of T_t, firm ought to invest in information security technology. For larger T^*, therefore, the investment timing becomes later because the firm must wait to investment until T_t reaches the larger value of T^*. On the other hand, the timing is sooner for smaller T^*. This T^* is called as optimal investment threshold.

The $V(T)$ function is nonlinear in that it has a kink at T^*. The shape is dependent on r, μ, σ, λ, and v. Then we have to economically interpret the dependency, which will be done in the next section.

11.4 The Optimal Solution: Numerical Illustrations

In this section we numerically calculate the optimal investment threshold T^* and the optimal level of investment z^*. To perform the calculation, we use $S^{\mathrm{I}} = v/(\alpha z + 1)^{\gamma}$, $(\alpha > 0,\ \gamma \in \mathbb{R})$ and $S^{\mathrm{II}} = v^{\alpha z + 1}$, $(\alpha > 0)$ for the remaining vulnerability function case I and II. Furthermore, we present a comparative statics analysis of the threshold and level of investment by changing parameters: volatility σ, drift μ, vulnerability v and the parameter of remaining vulnerability function α. Since the volatility σ represents the degree of uncertainty, among these it is the most important parameter in a real options model.

The drift μ represents the expected growth rate of the potential loss. The vulnerability v is interpreted to represent the urgency of information security investment. The parameter α is interpreted to represent the *efficiency* of the investment. Since these parameters are important, we focus these parameters in this section. We assume that the hypothetical base values of the parameters are as follows: $\sigma = 0.2$, $\mu = 0.02$, $r = 0.05$, $v = 0.5$, $\lambda = 1$, $\alpha = 1$ and $\gamma = 1$. Fig. 11.1 shows the difference of the efficiency of vulnerability reduction between case I and II.

11.4.1 Remaining Vulnerability Case I

In this case, we use $S^{\mathrm{I}} = v/(\alpha z + 1)^{\gamma}$, $(\alpha > 0, \gamma \in \mathbb{R})$. By solving the first order condition

$$-\frac{\partial}{\partial z}\frac{v\lambda T^*}{(\alpha z+1)^{\gamma}} = \frac{r-\mu}{r}, \tag{11.21}$$

after insertion of the function into Eq. (11.20), we have

$$z^* = \frac{(\frac{r}{r-\mu} v\gamma\alpha\lambda T^*)^{1/(\gamma+1)} - 1}{\alpha}, \tag{11.22}$$

which is the same as Gordon–Loeb's deterministic case if $\mu = 0$. Then, we have $T^* = 8.89$ and $z^* = 1.72$, under the hypothetical base values of the parameters. That is, suppose the potential loss reach \$8.89 (million) at the time of τ, the firm should start information security investment \$1.72 (million). After the investment, the remaining vulnerability will be reduced to 0.184 from the hypothetical vulnerability value 0.5.

Fig. 11.2 displays the value functions and the net present value (NPV). The value function $V(T)$ is a convex function and tangent to the NPV at $T^* = 8.89$. This shape resembles the payoff of an American call option before the maturity. For $T < T^*$, the firm wait to investment because the value of waiting ($V(T) - \mathrm{NPV}$) is positive. For $T^* \leq T$, the value of waiting is 0, so that $V(T)$ coincides with the NPV. It shows an orthodox shape in a real options model.

Figs. 11.3–11.6 display the comparative statics of the optimal investment threshold T^* and the optimal level of investment z^* with respect to σ, μ, v and α respectively. In Fig. 11.3, T^* and z^* are displayed with respect to σ. The relationship between T^* and σ is the same as that often observed in a real options model, which is high uncertainty leads to a high threshold, i.e., delay in investment. This is because the value of delaying investment increases in order to wait for new information under high uncertainty. On the other hand, we could see for z^* that high uncertainty σ requires larger amount of the investment expenditure.

In Fig. 11.4, we must distinguish the range of $\mu < 0$ from $\mu > 0$. For $\mu > 0$, high drift causes larger amount of the investment expenditure z^*, and hence forces the firm later investment. On the other hand, for $\mu < 0$, T^* is slightly decreasing with μ. There is a possibility that the expected potential loss will decrease in the future. This

implies that high negative drift makes the necessity of information security investment low. Hence, high negative drift causes later investment and lower investment expenditure. The consideration shows that our dynamic modeling of information security investment is properly formulated and yields reasonable conclusion.

In Fig. 11.5, we find an unique property that the vulnerability has no impact on the level of investment z^* but the investment threshold T^*. Because of the emergency, high vulnerability requires immediate investment. However, the required expenditure is not a variant, independently of the vulnerability. Important thing in this situation is timing, not amount.

Fig. 11.6, where T^* and z^* are displayed with respect to α, shows interestingly enough that high efficiency of vulnerability reduction α encourages the firm to invest earlier and induces cost reduction.

11.4.2 Remaining Vulnerability Case II

In this case, we use $S^{\mathrm{II}} = v^{\alpha z+1}$, $(\alpha > 0)$. By solving the first order condition again after insertion of the function into Eq. (11.20), we have

$$z^* = \frac{\ln \frac{r-\mu}{r} - \ln(-\alpha v \lambda T^* \ln v)}{\alpha \ln v}, \tag{11.23}$$

which is also the same as Gordon–Loeb's deterministic case if $\mu = 0$. Then, we have $T^* = 9.99$, $z^* = 2.53$ and $S(z*, v) = 0.087$. Comparing with case I, the firm needs more expenditure and later investment due to more efficient reduction of vulnerability as shown above in Fig. 11.1.

Figs. 11.7–11.10 display the comparative statics of the optimal investment threshold T^* and the optimal level of investment z^* with respect to σ, μ, v and α, respectively. While Figs. 11.7, 11.8 and 11.10 show the same property as in case I (in Fig. 11.8, the characteristic is more clearly), we can find a following interesting property in Fig. 11.9, from which we could say that case II has more natural result than case I.

Unlike in case I shown in Fig. 11.5, high vulnerability requires high investment expenditure in Fig. 11.9. This is due to the difference of the remaining vulnerability function in both cases. Furthermore, high vulnerability should require later investment, since T^* in Fig. 11.9 is U-shaped with respect to v. High vulnerability requires high expenditure and later investment.

11.5 Concluding Remarks

11.5.1 Summary

It is intuitively clear that firms have to respond quickly (very slowly) to immediate (remote) threat. We do not know how firms respond in the intermediate case. The current rigorous research provides the solution.

Positive drift of threat causes both larger amount of investment expenditure z^* and later investment, while high negative drift causes immediate investment in spite of smaller amount of investment expenditure (Figs. 11.4 and also 11.8). The efficiency of vulnerability reduction technology encourages firm to invest earlier and induces cost reduction (Figs. 11.6 and also 11.10). High vulnerability requires either immediate investment independently of the amount of investment expenditure (Fig. 11.5) or delayed and larger amount of investment (Fig. 11.9).

It has been seen in the last section that case I and case II yield different results although their functional forms look alike as shown in Fig. 11.1. We do not know which function is valid in the real world. It would be therefore concluded that the estimation of $S(z, v)$ is very important especially in the sight of high vulnerability.

11.5.2 Remaining Problems

Several remaining problems are explored in the following.

11.5.2.1 Dynamics Formulation

The investment expenditure z and therefore the vulnerability v could vary every period of time. Attackers come every moment from all over the world and with newer technologies armed. Defenders need to continuously execute information security investment in order to avoid defeated by the unrelenting attacks. Using dynamic control theory the optimal policy over time could be derived under the circumstance.

Furthermore a difficult problem remains left. Attackers strike suddenly. This might be described by a jump process of the threat. The forecasting of their arrival is not possible and any definite countermeasure could be hardly taken before attacks. The formulation of these phenomena will be our works to do next.

11.5.2.2 Attackers' Behavior Formulation

It is also necessary to take some initial steps toward a better understanding of not only how firms invest in information security technology, but also how firms are

faced with the threat (menace) of breach. To formulate attacker's behavior, we have to know what attackers are maximizing.

Attackers might threaten a firm to make an assault on the information system of the firm unless the firm pays money. Another objective of attackers might get the value of targeted firm. They try to lose the confidence of customers or affect reputation of the targeted firm by causing disorder or being mixed up. They do anything to injure the credit of the targeted firm and to rob of their customers. However there is no guarantee to succeed their attempt. It is also possible that he or she is a criminal who enjoys watching how people/firm react to what he or she has done.

If the attack continues for long time, the defender takes action. Firm would defend themselves against the violent attack. Then the payoff of the attacker will diminish and the attacker will change their strategy. This long run problem is certainly hard to be captured.

Hence it is one of the toughest tasks to formulate attacker's objective function. Once we could formulate the attacker's behavior, the equilibrium becomes the solution to a complex 2-player problem. This would not be zero sum 2 person game, some of which could be easily solved by mathematical programming. If we could solve this problem by mathematical programming, however, it has a practical use and helps firms greatly.

11.5.2.3 Empirical Analysis

We have to carry out empirical analysis by finding data of the threat and also vulnerability, and also estimating the distribution parameters of the probability of the threat and success probability of the attack. It would help understanding the real phenomena and constructing strategies of firms to know the parameter values.

As far as the mean and variability of the probability of the threat, we can rely on tools which have been developed in finance field. Having a strong foothold in finance, we could move the estimation forward. Once we know the character of the distribution, we could go back to the theory again and might be able to build a new theory with observations.

Acknowledgements Advices by anonymous reviewers are gratefully acknowledged. All remaining errors are our own.

References

1. Copeland, T., Antikarnov, V.: Real Options: A Practitioner's guide. Texere (2001)
2. Dixit, A.K., Pindyck, R.S.: Investment Under Uncertainty. Princeton University Press (1994)
3. Gordon L.A., Loeb, M.P.: The economics of information security investment. ACM Transactions on Information and System Security **5**(4), 438–457 (2002)
4. Gordon, L.A., Loeb, M.P., Lucyshyn, W.: Information security expenditures and real options: A wait-and-see approach. Computer Security Journal **19**(2), 1–7 (2003)

5. Gal-Or, E., Ghose, A.: The economic incentives for sharing security information. Information Systems Research **16**(2), 186–208 (2005)
6. Herath, H., Harath, T.: Investments in information security: A real options perspective with bayesian postaudit. Journal of Management Information Systems **25**(3), 337–375 (2009)
7. Pindyck, R. S. (1991). Irreversibility, uncertainty, and investment. Journal of Economic Literature **29**(3), 1110–1148.
8. Roundtable discussion in WEIS 2003
`http://www.cpppe.umd.edu/rhsmith3/agenda.htm`
9. Trigeorgis, L.: Real Options. MIT Press (1996)
10. Willemson, J.: On the Gordon & Loeb model for information security investment. In: Proceedings of the 5th Workshop on the Economics of Information Security (WEIS). Cambridge, UK (2006)

FIGURES

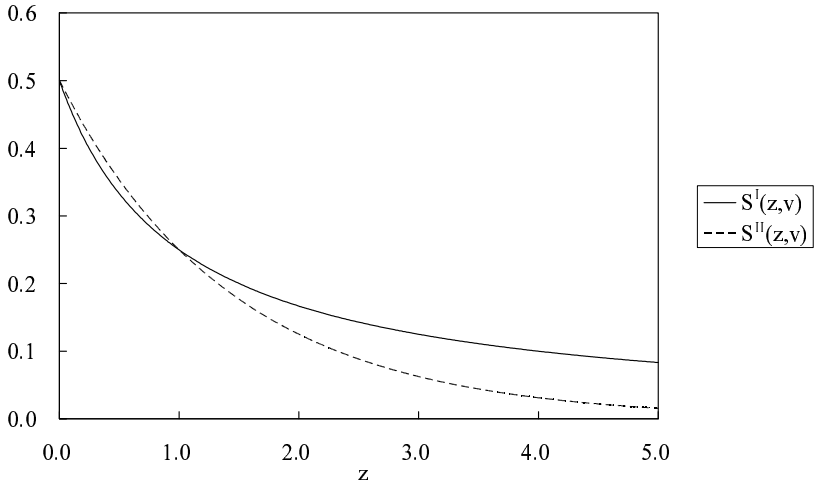


Fig. 11.1 The remaining vulnerability function in case I (solid curve) and II (dashed curve).

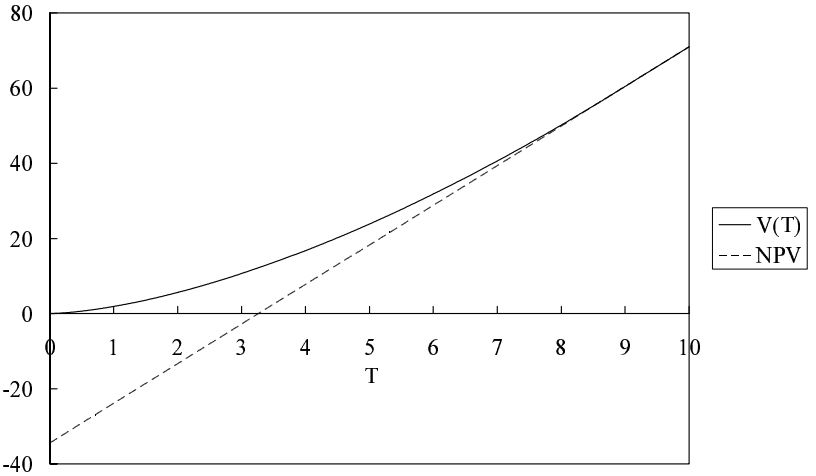


Fig. 11.2 The value function $V(T)$ (solid curve) and NPV (dashed line) in case I.

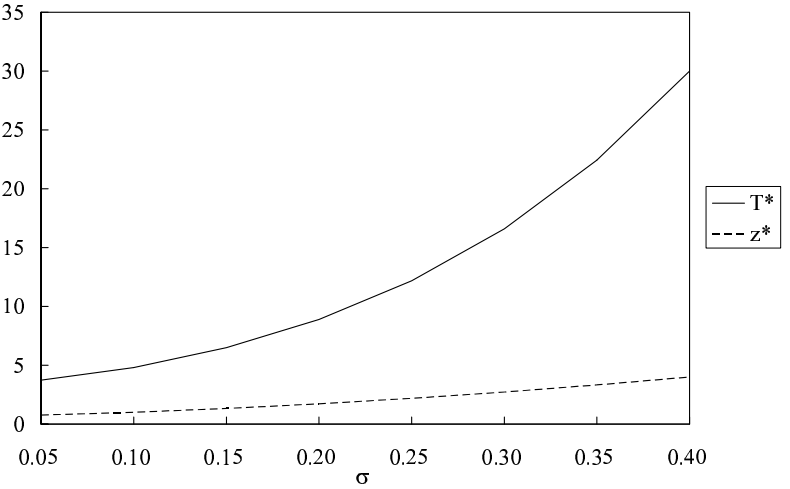


Fig. 11.3 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed line) with respect to σ in case I.

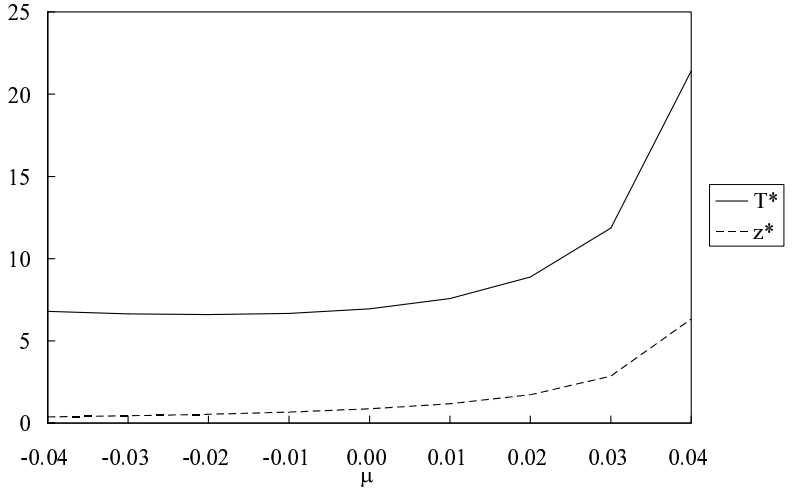


Fig. 11.4 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed curve) with respect to μ in case I.

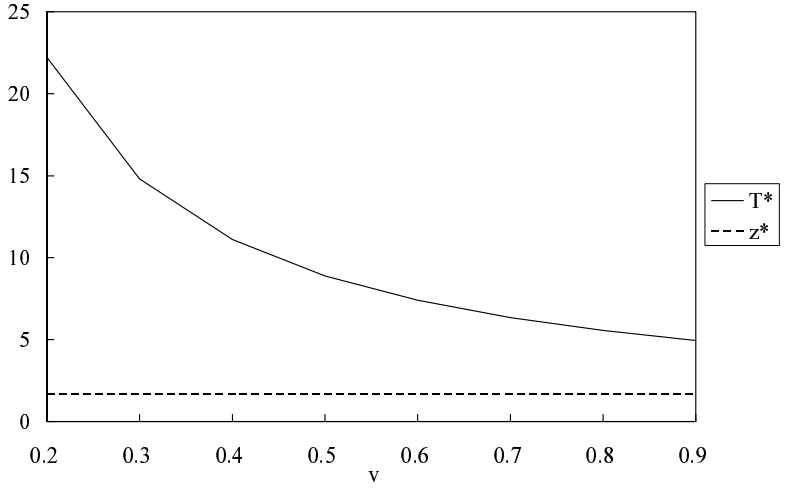


Fig. 11.5 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed line) with respect to v in case I.

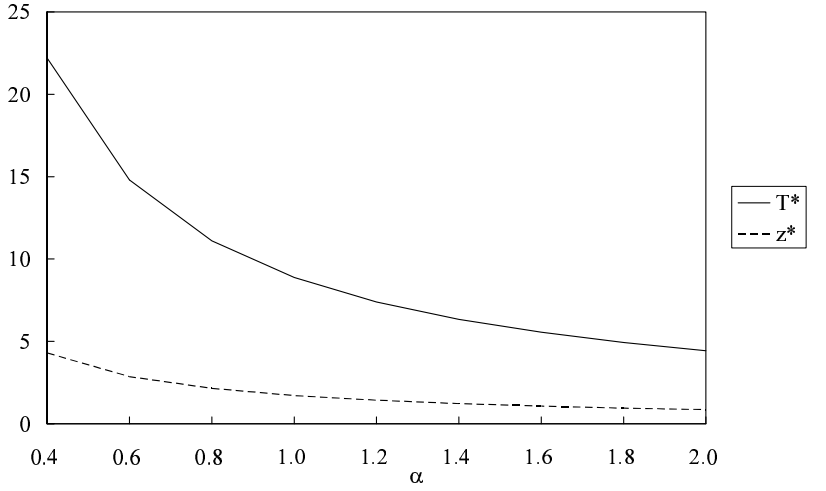


Fig. 11.6 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed curve) with respect to α in case I.

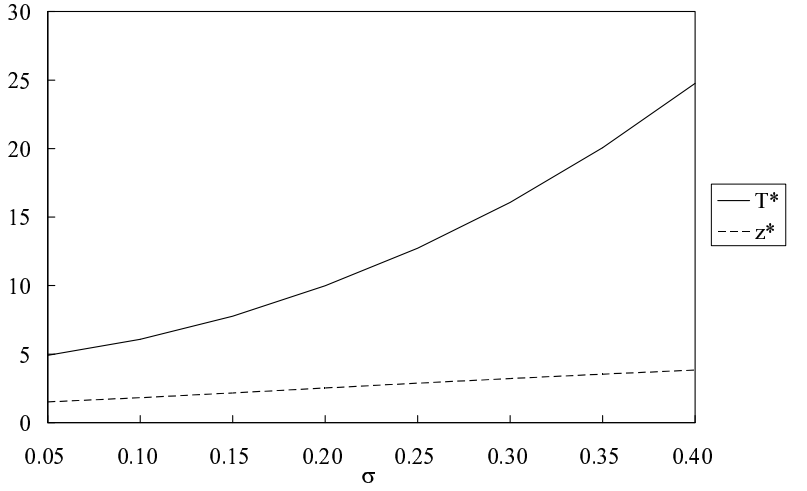


Fig. 11.7 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed line) with respect to σ in case II.

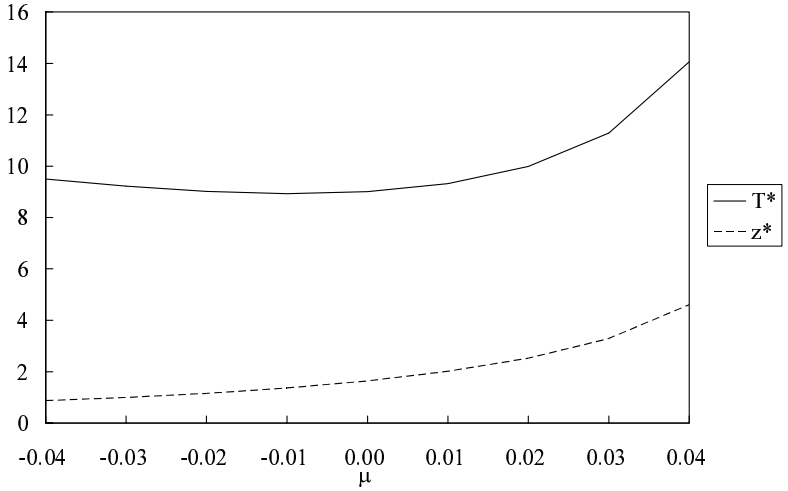


Fig. 11.8 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed curve) with respect to μ in case II.

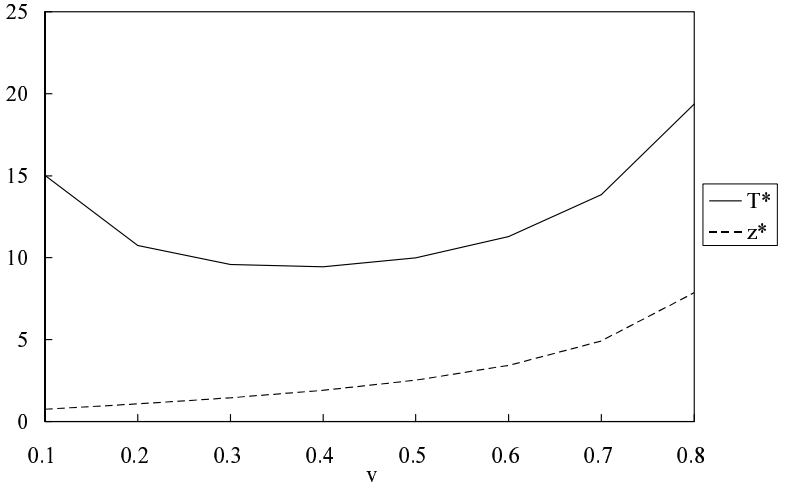


Fig. 11.9 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed curve) with respect to v in case II.

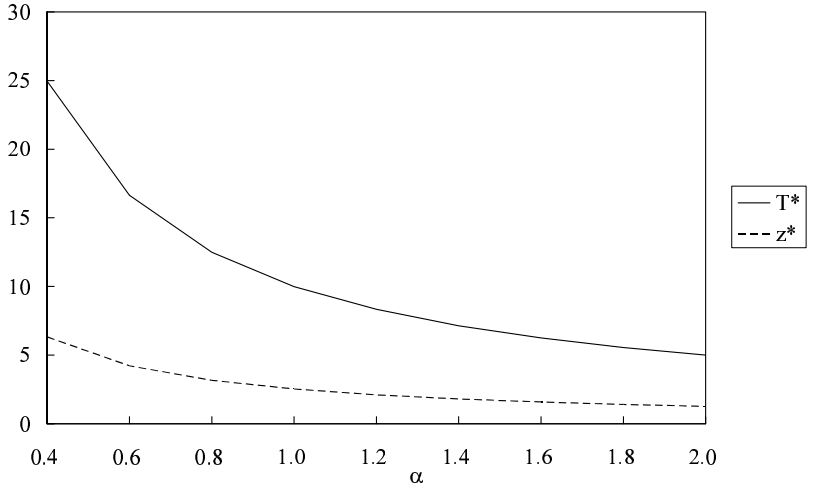


Fig. 11.10 The comparative statics of the optimal investment threshold T^* (solid curve) and the optimal level of investment z^* (dashed curve) with respect to α in case II.

Chapter 12
Competitive Cyber-Insurance and Internet Security

Nikhil Shetty, Galina Schwartz, Mark Felegyhazi, and Jean Walrand

Abstract This paper investigates how competitive cyber-insurers affect network security and welfare of the networked society. In our model, a user's probability to incur damage (from being attacked) depends on both his security and the network security, with the latter taken by individual users as given. First, we consider cyber-insurers who cannot observe (and thus, affect) individual user security. This asymmetric information causes moral hazard. Then, for most parameters, no equilibrium exists: *the insurance market is missing*. Even if an equilibrium exists, the insurance contract covers only a minor fraction of the damage; network security worsens relative to the no-insurance equilibrium. Second, we consider insurers with perfect information about their users' security. Here, user security is perfectly enforceable (zero cost); each insurance contract stipulates the required user security. The unique equilibrium contract covers the entire user damage. Still, for most parameters, network security worsens relative to the no-insurance equilibrium. Although cyber-insurance improves user welfare, in general, competitive cyber-insurers fail to improve network security.

Nikhil Shetty
UC Berkeley, Berkeley-94720, e-mail: nikhils@eecs.berkeley.edu

Galina Schwartz
UC Berkeley, Berkeley-94720, e-mail: schwartz@eecs.berkeley.edu

Mark Felegyhazi
ICSI, Berkeley-94704, e-mail: mark@icsi.berkeley.edu

Jean Walrand
UC Berkeley, Berkeley-94720, e-mail: wlr@eecs.berkeley.edu

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_12,

12.1 Introduction

In this paper,[1] we propose a model to study the effects of cyber insurance on user security and their welfare. Our model highlights how network externalities combined with information asymmetry lead to a *missing market for cyber insurance*.

The Internet serves as a ubiquitous communication platform for both individuals and businesses. Thus, an increasing amount of wealth is accessible online, and cyber-crime is becoming one of the most lucrative criminal activities. Cyber-crime is lucrative because network vulnerabilities are easy to exploit and persecution of cyber-criminals is plagued by enforcement problems. First, and importantly, criminals are relying on the anonymity of the Internet protocols to disguise their traces. Second, global Internet connectivity makes it difficult for law enforcement authorities to identify the origin of the attacks. Exploiting national differences in legal systems, criminals often operate safely from countries with the weakest legislations and enforcement. Third, criminals quickly adapt their attack strategies as new defenses are developed; thus, cyber-crime evolves to minimize the chance of persecution. Altogether, this situation results in formation of highly professional, mafia-style cyber-crime establishments, which are rapidly expanding, see [2].

Technology-based defense and enforcement solutions are available, but there is a consensus among security researchers [2] that the existing security problems cannot be solved by technological means alone. We concur that these security problems primarily result from misaligned incentives of the networked parties with respect to their security. Existing research [4, 7, 16, 18, 19] indicates that *risk management* in general and cyber-insurance in particular are potentially valuable tools for security management. Still, at present, risk management capabilities are virtually nonexistent in the network [2].

We model the effects of informational asymmetries in the presence of network externalities, and study their consequences for network security incentives. We believe that these features of the environment induce socially suboptimal network security, and complicate the management of security risks. We build on the seminal ideas of Akerlof [1], Rothschild and Stiglitz [17] and others,[2] which we combine with the ideas of interdependent security originated by Heal-Kunreuther [14], Gordon-Loeb [8] and Hausken [11].[3]

In our model, all users are identical, meaning that their wealth is identical and they suffer identical damage if successfully attacked. The user's probability of being attacked depends on both the *user security level* and the *network security level*, which individual users take as given. Thus, we have an externality. Indeed, due to this externality, individually optimal user security level is lower than the socially optimal one.

[1] This work was funded in part by the National Science Foundation under grant NSF-0433702. Any opinions, findings, conclusions, and recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding source.

[2] See [20] for the literature review.

[3] See also [3, 5–7, 9, 10, 12, 13, 21]. This list is by no means exhaustive.

Our setting emphasizes that interdependent security is a focal feature, which shapes the incentives for Internet security. Although security interdependence is present in other contexts (such as terrorist attacks [15]), network security is especially prone to these effects because everyone is interlinked.

First, we investigate the effects of information asymmetry in the presence of network effects. Though our model allows to study both moral hazard (when insurers are not aware of user security levels) and adverse selection (when insurers cannot distinguish different user types), in this paper, we address only moral hazard. We demonstrate that for a wide range of parameters, insurance market fails to exist, i.e., we observe a *missing market*.

Next, we assume no information asymmetry between the insurers and the insured (users). We demonstrate that user utility is higher with insurance, but the network security level is not necessarily higher. On reverse, in many cases network security worsens with insurers. Indeed, insurers only *manage* risks, but they do not necessarily reduce them.

Our homogeneity assumption is simplistic, and does not hold in the actual Internet. But, adding user and insurer heterogeneity to our setting only adds more informational asymmetries. Then, the lemon problem becomes likely, which itself could cause missing markets [1]. Thus, with heterogeneity, one expects adverse selection problems, which would also contribute to missing markets.

We make two main contributions to the literature. First, we observe that even with no heterogeneity (of users and insurers), information asymmetries complicate the formation of viable cyber-insurance markets. Second, we demonstrate that even in the absence of informational asymmetries, competitive cyber-insurers fail to improve network security. The significant implication is that in the existing network environment, cyber-insurance markets cannot serve as a catalyst for improvement of network security.

The paper is organized as follows. In Section 12.2, we propose a base model, derive its Nash equilibrium, and compare it with socially optimal allocation. In Section 12.3, we add competitive insurers to our base model, analyze the equilibrium with insurers. We consider two cases: when individual security levels are non-contractible and when insurers include the requirement about individual security level into the contract. In Section 12.4, we summarize our findings and conclude. The technical details are relegated to Appendix.

12.2 Model

In this section, we present our base model, which highlights the interdependence of user and network security. We consider a network populated by identical users. Each user i has two choice variables: the *convenience* level $a_i > 0$ of his network activity, and his *security* level $s_i \in [0,1]$. The *convenience* level a_i can be, for example, characterized by the number of applications utilized by the user, such as emails, Web, IM, P2P, etc. If there are no security problems, the user derives utility from

his wealth and from network usage. We assume that both these components of user utility U_i are additively separable:

$$U_i = K_1 \cdot f(W) + K_2 \cdot g(a_i) - K_3 \cdot a_i,$$

where K_1, K_2 and K_3 are positive constants, and $W > 0$ denotes user's wealth. We assume that the functions f and g are increasing and concave, reflecting that user wealth W and *convenience* level a_i have a positive but decreasing marginal benefit for the user. To increase his *convenience* level, user incurs a linear cost (cost of effort).

In the presence of network attacks, we assume that, if the attack on the user is successful, the user incurs a monetary damage $D \in (0, W)$. Let p_i be the probability that user i suffers such an attack. This probability depends on two factors: the network security level $\bar{s} \in [0,1]$, which determines the probability of a user being attacked, and the user security level s_i, which determines the probability of success of such an attack. This justifies our expression for p_i:

$$p_i = (1 - s_i) \cdot (1 - \bar{s}) = v_i \cdot \bar{v}, \tag{12.1}$$

where for mathematical convenience, we introduce the *user vulnerability* level $v_i = 1 - s_i$ and the *network vulnerability* level $\bar{v} = 1 - \bar{s}$. Further, assume that $\bar{s}$ is equal to the average security levels of its users:

$$\bar{s} = \frac{\sum_{i=1,\ldots N} s_i}{N}, \tag{12.2}$$

and we let the number of users N be large enough so that a single user has a negligible effect on the network security level. Thus, each user takes the network security level as a given parameter.

We assume that user's choice of a higher security requires a higher user cost (in terms of effort), and this cost is proportional to the convenience level. Again, assuming additive separability, we express the expected utility of user i in the presence of network insecurity as:

$$E[U_i] = K_1 \{(1 - p_i) \cdot f(W) + p_i \cdot f(W - D)\} + K_2 \cdot g(a_i) - K_3 \cdot a_i \cdot (h(s_i) + 1), \tag{12.3}$$

where the security cost function, $h(\cdot)$ is increasing and convex ($h', h'' > 0$) with $h(0) = 0$ corresponding to zero security level and $h(1) = \infty$, corresponding to a hypothetical "perfectly secure" system. Thus, it becomes increasingly costly to improve the security level at a higher level of security.

For simplicity, we let $f(x) = g(x) = \sqrt{x}$ and $h(x) = \frac{1}{\sqrt{1-x}} - 1$ and solve the problem for these specific functions. Then, (12.3) becomes:

$$E[U_i] = K_1 \left\{(1 - p_i)\sqrt{W} + p_i\sqrt{W - D}\right\} + K_2\sqrt{a_i} - K_3 a_i \frac{1}{\sqrt{v_i}}). \tag{12.4}$$

Since we assume that the convenience level of user i's network usage a_i is not affected even when this user is attacked, this model may be more suitable for attacks like phishing, eavesdropping, etc. rather than for attacks like denial-of-service.

12.2.1 Analysis

We start by deriving the optimal convenience level a_i^* by taking the partial derivative of (12.4) with respect to a_i:

$$\frac{\partial E[U_i]}{\partial a_i} = K_2 \frac{1}{2} \frac{1}{\sqrt{a_i}} - K_3 \frac{1}{\sqrt{v_i}},$$

from which a_i^* is:

$$a_i^* = \frac{1}{4} \frac{K_2^2}{K_3^2} v_i. \tag{12.5}$$

Thus, the user's a_i^* depends only on her choice of v_i, but not on network vulnerability level $\overline{v}$. Next, we substitute (12.5) in (12.4) to obtain:

$$E[U_i] = \frac{1}{4} \frac{K_2^2}{K_3} [\sqrt{v_i} - v_i \overline{v} K (\sqrt{W} - \sqrt{W - D}) + K\sqrt{W}] \tag{12.6}$$

where $K = \frac{4K_1K_3}{K_2^2}$. To simplify, we let $\frac{1}{4}\frac{K_2^2}{K_3} = 1$, and obtain a normalized utility:

$$E[U_i] = \sqrt{v_i} - v_i \overline{v} K (\sqrt{W} - \sqrt{W - D}) + K\sqrt{W}. \tag{12.7}$$

The constant K characterizes how users value their wealth relative to the utility from the network.

12.2.1.1 Nash Equilibrium

To find the user i's best response $v_i^*(\overline{v})$ to a given network vulnerability $\overline{v}$, we optimize (12.7) with respect to v_i (subject to $v_i \leq 1$) and express $v_i^*(\overline{v})$ as

$$v_i^*(\overline{v}) = \min \left\{ \frac{1}{[2\overline{v}K(\sqrt{W} - \sqrt{W - D})]^2}, 1 \right\}. \tag{12.8}$$

From (12.8), $v_i^*(\overline{v})$ is identical for all users, from which any Nash equilibrium is symmetric, and let $v_i^*(\overline{v}) = v_j^*(\overline{v}) = v^*$ for any users i and j. Then, from (12.2), we have $\overline{v} = v^*$ and hence,

$$v^* = \min \left\{ \frac{1}{[2v^*K(\sqrt{W} - \sqrt{W - D})]^2}, 1 \right\},$$

from which we obtain Nash equilibrium vulnerability v^*:

$$v^* = 1 - s^* = \min\left\{\frac{1}{[2K(\sqrt{W} - \sqrt{W-D})]^{2/3}}, 1\right\}. \tag{12.9}$$

From (12.9), $v^* < 1$ only if $\sqrt{W} - \sqrt{W-D} > \frac{1}{2K}$ and thus, all else equal, users invest in security only when their damage D or K become sufficiently high, or when user wealth W is low.

12.2.1.2 Social Optimum

We assume that a social planner unilaterally dictates user vulnerability, $v_i = v$, and maximizes cumulative utility of the users. Since users are identical, this maximization is identical to a representative user utility maximization with $\bar{v} = v$. From (12.7), the representative user utility is:

$$E[U] = \sqrt{v} - v^2 K(\sqrt{W} - \sqrt{W-D}) + K\sqrt{W}. \tag{12.10}$$

Maximizing (12.10), subject to $v \leq 1$, we obtain the socially optimal vulnerability v^{soc} as:

$$v^{soc} = 1 - s^{soc} = \min\left\{\frac{1}{[4K(\sqrt{W} - \sqrt{W-D})]^{2/3}}, 1\right\}. \tag{12.11}$$

Thus, $v^{soc} < 1$ only if $(\sqrt{W} - \sqrt{W-D}) > \frac{1}{4K}$. As expected, $v^{soc} \leq v^*$, which allows us to formulate the following proposition:

Proposition 12.1. *When the socially optimal security level is strictly positive, it is strictly higher than the individually optimal one: $s^{soc} > s^*$. Users are strictly better off in the social optimum than in the Nash equilibrium.*

In the next section, we extend this model to the presence of competitive insurers. We will investigate how insurer information about user security level (or lack of such information) impacts network security.

12.3 Insurance Model

We define market equilibrium similar to the model of Rothschild and Stiglitz [17], who pioneered the examination of equilibria in insurance markets with information asymmetries. We assume that each insurer offers a single insurance contract in a *class of admissible contracts*, or does nothing. A Nash equilibrium is defined as a set of admissible contracts such that: i) all contracts result in a non-negative utility for the insurers, ii) taking as given the contracts offered by incumbent insurers (those offering contracts), there is no additional contract which an entrant-insurer (one not offering a contract) can offer and make a strictly positive profit and iii) taking as

given the set of contracts offered by other incumbent insurers, no incumbent can increase its profits by altering his offered contract. The literature referred to such contracts as "competitive", because entry and exit are free, and because no barrier to entry or scale economies are present.

We consider risk neutral insurers who compete with each other. Let ρ be the premium charged to a user and $L > 0$ be his loss covered by the insurer. We do not consider $L < 0$ because it is unrealistic to expect a fine when a user suffers a damage. Let v and $\bar{v}$ be the user and network vulnerability. Then, we denote the respective user utility by $U(v,\bar{v},\rho,L)$, and from (12.7) and (12.1), we have:

$$U(v,\bar{v},\rho,L) = \sqrt{v} + v\bar{v}K\sqrt{W-D+L-\rho} + (1-v\bar{v})K\sqrt{W-\rho}. \tag{12.12}$$

If v,ρ,L are identical for all users, then $v = \bar{v}$, and we obtain

$$U(v,v,\rho,L) = \sqrt{v} + v^2K\sqrt{W-D+L-\rho} + (1-v^2)K\sqrt{W-\rho}. \tag{12.13}$$

Additionally, we will assume that insurers take network security $\bar{v}$ as given. This assumption reflects that individual insurers cannot affect $\bar{v}$ on their own.

12.3.1 Insurance with Non-Contractible Security

In this section, we assume that it is impossible (or too costly) for the insurers to monitor the users' security level. Indeed, even if v is included in the contract and user compliance is observable by the insurer, but unverifiable in court (due to the prohibitively high costs), the insurer would effectively operate as if no requirement on v is imposed. Thus, we consider the contracts of the form (ρ,L) only. In addition, we will assume that contracts stipulate that purchase of extra coverage from outside parties is prohibited. Further, since the users are homogeneous, we will restrict our attention to a symmetric equilibrium, i.e., the equilibria with identical user actions. Henceforth, we will use the superscript ‡ to distinguish the values in such an equilibrium.

Let user i purchase a contract (ρ,L). Then, he will choose his vulnerability v_i to maximize his utility (taking $\bar{v}$ as given):

$$E[U_i] = \sqrt{v_i} - v_i\bar{v}K(\sqrt{W-\rho} - \sqrt{W-D+L-\rho}) + K\sqrt{W-\rho}. \tag{12.14}$$

Any contract which improves user utility $U(v,\bar{v},\rho,L)$ is preferred by users to any other contract. Hence, in equilibrium, there should exist no such deviating contract that makes non-negative profits for an insurer. Further, the equilibrium contract is constrained by user participation - a user must prefer to buy insurance, assuming that others already did so, to staying without insurance. In Appendix, we show that this participation constraint never binds, and, in equilibrium, due to competition, insurers' profits are zero: $\rho^{\ddagger} = \left(v^{\ddagger}\right)^2 L^{\ddagger}$. Further, we demonstrate that, in any equilibrium:

$$L^{\ddagger} < D,$$

and from user optimization, we have:

$$v^{\ddagger} = \frac{1}{\left[2K(\sqrt{W-\rho^{\ddagger}} - \sqrt{W-D+(L^{\ddagger}-\rho^{\ddagger})}\right]^{2/3}}. \tag{12.15}$$

Comparing (12.15) with (12.9), we infer that in any equilibrium:

$$v^{\ddagger} > v^{*}. \tag{12.16}$$

Although the availability of insurance may allow users to reach a higher utility, the network security is strictly lower with insurance. In Appendix, we prove the following proposition:

Proposition 12.2. *If $D < \frac{8}{9}W$, any insurance contract with security levels unobservable by the insurers strictly decreases the utility of the users. Hence, no insurance is offered and no insurance market exists. If $D > \frac{8}{9}W$, there could exist an equilibrium in which all users purchase insurance contract $(\rho^{\ddagger}, L^{\ddagger})$. This insurance improves users' utility relative to the no insurance case, but decreases their security (i.e., $v^{\ddagger} > v^{*}$ is always true).*

From Proposition 12.2, the presence of insurers negatively affects network security. Indeed, here, security is chosen by the users, and insured users have meager incentives to secure themselves. This is a typical manifestation of a moral hazard. In this case, the expected per user loss due to network insecurity increases by:

$$\Delta^{\ddagger} = \left[(v^{\ddagger})^2 - (v^{*})^2\right] D.$$

12.3.2 Insurance with Contractible Security

In this section, we assume that insurers can enforce a desired security level for the insured users at zero cost. Thus, we permit contracts (v, ρ, L) to specify a user's required vulnerability v. In reality, this may be achieved, for example, by deploying tamper-proof security software that monitors and enforces user security.

12.3.2.1 Social Planner

Next, we derive the social planner choice of contract when security is contractible. Let $(v^{\dagger}, \rho^{\dagger}, L^{\dagger})^{soc}$ be the contract chosen by a social planner. The social planner objective is to maximize the user utility, subject to the constraint of non-negative profits:

$$\max_{\rho,v,L} U(v,v,\rho,L)$$
$$s.t. \quad v^2 L \leq \rho \text{ and } v \leq 1.$$

In Appendix, we solve this optimization problem, and derive the following social planner's equilibrium:

$$\rho^{\dagger soc} = (v^{\dagger soc})^2 L^{\dagger soc},$$

and full coverage will be offered since users prefer it:

$$L^{\dagger soc} = D.$$

If the equilibrium vulnerability $v^{\dagger soc} < 1$, then it must be a solution of:

$$\frac{v^3}{W - v^2 D} = \frac{1}{(2KD)^2}, \tag{12.17}$$

which we have proven to be unique.

12.3.2.2 Competitive Insurers

Any insurance contract (v,ρ,L) that achieves a higher user utility $U(v,\bar{v},\rho,L)$ would be preferred to other contracts. In equilibrium, there should exist no contract that permits non-negative insurer profits and yields a higher user utility than the equilibrium contract does. In addition, we modify the definition of insurance market equilibrium in Section 12.3 and assume that no single insurer affects the network vulnerability. This assumption is realistic since competitive insurers lack market power. The participation constraint must hold in equilibrium, i.e., insured users must obtain at least the same utility with insurance than by staying uninsured. In Appendix, we show that only a unique contract can exist in equilibrium. Let this equilibrium contract be denoted by $(v^\dagger, \rho^\dagger, L^\dagger)$.

In Appendix, we demonstrate that, in equilibrium, insurers make zero profits and offer full coverage since users prefer it.

$$\rho^\dagger = (v^\dagger)^2 L^\dagger, \text{ and } L^\dagger = D.$$

If the equilibrium vulnerability $v^\dagger < 1$, then it must be a solution of:

$$\frac{v^3}{W - v^2 D} = \frac{1}{(KD)^2}, \tag{12.18}$$

which we have proven to be unique. From (12.17) and (12.18), we conclude that the vulnerability in the competitive insurer equilibrium is higher than that in the social optimum: $v^\dagger > v^{\dagger soc}$. In Appendix, we also derive the condition for $v^\dagger < v^*$. We find that equilibrium vulnerability only improves (relative to the Nash equilibrium without insurance) when $\frac{D}{W}$ is lower than some critical value. This critical value is achieved only when v^* is close to 1, i.e., when user security is close to zero in the

no-insurance Nash equilibrium. Thus, for a large range of parameters, $v^\dagger > v^*$, i.e., the presence of insurance leads to a higher vulnerability.

This permits us to formulate the following proposition:

Proposition 12.3. *With insurers present, and security levels contractible, in any equilibrium, full coverage $L^\dagger = D$ is offered. For most parameters, equilibrium network security is lower than in the no-insurance equilibrium. Only when user security is low in the no-insurance Nash equilibrium (i.e., v^* close to* 1*), the presence of insurers improves network security.*

From Proposition 12.3, with security levels observable by the insurers, the insurers' presence allows to improve user welfare, but hardly improves network security. When $v^\dagger < v^*$, the insurers' presence reduces the per user expected loss from network insecurity by $\Delta^\dagger$, where:

$$\Delta^\dagger = \left[(v^*)^2 - (v^\dagger)^2\right] D.$$

Else, the per user expected loss increases by

$$\Delta^\dagger = \left[(v^\dagger)^2 - (v^*)^2\right] D.$$

Figure 12.1(a) depicts the equilibrium security level of users (and hence the network security level) as a function of the damage D while Figure 12.1(b) depicts the equilibrium utility of users as a function of D. The parameter values used are $K = 1$ and $W = 100$.

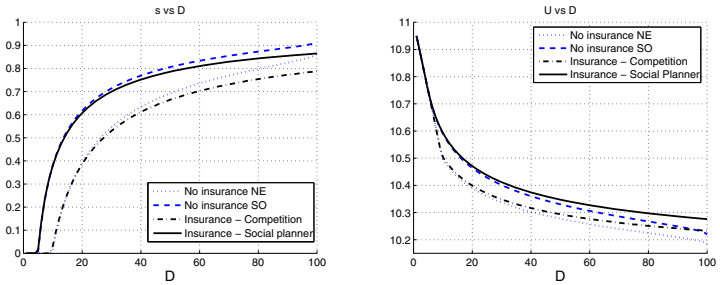


Fig. 12.1 (a) Security level and (b) utility of homogeneous users in equilibrium as a function of the damage $0 < D < W$. Here $W = 1000$ and $K = 1$.

12.4 Conclusion

In this paper, we investigate the effects of competitive cyber-insurers on network security and welfare. We highlight the impact of asymmetric information in the pres-

ence of network externalities and address the effects of interdependent security on the market for cyber-risks. The existing literature attributes cyber-insurance a significant role in cyber-risk management; it especially emphasizes positive effects of cyber-insurance market on security incentives. We find that, on reverse, the presence of competitive cyber-insurers weakens user incentives to improve security.

First, we consider insurers who cannot observe (and thus, cannot contract) user security; here, insurers observe the network security only. Then, the moral hazard problem is present, i.e., with more insurance coverage, the users' incentives to invest in security become meager. In this case, for most parameters, the insurance market collapses, i.e., no insurance is offered in equilibrium. Even if cyber-insurance exists, it covers a minor fraction of damages only. Our findings are in line with the existing Internet, where cyber-insurance is scantly observed.

Second, we consider insurers who observe (and thus, can contract) user security. Here, insurers' contracts include user security level which insurers enforce at zero cost, and thus, no moral hazard is present. Still, in general, competitive insurers fail to improve upon the security level of the no-insurance equilibrium. Though insurance improves the utility for risk-averse users, it does not serve as an incentive device for improving security practices. Indeed, insurance is a tool for risk management and redistribution, not necessarily a tool for risk reduction.

To sum up, we argue that a combination of network effects and information asymmetries leads to difficulties in formation of viable insurance markets for cyber risks. Thus, our results dash the hopes for both, expectations of development of cyber insurance markets under the current network environment, and for the beliefs that such markets may serve as a catalyst for improvement of network security.

12.5 Appendix

Proof of Proposition 12.2

When the user vulnerability v is non-contractible, the contracts have the form (ρ, L), and v is selfishly chosen by the users. Since our users are homogeneous, we will restrict our attention to a symmetric equilibrium, i.e., user actions in equilibrium are identical. Let $(\rho^\diamond, L^\diamond)$ be such an equilibrium insurance contract and $\overline{v}^\diamond$ be the resulting network vulnerability. First, we show that in any equilibrium $L^\diamond < D$.

Assume the reverse and let $(\rho^\diamond, L^\diamond = D)$ be an equilibrium. In this case, it is optimal for each user to choose $v = 1$. Hence, $\overline{v}^\diamond = 1$ and $\rho^\diamond = D$ for non-negative insurer profits. From (12.13), $U(1,1,D,D) = U(1,1,0,0)$, which implies that the user is indifferent between buying and not buying insurance. If the vulnerability in the no-insurance Nash equilibrium $v^* < 1$, then the user's participation constraint does not hold: $U(v_i,1,0,0) > U(1,1,D,D)$ for some $v_i < 1$. This is a contradiction since user i is better off not purchasing such an insurance, and therefore $L^\diamond < D$.

To determine the vulnerability that the insured user chooses selfishly, we differentiate his utility with respect to v, keeping $\overline{v}$ fixed:

$$\frac{\partial U(v,\overline{v^\diamond},\rho^\diamond,L^\diamond)}{\partial v}=0,$$

and we have

$$v=\frac{1}{(2\overline{v^\diamond}K(\sqrt{W-\rho}-\sqrt{W^D+L-\rho}))^2}. \tag{12.19}$$

where $W^D=W-D$. In equilibrium, $v=\overline{v^\diamond}$ and from (12.19), we obtain (similar to (12.9)):

$$\overline{v^\diamond}=\frac{1}{\left[2K(\sqrt{W-\rho^\diamond}-\sqrt{W^D+(L^\diamond-\rho^\diamond)}\right]^{2/3}}, \tag{12.20}$$

Comparing (12.20) with (12.9), we infer that:

$$v^*<\overline{v^\diamond}, \tag{12.21}$$

because

$$\sqrt{W-\rho^\diamond}<\sqrt{W} \text{ and } \sqrt{W^D+(L^\diamond-\rho^\diamond)}\geq\sqrt{W^D}.$$

Next, let us make sure that no user deviates and stays without insurance, that is the participation constraint holds. For the uninsured user i, utility is maximized at

$$v_i=\frac{1}{(\overline{v^\diamond})^2\left[2K(\sqrt{W}-\sqrt{W^D})\right]^2}. \tag{12.22}$$

Comparing this with (12.9), we have

$$v_i(\overline{v^\diamond})^2=(v^*)^3,$$

and from (12.21),

$$v_i=\left(\frac{v^*}{\overline{v^\diamond}}\right)^2 v^*<v^*,$$

and his maximum attainable utility is

$$\begin{aligned}U_i&=\sqrt{v_i}+v_i\overline{v^\diamond}\left[2K(\sqrt{W}-\sqrt{W^D})\right]+K\sqrt{W}\\&=\left(\frac{v^*}{\overline{v^\diamond}}\right)\sqrt{v^*}+\left(\frac{v^*}{\overline{v^\diamond}}\right)(v^*)^2\left[2K(\sqrt{W}-\sqrt{W^D})\right]+K\sqrt{W}<U^*.\end{aligned} \tag{12.23}$$

Note that $U^*=U(v^*,v^*,0,0)$. Hence, for $(\rho^\diamond,L^\diamond)$ to be an equilibrium contract, $U(v^\diamond,\overline{v^\diamond},\rho^\diamond,L^\diamond)>U(v^*,v^*,0,0)=U^*$. Then, from (12.23),

$$U_i<U^*<U(v^\diamond,\overline{v^\diamond},\rho^\diamond,L^\diamond),$$

and we infer that the participation constraint does not bind.

Next, we show that, if $D<\frac{8}{9}W$, the only equilibrium contract is $(0,0)$. Consider a contract (ρ,L) and let $\tilde{v}$ be the vulnerability obtained from (12.20). Due to insurer

competition, in any equilibrium

$$\rho = \tilde{v}^2 L. \tag{12.24}$$

If not, an entrant insurer could design another contract that yields lower profits, which users prefer since it maximizes their utility. The user utility is obtained by substituting (12.20) in (12.12). Then, we have

$$U = K\sqrt{W-\rho} + \frac{1}{(16K(\sqrt{W-\rho} - \sqrt{W^D + L - \rho}))^{1/3}}. \tag{12.25}$$

Using (12.24), we rewrite (12.25) as $K\sqrt{W-\tilde{v}^2 L} + \frac{1}{(16K(\sqrt{W-\tilde{v}^2 L} - \sqrt{W-D+L-\tilde{v}^2 L}))^{1/3}}$.

Let $\dot{\tilde{v}}$ denote $\frac{\partial \tilde{v}}{\partial L}$, and let $\tilde{W}^D = W^D + (L-\rho)$ and $\tilde{W} = W - \rho$. Next, we demonstrate that $\dot{\tilde{v}} > 0$. From (12.20),

$$\begin{aligned}
\frac{\partial \tilde{v}^3}{\partial L} &= \frac{\partial}{\partial L} \frac{1}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^2} \\
3\tilde{v}^2 \dot{\tilde{v}} &= \frac{-2}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^3} \left(\frac{1}{2\sqrt{\tilde{W}}} \frac{\partial \tilde{W}}{\partial L} - \frac{1}{2\sqrt{\tilde{W}^D}} \frac{\partial \tilde{W}^D}{\partial L} \right) \\
&= \frac{-2}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^3} \left(\frac{(-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{\tilde{W}}} - \frac{(1-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{\tilde{W}^D}} \right) \\
&= \frac{1}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^3} \left(\frac{(\tilde{v}^2 + 2\tilde{v}\dot{\tilde{v}}L)}{\sqrt{\tilde{W}}} + \frac{(1-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{\sqrt{\tilde{W}^D}} \right) \\
\therefore \dot{\tilde{v}} \Bigg(3\tilde{v}^2 &+ \frac{2\tilde{v}L}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^3} \left[\frac{1}{\sqrt{\tilde{W}^D}} - \frac{1}{\sqrt{\tilde{W}}} \right] \Bigg) \\
&= \frac{1}{(2K(\sqrt{\tilde{W}} - \sqrt{\tilde{W}^D}))^3} \left(\frac{\tilde{v}^2}{\sqrt{\tilde{W}}} + \frac{(1-\tilde{v}^2)}{\sqrt{\tilde{W}^D}} \right),
\end{aligned}$$

where the last step is obtained by moving all the terms involving $\dot{\tilde{v}}$ to the LHS. The RHS is obviously positive while the coefficient of $\dot{\tilde{v}}$ on the LHS is also positive (since $\tilde{W} > \tilde{W}^D$) and $\dot{\tilde{v}} > 0$ is proven.

Next, we differentiate the utility w.r.t. L,

$$\begin{aligned}
\frac{\partial U}{\partial L} &= \frac{K}{2\sqrt{W-\tilde{v}^2 L}}(-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L) + \frac{\frac{-1}{3}}{(16K)^{1/3}(\sqrt{W-\tilde{v}^2 L} - \sqrt{W-D+L-\tilde{v}^2 L})^{4/3}} \\
&\quad \ldots \times \left(\frac{(-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{W-\tilde{v}^2 L}} - \frac{((1-\tilde{v}^2) - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{W-D+L-\tilde{v}^2 L}} \right) \\
&= \frac{K(-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{W-\tilde{v}^2 L}} - \frac{K\tilde{v}^2}{3} \left(\frac{(-\tilde{v}^2 - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{W-\tilde{v}^2 L}} - \frac{((1-\tilde{v}^2) - 2\tilde{v}\dot{\tilde{v}}L)}{2\sqrt{W-D+L-\tilde{v}^2 L}} \right)
\end{aligned}$$

Collecting the terms and simplifying we obtain:

$$\begin{aligned}\frac{2}{K}\frac{\partial U}{\partial L} &= \frac{-\tilde{v}^2}{\sqrt{\tilde{W}}} - \frac{2\tilde{v}\dot{\tilde{v}}L}{\sqrt{\tilde{W}}} + \frac{\tilde{v}^2}{3\sqrt{\tilde{W}^D}} - \frac{\tilde{v}^2}{3}\left(\frac{(-\tilde{v}^2-2\tilde{v}\dot{\tilde{v}}L)}{\sqrt{\tilde{W}}} - \frac{(-\tilde{v}^2-2\tilde{v}\dot{\tilde{v}}L)}{\sqrt{\tilde{W}^D}}\right)\\ &= -\tilde{v}^2\left(\frac{1}{\sqrt{\tilde{W}}} - \frac{1}{3\sqrt{\tilde{W}^D}}\right) - \frac{2\tilde{v}\dot{\tilde{v}}L}{\sqrt{\tilde{W}}} + \frac{\tilde{v}^2(\tilde{v}^2+2\tilde{v}\dot{\tilde{v}}L)}{3}\left(\frac{1}{\sqrt{\tilde{W}}} - \frac{1}{\sqrt{\tilde{W}^D}}\right)\\ &= -\tilde{v}^2\left(\frac{3\sqrt{\tilde{W}^D}-\sqrt{\tilde{W}}}{3\sqrt{\tilde{W}}\sqrt{\tilde{W}^D}}\right) - \frac{2\tilde{v}\dot{\tilde{v}}L}{\sqrt{\tilde{W}}} + \frac{\tilde{v}^2(\tilde{v}^2+2\tilde{v}\dot{\tilde{v}}L)}{3}\left(\frac{\sqrt{\tilde{W}^D}-\sqrt{\tilde{W}}}{\sqrt{\tilde{W}}\sqrt{\tilde{W}^D}}\right) \qquad (12.26)\end{aligned}$$

Since $2\tilde{v}\dot{\tilde{v}}L > 0$ and $\sqrt{\tilde{W}^D} < \sqrt{\tilde{W}}$, the last two terms of (12.26) are strictly negative for any $L \geq 0$.

Let $D < \frac{8}{9}W$. Then, $W < 9(W-D)$, and taking the square root we obtain:

$$3\sqrt{W^D} - \sqrt{W} > 0, \qquad (12.27)$$

and since $\tilde{W}^D = W^D + (L^\ddagger - \rho^\ddagger) > W^D$ and $\tilde{W} = W - \rho < W$ from (12.27) we have:

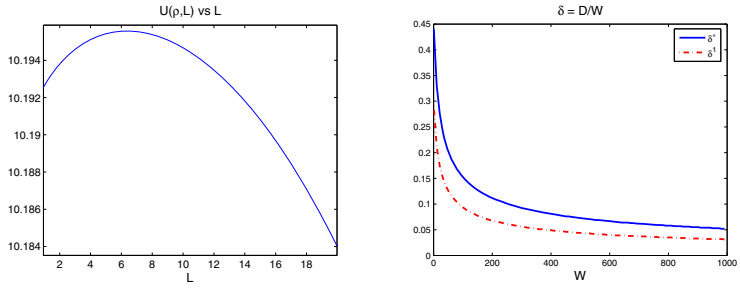
$$3\sqrt{\tilde{W}^D} - \sqrt{\tilde{W}} > 3\sqrt{W^D} - \sqrt{W} > 0.$$

Hence, we have proven that if $D < \frac{8}{9}W$, $3\sqrt{\tilde{W}^D} - \sqrt{\tilde{W}} > 0$. In this case, the first term of (12.26) is negative as well, which leads to:

$$\frac{2}{K}\frac{\partial U}{\partial L} < 0.$$

Thus, we have proven that if $D < \frac{8}{9}W$, utility is maximized at $L = 0$. Thus, the only equilibrium insurance contract is $(0,0)$.

If $D > \frac{8}{9}W$, there could exist an insurance contract, which improves user utility relative to U^*. See Fig. 12.2(a) for an example which shows how $U(\rho, L)$ is maximized at $L > 0$, and users may reach a higher utility with insurance.



(a) (b)

Fig. 12.2 (a) Unobservable case: U vs L ($L \in [0,20]$, $K = 1$, $W = 100$, $D = 99$) and (b) Observable case: $\delta^\dagger$ and δ^1 vs W.

Proof of Proposition 12.3

First, we notice that in any equilibrium, $L^\dagger = D$ and insurer profit is zero due to competition, as in Proposition 12.2. Hence, we restrict our analysis to full coverage only.

Second, in any equilibrium, user utility from deviation to no-insurance gives user a strictly lower utility. Indeed, assume the reverse. Suppose a user can deviate to v_i with no insurance and his utility without insurance is equal to his utility with insurance, i.e., $U(v_i, v^\dagger, 0, 0) \geq U(v^\dagger, v^\dagger, v^{\dagger 2}D, D)$. Consider an entrant insurer who offers him a contract $(v_i, v_i v^\dagger D, D)$ that offers non-zero coverage at actuarially fair price. By adopting this contract, the user improves his utility, which conflicts our assumption about the equilibrium. Therefore, the utility from deviation must be strictly lower and all users strictly prefer to buy insurance.

Lastly, we prove that in any equilibrium, all user contracts are identical. Assume the reverse, and let $(v_1, v_1\overline{v}D, D)$ and $(v_2, v_2\overline{v}D, D)$ be two contracts in equilibrium, with non-zero fraction of users buying each contract. Without loss of generality, we let $v_1 < v_2$, and thus $v_1 < \overline{v} < v_2$. From Section 12.3.2.2, we assume that insurers take $\overline{v}$ as given. Consider the contract $(\tilde{v}, \tilde{v}\overline{v}D, D)$ offered by an entrant insurer. Suppose this contract maximizes $U(\tilde{v}, \tilde{v}\overline{v}D, D)$:

$$\frac{\partial}{\partial \tilde{v}}\left(\sqrt{\tilde{v}} + K\sqrt{W - \tilde{v}\overline{v}D}\right) = 0.$$

$$\begin{aligned}
\frac{\partial}{\partial \tilde{v}}\left(\sqrt{\tilde{v}} + K\sqrt{W - \tilde{v}\overline{v}D}\right) &= 0 \\
\frac{1}{2\sqrt{\tilde{v}}} - \frac{K\overline{v}D}{2\sqrt{W - \tilde{v}\overline{v}D}} &= 0 \\
\frac{\sqrt{W - \tilde{v}\overline{v}D}}{\sqrt{\tilde{v}}} &= K\overline{v}D \qquad (12.28)
\end{aligned}$$

From (12.28), there is a unique solution for $\tilde{v}$ since the LHS is monotone decreasing. Hence, $\tilde{v} \neq v_1$ and $\tilde{v} \neq v_2$ since if either were true, then $U(v_1, v_1\overline{v}D, D) \neq U(v_2, v_2\overline{v}D, D)$, which is a contradiction. Thus, $U(\tilde{v}, \tilde{v}\overline{v}D, D) > U(v_1, v_1\overline{v}D, D) = U(v_2, v_2\overline{v}D, D)$ and insured users will be willing to deviate to this new contract. Thus, we have shown that two different contracts cannot be present in equilibrium, and we have proven that in any equilibrium, all users buy an identical contract.

Next, we prove that the equilibrium is unique. From (12.28), in any equilibrium, $\tilde{v} = \overline{v} = v^\dagger$, and we have

$$\begin{aligned}
\frac{\sqrt{W - v^{\dagger 2}D}}{\sqrt{v^\dagger}} &= Kv^\dagger D \\
\sqrt{W - v^{\dagger 2}D} &= Kv^\dagger\sqrt{v^\dagger}D \\
\frac{v^{\dagger 3}}{W - v^{\dagger 2}D} &= \frac{1}{(KD)^2}. \qquad (12.29)
\end{aligned}$$

From (12.29), there is a unique solution for the equilibrium $v^\dagger$, since the LHS is monotone decreasing. Thus, the equilibrium is unique.

Next, we determine how this unique $v^\dagger$ compares to v^*. When both $v^\dagger$ and $v^* < 1$, we can equate v^3 from (12.9) and (12.29) to get

$$\begin{aligned}\frac{1}{[2K(\sqrt{W}-\sqrt{W-D})]^2} &= \frac{W-v^2D}{(KD)^2}\\ \frac{D^2}{[2(\sqrt{W}-\sqrt{W-D})]^2} &= W-v^2D\\ \frac{W}{D}-\frac{D}{[2(\sqrt{W}-\sqrt{W-D})]^2} &= v^2\end{aligned}$$

Using (12.9) for $v^* < 1$ and denoting $\frac{D}{W}$ by δ, we have

$$\begin{aligned}\frac{W}{D}-\frac{D}{[2(\sqrt{W}-\sqrt{W-D})]^2} &= \frac{1}{[2K(\sqrt{W}-\sqrt{W-D})]^{4/3}}\\ \frac{1}{\frac{D}{W}}-\frac{\frac{D}{W}}{[2(1-\sqrt{1-\frac{D}{W}})]^2} &= \frac{1}{[2K\sqrt{W}(1-\sqrt{1-\frac{D}{W}})]^{4/3}}\\ \frac{1}{\delta}-\frac{\delta}{[2(1-\sqrt{1-\delta})]^2} &= \frac{1}{[2K\sqrt{W}(1-\sqrt{1-\delta})]^{4/3}}.\end{aligned}$$

Thus, we obtain an equation for δ:

$$\begin{aligned}(1-\sqrt{1-\delta})^{1/3}\left(\frac{(1-\sqrt{1-\delta})}{\delta}-\frac{\delta}{4(1-\sqrt{1-\delta})}\right) &= \frac{1}{[2K\sqrt{W}]^{4/3}}\\ (1-\sqrt{1-\delta})^{1/3}\left(\frac{1}{(1+\sqrt{1-\delta})}-\frac{(1+\sqrt{1-\delta})}{4}\right) &= \frac{1}{[2K\sqrt{W}]^{4/3}}\end{aligned} \quad (12.30)$$

We observe that the LHS is an increasing function of δ, which gives us a unique solution δ^* of (12.30). For $\delta \le \delta^*$, we have $v^\dagger \le v^*$, i.e., insurance improves the security level in the no-insurance Nash equilibrium. From (12.9), we know that when δ is low, v^* is high. This implies that insurance improves upon the no-insurance security level only when v^* is high. Let δ^1 denote the δ at which $v^* = 1$. Fig. 12.2 (b) depicts δ^1 and δ^* as a function of the wealth W ($K = 1$).

Social Planner

The contract offered by a social planner must be a solution to the following optimization problem:

$$\begin{aligned}&\max_{v,\rho,L} U(v,v,\rho,L)\\ &\text{s.t.} \quad v^2L \le \rho \text{ and } v \le 1.\end{aligned}$$

Next, we write the Lagrangian:

$$LAN = U(v,v,\rho,L) - \lambda_1(v^2L - \rho) - \lambda_2(v-1)$$

Taking the derivatives of LAN w.r.t. v, L and ρ and equating to 0 gives us the following equations.

$$\frac{\partial LAN}{\partial v} = \frac{\partial U(v,v,\rho,L)}{\partial v} - 2\lambda_1 vL - \lambda_2 = 0$$
$$(\frac{1}{2\sqrt{v}} - 2vK(\sqrt{W-\rho} - \sqrt{W-D+L-\rho})) - 2\lambda_1 vL - \lambda_2 = 0 \quad (12.31)$$

$$\frac{\partial LAN}{\partial L} = \frac{\partial U(v,v,\rho,L)}{\partial L} - \lambda_1 v^2 = 0$$
$$\frac{Kv^2}{2\sqrt{W-D+L-\rho}} - \lambda_1 v^2 = 0 \quad (12.32)$$

$$\frac{\partial LAN}{\partial \rho} = \frac{\partial U(v,v,\rho,L)}{\partial \rho} + \lambda_1 = 0$$
$$-\frac{Kv^2}{2\sqrt{W-D+L-\rho}} - \frac{K(1-v^2)}{2\sqrt{W-\rho}} + \lambda_1 = 0 \quad (12.33)$$

Further, from complementary slackness, we have

$$\lambda_1(v^2L - \rho) = 0, \quad (12.34)$$
$$\text{and } \lambda_2(v-1) = 0 \quad (12.35)$$

Note that $v \neq 0$, since that would require infinite security costs for the users. From (12.32), we conclude that $\lambda_1 > 0$ and thus the constraint (12.34) binds:

$$v^2L = \rho \quad (12.36)$$

Equating λ_1 from (12.32) and (12.33), we obtain:

$$\frac{K}{2\sqrt{W-D+L-\rho}} = \frac{Kv^2}{2\sqrt{W-D+L-\rho}} + \frac{K(1-v^2)}{2\sqrt{W-\rho}}$$

Canceling out $K/2 > 0$, we obtain:

$$\frac{1}{\sqrt{W-D+L-\rho}} = \frac{v^2}{\sqrt{W-D+L-\rho}} + \frac{(1-v^2)}{\sqrt{W-\rho}},$$

or

$$\frac{(1-v^2)}{\sqrt{W-D+L-\rho}} = \frac{(1-v^2)}{\sqrt{W-\rho}},$$

which leads to:

$$L = D \text{ if } v < 1. \tag{12.37}$$

Now, if $v < 1$, we can substitute (12.36) and (12.37) into (12.32) to get $\lambda_1 = \frac{K}{2\sqrt{W - v^2 D}}$. Substituting this value of λ_1, $\lambda_2 = 0$ (since $v < 1$) and (12.37) into (12.31), we get

$$\begin{aligned} \frac{1}{2\sqrt{v}} &= \frac{K}{\sqrt{W - v^2 D}} vD \\ \frac{v^3}{W - v^2 D} &= \frac{1}{(2KD)^2} \end{aligned} \tag{12.38}$$

Thus, if $v < 1$, it is the unique solution to (12.38) (since the LHS is monotone increasing).

References

1. Akerlof, G.A.: The market for 'lemons': Quality uncertainty and the market mechanism. The Quarterly Journal of Economics **84**(3), 488–500 (1970). URL `http://ideas.repec.org/a/tpr/qjecon/v84y1970i3p488-500.html`
2. Anderson, R., Böehme, R., Clayton, R., Moore, T.: Security economics and european policy. In: Proceedings of WEIS'08. Hanover, USA (2008)
3. Baer, W.S., Parkinson, A.: Cyberinsurance in it security management. IEEE Security and Privacy **5**(3), 50–56 (2007). DOI http://dx.doi.org/10.1109/MSP.2007.57
4. Böhme, R.: Cyber-insurance revisited. In: Proceedings of WEIS'05. Cambridge, USA (2005)
5. Bolot, J., Lelarge, M.: A new perspective on internet security using insurance. INFOCOM 2008. The 27th Conference on Computer Communications. IEEE pp. 1948–1956 (2008). DOI 10.1109/INFOCOM.2008.259
6. Fisk, M.: Causes and remedies for social acceptance of network insecurity. In: Proceedings of WEIS'02. Berkeley, USA (2002)
7. Gordon, L.A., Loeb, M., Sohail, T.: A framework for using insurance for cyber-risk management. Communications of the ACM **46**(3), 81–85 (2003)
8. Gordon, L.A., Loeb, M.P.: The economics of information security investment. ACM Trans. Inf. Syst. Secur. **5**(4), 438–457 (2002). DOI http://doi.acm.org/10.1145/581271.581274
9. Grossklags, J., Christin, N., Chuang, J.: Secure or insure?: a game-theoretic analysis of information security games. In: WWW '08: Proceeding of the 17th international conference on World Wide Web, pp. 209–218. ACM, New York, NY, USA (2008). DOI http://doi.acm.org/10.1145/1367497.1367526
10. H. Ogut, N.M., Raghunathan, S.: Cyber insurance and it security investment: Impact of interdependent risk. In: Proceedings of WEIS'05. Cambridge, USA (2005)
11. Hausken, K.: Returns to information security investment: The effect of alternative information security breach functions on optimal investment and sensitivity to vulnerability. Information Systems Frontiers **8**(5), 338–349 (2006). DOI http://dx.doi.org/10.1007/s10796-006-9011-6
12. Hofmann, A.: Internalizing externalities of loss prevention through insurance monopoly: an analysis of interdependent risks. Geneva Risk and Insurance Review **32**(1), 91–111 (2007)
13. Honeyman, P., Schwartz, G., Assche, A.V.: Interdependence of reliability and security. In: Proceedings of WEIS'07. Pittsburg, PA (2007)
14. Kunreuther, H., Heal, G.: Interdependent security. Journal of Risk and Uncertainty **26**(2-3), 231–49 (2003). URL `http://ideas.repec.org/a/kap/jrisku/v26y2003i2-3p231-49.html`

15. Kunreuther, H.C., Michel-Kerjan, E.O.: Evaluating the effectiveness of terrorism risk financing solutions. NBER Working Papers 13359, National Bureau of Economic Research, Inc (2007). URL http://ideas.repec.org/p/nbr/nberwo/13359.html
16. Majuca, R.P., Yurcik, W., Kesan, J.P.: The evolution of cyberinsurance. Tech. Rep. CR/0601020, ACM Computing Research Repository (2006)
17. Rothschild, M., Stiglitz, J.E.: Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. The Quarterly Journal of Economics **90**(4), 630–49 (1976). URL http://ideas.repec.org/a/tpr/qjecon/v90y1976i4p630-49.html
18. Schechter, S.E.: Computer security strength and risk: a quantitative approach. Ph.D. thesis, Cambridge, MA, USA (2004). Adviser-Smith,, Michael D.
19. Soohoo, K.: How much is enough? a risk-management approach to computer security. Ph.D. thesis, Stanford University
20. Stiglitz, J.E.: Information and the change in the paradigm in economics. American Economic Review **92**(3), 460–501 (2002). URL http://ideas.repec.org/a/aea/aecrev/v92y2002i3p460-501.html
21. Varian, H.: System reliability and free riding. In: Workshop on the Economics of Information Security, WEIS 2002. Cambridge, USA (2002)

Chapter 13
Potential Rating Indicators for Cyberinsurance: An Exploratory Qualitative Study

Frank Innerhofer–Oberperfler, Ruth Breu

Abstract In this paper we present the results of an exploratory qualitative study with experts. The aim of the study was the identification of potential rating variables which could be used to calculate a premium for Cyberinsurance coverages. For this purpose we have conducted semi-structured qualitative interviews with a sample of 36 experts from the DACH[1] region. The gathered statements have been consolidated and further reduced to a subset of indicators which are available and difficult to manipulate. The reduced set of indicators has been presented again to the 36 experts in order to rank them according to their relative importance. In this paper we describe the results of this exploratory qualitative study and conclude by discussing implications of our findings for both research and practice.

13.1 Introduction

The increased dependency on information technologies poses a variety of risks to organizations. The Congressional Research Service Report for Congress summarizes some surveys estimating the losses due to cyber attacks ranging *"[. . .] from $13 billion (worms and viruses only) to $226 billion (for all forms of overt attacks)"* [10]. Particularly since the Internet and the continuing cross-linking of information systems have become a backbone of modern business the headlines are constantly filled with news about devastating information security incidents. The resulting potential loss of reputation or brand image is a major driver for information security [14].

Research Group Quality Engineering
Institute of Computer Science
University of Innsbruck
A–6020 Innsbruck, Austria
e-mail: frank.innerhofer-oberperfler@uibk.ac.at, ruth.breu@uibk.ac.at

[1] DACH is the combination of the abbreviations of the countries of Germany (D), Austria (A) and Switzerland (CH).

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_13, © Springer Science+Business Media, LLC 2010

But not only the potential economic impact of information security moved the topic on the agenda of executives, where it continues to rise [13]. Especially the need to be compliant with the emerging regulatory landscape of the recent years (e.g., Sarbanes-Oxley Act in the US, the Basel II accord for financial services companies and others) requires executives to implement a proper risk management in their organizations. Adding to this, rating agencies recently announced that as part of their evaluation of the quality of management they will start to incorporate a review of enterprise risk management [11].

ISO/IEC 73:2002 defines risk management as *co-ordinated activities to direct and control an organization with regard to risk.* [21]. Risks can be managed through a combination of the following four strategic options [8]:

- reduce the risk,
- avoid the risk,
- transfer the risk,
- knowingly and objectively accept the risk.

One option to transfer the risk related to information technologies and information security which has emerged in the last years is Cyberinsurance [5, 16]. Cyberinsurance has been proposed as a market based solution for information security by different authors in the field [24, 36, 44]. Cyberinsurance coverage compensates the insured parties for a wide range of losses including but not limited to data loss, third party liabilities and others[2]. Estimates of the market for Cyberinsurance in the United States range from $450 to 500 million dollar annual gross written premium [4].

However, the insurance carriers still struggle to determine appropriate premium rates for covering cyber risks [14]. The reasons for these difficulties are missing actuarial loss data [14, 26] and the general lack of statistical data about information security incidents [25]. Therefore insurers put a range of exclusions in these policies, which again is a hindrance for wider market adoption Cyberinsurance as an instrument for risk transfer [14].

In this paper the results of an exploratory qualitative study with 36 experts from the DACH region will be presented. The objective of this study was the identification of potential rating indicators which could be used as a basis for the development of a risk classification system for Cyberinsurance.

The paper is structured as follows: In Section 13.2 the relevant notions and the context of this research are described. In Section 13.3 the research problem and our contribution will be outlined. In Section 13.4 the whole research is outlined using a step-by-step description. In Section 13.5 we present and discuss the results of our exploratory qualitative study. In Section 13.6 we discuss the limitations of this study. Section 13.7 positions this study with regard to related work from different fields. The paper is concluded with a conclusion and an outlook of the possible implications of our findings.

[2] For an overview of different insurance offerings the reader is referred to Baer [3].

13.2 Background

The business of insurance presumes an exposure – the possibility of a loss. If there is no chance of loss, there is no need for insurance [7]. Even more, if there would be no economic uncertainty regarding the occurrence, timing and magnitude of an event, there would be no reason for insurance neither [1]. An insurer has to determine the price – the insurance premium or rate – for assuming potential losses of a certain type. Therefore, risk rating is a very important aspect – if not the most important – of insurance.

The business impact of cyber risks can materialize in different ways [9]. First of all businesses may be exposed to a loss of property if hardware breaks down, is damaged or stolen (*physical resources exposures*). The range extends to a wide spectrum of financial losses (*financial resources exposures*) due to business interruptions or recovery expenses that are related to information technology failures or successful cyber attacks. Even damage to persons (*human resources exposures*) can be a result of information technology incidents (e.g. traffic management or clinical systems). This distinction according to *Loss Type* is made from a legal perspective with regard to the problem of economic loss which can be the result of a pure financial loss, a physical damage to property or personal injury [28, p 169].

Another dimension which is used to classify the losses due to cyber risks is the notion of *Loss Centre*. The dimension of *Loss Centre* describes to whom the loss happens [5]: *First-party losses* are those losses occurring directly to the insured organization, while *third-party losses* are losses which occurred to other parties (e.g., a customer, vendor or another third party). The dimension of *Loss Centre* is useful to distinguish coverages in the context of Cyberinsurance [4]. Third-party coverages are often also labeled as *liability* coverages.

There are different methods for identifying loss exposures which can be categorized along the following groups [2]:

- Document analysis,
- Compliance review,
- Personal inspections,
- Expertise within and beyond the organization.

In the domain of Cyberinsurance the methods of choice for identifying loss exposures are document analysis and personal inspections. Depending on the size and coverage of the insurance contract either more economic questionnaires or very costly evaluations conducted by a third party which is specialized in information security and performs personal and physical inspections on the clients site are used [33].

No matter what method is used for pricing and rating risks, they all have in common the need to determine factors that have an influence on the expected losses. These factors can be divided in two groups [7]: the *exposure base* (consisting of only one factor) and *rating variables*.

The exposure base is the basis on which the premium is calculated and it should therefore accurately reflect the expected losses. It is important to note, that the ex-

posure base is not the real exposure, but rather a proxy for the real exposure [7]. Examples for exposure bases are car-month or car-mile in automobile insurance, the sum of coverage provided, or total payroll for workers compensation. The insurance premium is normally calculated as a rate per exposure base. Consider as an example the premium property insurance, that is calculated as a fraction of the value of the insured property. If the value of property is used as an exposure base and one homeowner insures his 100 000 Euro home and pays 1 000 Euro premium (i.e. a rate of 1%), then a homeowner insuring his 1.000.000 Euro should pay 10.000 Euro premium (i.e. a rate of 1%). Exposure bases should be correlated proportionally to the expected losses [15].

However, in practice there are many other factors influencing the exposure to loss and therefore insurers use additional rating variables that allow to classify risks and adjust the premium accordingly [7]. Examples for premium-rating variables used e.g. in car insurance include among others age, gender[3], marital status, use of the automobile (pleasure or work), geography (location and area) and other criteria like the type, make and age of the automobile, multiple-car discounts and others [15,40].

13.3 Research Problem and Contribution

Since the market for Cyberinsurance is a relatively new one, there seems to be a lack of sophisticated models. While some authors propose approaches and frameworks for pricing cyber-insurance [16, 17, 29], there is – to the best knowledge of the authors – no research available about rating variables and indicators which could play a role in the premium-rating process.

In this regard the situation of Cyberinsurance is similar to the field of operational risk rating. Power denotes some of the controversies and discussions with regard to operational risk, that in the authors opinion are reflecting also the main problems related to the insurance of cyber risks: *"[. . .] three key domains of policy controversy have been, and remain, particularly visible: definitional issues, data collection and the limits of quantification."* [34]

In an emerging market like Cyberinsurance the insurance companies compete also with the quality of their premium-rating models. According to economic theory in the long run these premium-rating models should converge with the actual security risks, because competition sets an upper and profitability a lower bound to the premiums [6].

The authors gathered several questionnaires which are actually used by insurance companies to assess the exposure to cyber risks. While a comparison of these

[3] The use of rating variables like age, marital status and gender is part of enduring ongoing debate in the light of discrimination [42]. In the European Union in 2004 a Directive regarding equal treatment between men and women has been released, which states: *"(18) The use of actuarial factors related to sex is widespread in the provision of insurance and other related financial services. In order to ensure equal treatment between men and women, the use of sex as an actuarial factor should not result in differences in individuals' premiums and benefits. [. . .]"* [32]

questionnaires gave an idea about what factors are to be considered in the premium-rating process, the risk model and actuarial tables behind these questionnaires remain a business secret of the insurance carriers.

This paper aims to make a first step in the direction of developing better risk models and rating frameworks in the context of Cyberinsurance by addressing the following research question: **What are potential rating indicators for Cyberinsurance?**

In this paper the results of an exploratory qualitative study addressing this research question will be presented. The results of this study might be a useful resource for insurers who seek additional rating variables to further refine their premium-rating models. From a theoretical perspective the identified indicators provide a starting point for further research into influential risk factors and the development of risk assessment models.

13.4 Research Method

To answer the research question we have chosen a qualitative research approach. Due to the lack of statistical data about information security incidents [25] we have chosen to collect potential rating indicators from experts. Based on semi-structured expert interviews from the DACH region a list of such potential indicators which could be used for rating Cyberinsurance premiums was identified.

In this section the research method is described using a step-by-step description of the whole process, from the preparation to the final ranking of indicators (cf. Figure 13.6). In the description of the process we partly follow the guidelines of Myers and Newman for conducting qualitative interviews [30].

13.4.1 1. Step: Preparation, Constructs

Before conducting the qualitative expert interviews a literature review about the related work in the field of Cyberinsurance was conducted, which is partly subsumed in Section 13.2 and in Section 13.7. To achieve the objectives of this research and owing to the exploratory nature of the research problem, the authors decided to conduct semi-structured expert interviews.

We wanted to explore the ideas about rating indicators that the experts might come up with as openly as possible. Therefore we have used a minimal structure for the interview using several sections, which each address a particular aspect of rating indicators. In this Section we will outline the constructs and concepts which have been used to structure the interviews.

13.4.1.1 Exposure and Quality

The first concept which has been used for structuring the interviews is the distinction between *exposure* and *quality*. With exposure we characterize the inherent risk level of an organization. The concept of quality stands for the quality of the IT risk management in an organization. This distinction was based on practical input from our project partners, who outlined the necessity to not only focus on the inherent risk of an organization but also the quality of its security and risk program [38, p. 347].

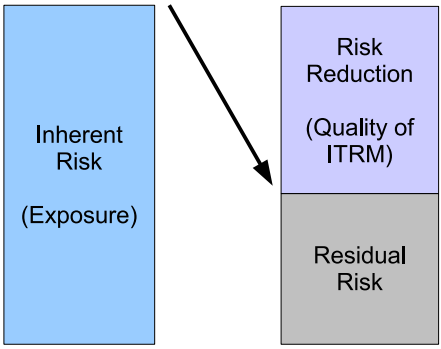


Fig. 13.1 Concepts of exposure and quality (adapted from ISO/IEC 13335-1:2004 [20]).

Figure 13.1 outlines these two concepts. The exposure stands for the inherent risk level and the quality of IT risk management acts as a proxy for the risk reduction capabilities in an organization. In a premium-rating framework, variables which indicate a high quality of IT risk management would lead to a reduction of the premium.

13.4.1.2 Loss Centre

The second concept which we have introduced is the concept of loss center which was already described in Section 13.2. The loss center can be either first-party or third-party. This distinction was also introduced based on comments from practitioners, since these two types of coverages often present different lines of business of an insurer, are treated differently from a legal perspective and impose different requirements [4].

The concept of loss center was applied to further distinguish the initial concept of exposure. We now have the exposure to first-party losses and the exposure to third-party losses. The concept of quality remained untouched by a further classification using the concept of loss center. The authors believe that a high quality risk

management is evenly capable to reduce both first-party and third-party losses to an acceptable level.

13.4.1.3 Layer Model

The last construct we have introduced to structure the interview is a layer model for information management. This concept was introduced to further classify the concept of third-party loss exposure. Why only the concept of third-party loss exposure and not also the one of first-party losses? The rationale behind this further classification of third-party loss exposure is based on the fact, that it is mainly companies from the IT sector (labeled as IT-Providers in the questionnaire) who are requesting this type of insurance coverage.

To account for different types of IT businesses we have explored different constructs for classifying the third-party loss exposure according to the type of offered IT service or product. The solution we came up with is a layer model which is present in many enterprise architecture frameworks. Examples for layered models in the technology domain include e.g. the well known Open Systems Interconnection Basic Reference Model [45] (OSI Model).

From the domain of information management, layered models include Wollnik's Three-Layer-Model of Information Management [43] or Krcmar's Layer-Model [27]. Both models[4] have in common three different layer of abstraction to distinguish activities of information management in enterprises.

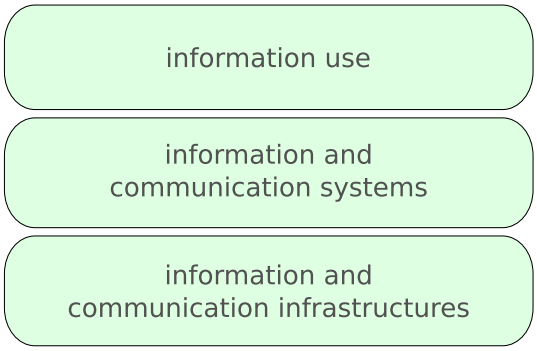


Fig. 13.2 Layer model of information management [27, 43].

The three layers are organized as follows from top to bottom (cf. Figure 13.2): The top layer represents the abstraction of *information use* in an organization, the middle layer depicts *information and communication systems* while the bottom layer

[4] Krcmar's layer model of information management includes a fourth "layer" that encompasses all three abstraction layers and represents the managerial functions related to information management [27].

is used to handle *information and communication infrastructures* [43]. These three layers have been used as an abstraction to further distinguish different types of IT services and products and the related indicators for third-party loss exposures.

13.4.1.4 The Resulting Questionnaire

The questionnaire we have used to structure the interview is the result of an initial pre-test during which we tested the feasibility of the first versions of our questionnaire. We conducted the pre-test with 5 interviewees. During the pre-test it became clear, that some questions where very difficult to answer and these have therefore either been eliminated or reformulated. Also the ordering of questions has been changed to a more comprehensive logical order.

Table 13.1 The resulting questionnaire.

Section 3: IT Business Risk Exposure Indicators (first-party losses)
1. What are in your opinion relevant drivers and indicators for the IT Business Risk Exposure of an organization?
Section 4: Indicators for the Quality of the IT Risk Management
1. What are in your opinion indicators for the quality of the IT Risk Management efforts in an organization?
Section 5: IT Business Risk Exposure Indicators (IT-Providers with regard to third-party losses)
in general   1. Which indicators reflect the potential of IT-Providers in general to cause third party losses due to IT Business Risks?
IT-Infrastructure   2. Which indicators reflect the potential of IT-Infrastructure Providers to cause third party losses due to IT Business Risks?
Information Systems   3. Which indicators reflect the potential of Information Systems and Application Providers to cause third party losses due to IT Business Risks?
Information Use   4. Which indicators reflect the potential of Information Providers and Processors to cause third party losses due to IT Business Risks?

The resulting questionnaire which we have finally used for the interviews is outlined below (cf. Table 13.1). In the beginning of the interview we had two additional Sections with a general introduction of the important terms (exposure, quality, loss center) and additional questions about the interviewees to create a profile of the participants.

The questions of the interview which targeted towards the research question were organized in three Sections. The first Section questioned indicators related to the

first-party loss exposure. The second Section focused on the quality of the IT risk management. The third Section aimed at indicators related to the third-party loss exposure. The third Section was further subclassified using a general class and the three layers of the layer model (cf. Figure 13.2).

13.4.2 2. Step: Selection of Experts

Based on the targeted research objective and the aim of the study we started to design candidate profiles for the interviews. The research sample selected for this study was built using a combination of purposive and snowball sampling. Taking into consideration the sensitive nature of the research topic [25] the authors decided to interview candidates which have an internal or external perspective on the topic.

The internal perspective was captured by professionals who are working in an organization having the role of CISO (Chief Information Security Officer), Risk Manager or General Management and IT Management. The external perspective was captured by professionals who are either working as auditors, consultants in the field of IT risk management or information security experts and have thus a broad experience from many different organizations. Figure 13.4 outlines the experience profile of the participants. 70% of the sample have professional experience in the field of information security or risk management of more than 10 years. Figure 13.3 outlines the designation profile of the entire sample.

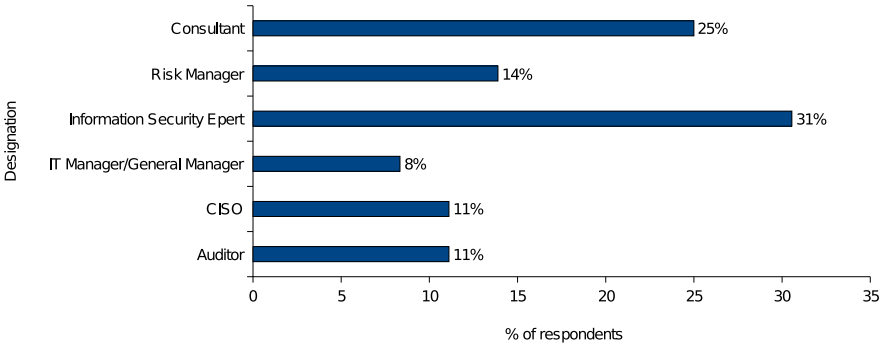


Fig. 13.3 Designation profile of the respondents.

During the process of acquiring experts for the interviews, we put special emphasis on the fact that the interviews are not aiming towards any kind of sensitive information about the risks in the respective organizations. This was a very important issue, which was brought up by most of the candidate experts.

By emphasizing that we were interested only in their expert knowledge and what indicators they deem important from a general point of view, we overcame this

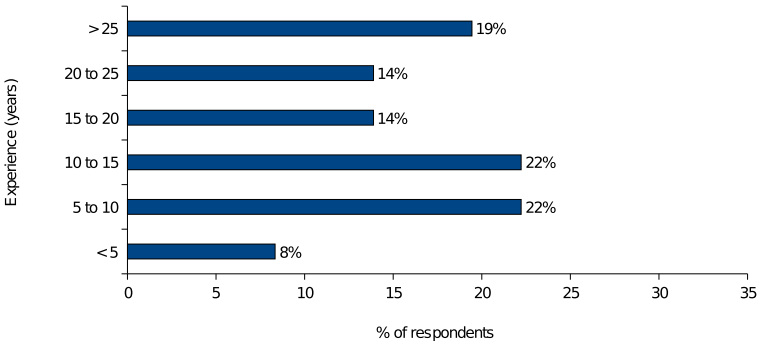


Fig. 13.4 Experience profile of the respondents.

critical point in all cases. In general the willingness to contribute and support the research project was very high and we had only few refusals to participate in the interviews, and these few were mainly attributable to time constraints on behalf of the candidates.

The first group of experts were purposely identified from a pool of contacts which our research group has built up as part of its ongoing activities to create a regional forum of information exchange about information security involving practitioners and academics. The second group of experts was also selected purposely from a pool of participants who attended the expert forum about IT and Internet Risks in 2005 organized by Swiss Re in Munich.

Beginning with these two initial groups of participants from Austria and Germany, we employed snowball sampling to identify further suitable candidates for the interviews based on the recommendation of the experts from the first two groups. At the closing of each interview the experts were asked whether they could name additional professionals who match our requirements. The requirements for being selected in the sample were the following:

- The candidates are involved in IT or information security risk management activities.
- Each of the candidates should have a minimum of three years of relevant experience working with IT risks.

The interviews were conducted in the period between April 2006 and October 2007. We have interviewed a total number of 36 experts.

13.4.3 3. Step: Generation of Statements

The interviews lasted between 50 minutes and 90 hours depending on the time constraints of the interviewees. 26 of the 36 interviews were conducted in a personal

face-to-face meeting with the participants, typically in their office and in a few cases at other meeting places like airports. The other 10 interviews were conducted via phone conversations.

Most of the interviews – given explicit allowance – have been tape-recorded to allow for a later analysis and transcription. The interviews that were not recorded were registered by taking notes and creating a mind protocol shortly after the interview. The recorded interviews were transcribed to identify the main statements, their relations and examples.

The interviews started with an opening which involved an introduction. During this introduction the purpose of the interview and the research project context were explained. After gathering information about the interviewee's role and experience (cf. Figure 13.3 and 13.4), the interview entered in its main stage with the key questions (cf. Table 13.1).

The semi-structured interviews contained just the three main sections about indicators which highlight the first-party loss exposure, the quality of the IT risk management and the third-party loss exposure. The question about the third party loss exposure was further subdivided using the different layers of abstraction of the layer model (cf. Figure 13.2).

The questions enabled the 36 participants to generate a total of 976 statements, which were written down on paper during the interviews. At the end of the interview the interviewees were presented with the list of the statements they generated to check for inconsistencies and completeness. In the case of face-to-face meetings the statements were highlighted on paper and presented to the participants. In the case of phone conversations the complete list of statements was repeated at the end of the interview.

13.4.4 4. Step: Interpretation and Consolidation of Statements

Shortly after the interviews took place they were transcribed and the relevant text sequences were highlighted. After the transcription we have used concept mapping for structuring and organizing the gathered knowledge. Concept mapping is an approach which supports the graphical representation of statements [12, 31]. Concept mapping offers some additional possibilities compared to a pure text based analysis [22]. Especially the possibility to outline connections between the mapped statements has been useful in the process of consolidating the gathered knowledge.

For each of the 36 interviews a concept map was created, which included the main statements and included also additional explanations, examples and clarifications. For creating the concept maps we used the software CmapTools from the Institute of Human and Machine Cognition[5]. Figure 13.5 shows an excerpt from one of the concept maps.

[5] Download of IHMC CmapTools: `http://cmap.ihmc.us/`

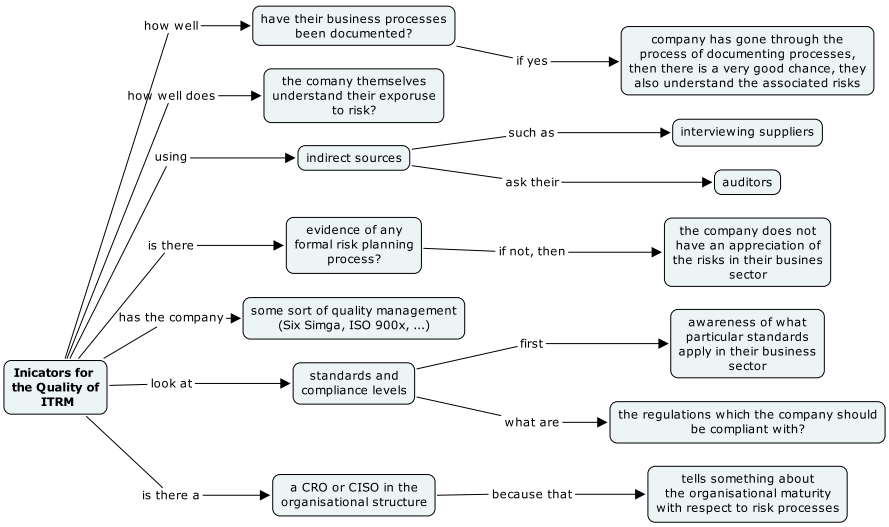


Fig. 13.5 Example concept map of an interview (excerpt).

Once all the interviews transcripts and protocols protocols have been transformed to concept maps, these single concept maps were exported to create a unified concept-map including the concepts and propositions of all interviewed experts. The purpose of this combined concept map was the step-wise consolidation of the 976 statements of all 36 interviewees.

The explanations and examples related to each of the statements facilitated the process of interpreting and consolidating the statements in similar groups. The criteria for grouping the statements were syntactic similarity and semantic similarity. Concept mapping has proved to be a valuable tool for this consolidation process. However, we used concept mapping only as a tool for representation of the gathered knowledge, we did not employ multivariate statistical analyzes as described by Daley [12] or Trochim et al. [22, 39].

By linking the statements of the single experts to similar groups it was possible to identify emerging themes and levels of hierarchy [12]. However, the purpose of interpreting and consolidating the statements was not to analyze the hierarchical relations of the statements and their connecting links, but to compile a reduced and consolidated list of statements. After the systematic grouping and categorization of the statements we have reduced the list of 976 statements to a list of 198 consolidated indicators (cf. Appendix 13.9).

13.4.5 5. Step: Reducing the Resulting List of Indicators

This list of 198 consolidated indicators was discussed and presented to actuaries who are experienced in the external risk assessment of organizations in the context of Cyberinsurance. The intention was to reduce the list of 198 indicators to a list of variables, which are considered useful for practical uses in the context of Cyberinsurance. During a workshop held with three actuaries of our project partner the 198 were presented, discussed and a selection based on specific criteria was made.

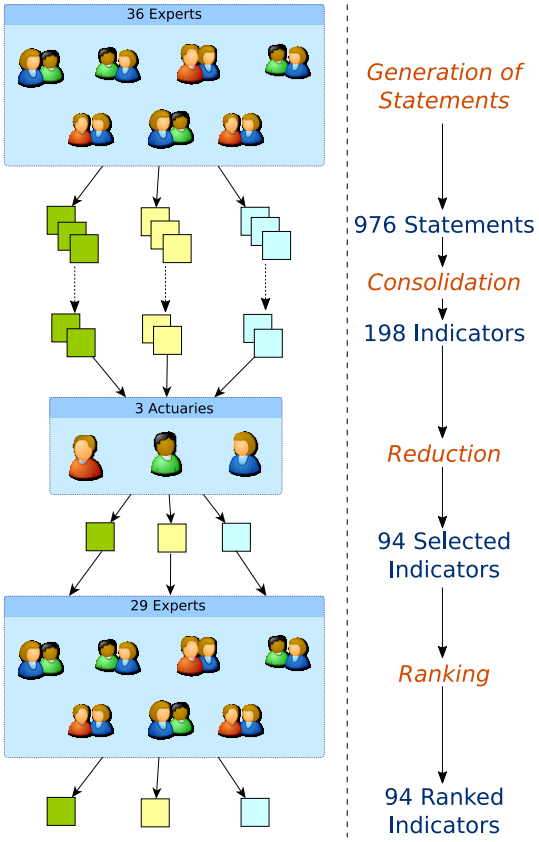


Fig. 13.6 Schematic overview of the research process.

The American Academy of Actuaries lists the following basic principles that should be present in any sound risk classification system and therefore also in

the selection of rating variables[6]. The principles state that a classification system should [1]:

- reflect expected cost differences,
- distinguish among risks on the basis of relevant cost-related factors,
- be applied objectively,
- be practical and cost-effective,
- be acceptable to the public.

For selecting the indicators in this research project we used only two criteria which were taken from a paper of Bouska. Bouska cites Webb who uses the following three criteria for selecting exposure bases: *"First and foremost, of course, it should be an accurate measure of the exposure to loss. Second, it should be easy for the insurer to determine. Finally, it should be difficult for the insured to manipulate."* [7] The last two criteria were actually used to filter the indicators that were identified in the first round:

- *Are the indicators measurable?*
- *Are the indicators unmistakable and difficult to manipulate?*

The workshop with the three actuaries resulted in the selection of 94 indicators which were deemed useful for premium-rating in the context of Cyberinsurance, since they were considered measurable and objectively answerable.

13.4.6 6. Step: Ranking Indicators

The 94 indicators which were selected by the three actuaries were again sent to the initial 36 experts asking them to rank the indicators according to their relative importance. For ranking the indicators we used a 10-point Likert scale as illustrated in Figure 13.7. The ranking of the 94 indicators was collected using a web-based questionnaire (cf. Figure 13.8).

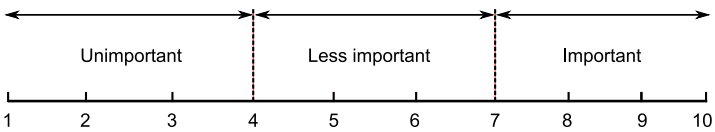


Fig. 13.7 Intervals of rating [18].

In the final step of ranking the indicators 29 of the initial 36 experts have participated. Seven experts were not able to participate in the final ranking due to time

[6] For a more detailed treatise see Finger, who provides an overview and a discussion of criteria for selecting rating variables. Finger groups the criteria in four categories, namely: *"actuarial, operational, social, and legal"* [15].

constraints. Finally, descriptive statistics were used to analyze the ranking of the indicators.

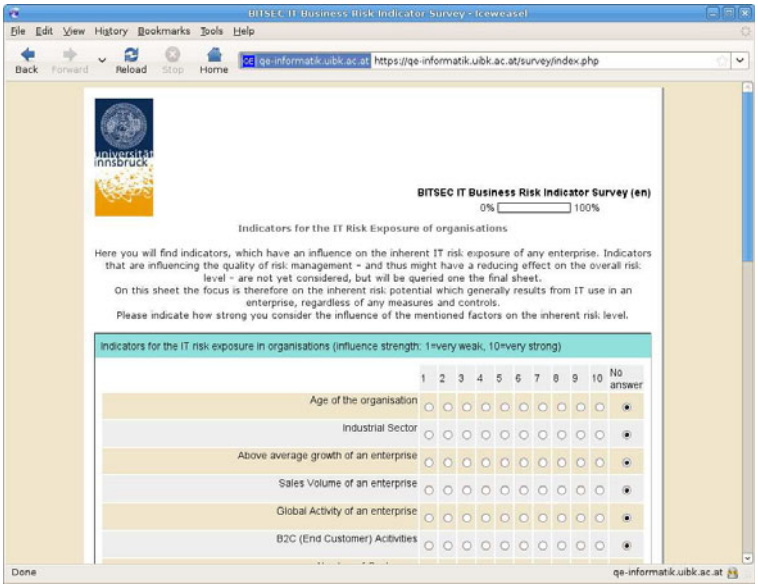


Fig. 13.8 Screenshot of the web based survey.

13.5 Results

The results of the whole research process are outlined in Table 13.2, 13.3 and 13.4 on the previous pages. Since the questions of the interview targeted towards the identification of indicators one might have expected, that in the final ranking all of the selected indicators should have been attributed a high importance rating.

Regarding the first-party loss exposure indicators indicators which focus on the dependency of the business on IT have been ranked highest. The reader might remark that the indicators are rather abstract and already known or obvious. This is especially true for the highest ranked indicators like *Critical dependency of business processes on IT* and *Low failure tolerance with regard to IT*. However, we have willingly included these indicators, as related statements were recurring again and again. The highest ranked indicator (cf. Table 13.2, Rank #1) has emerged out of the consolidation of 51 statements from the total of 976 statements (cf. Appendix 13.9).

What was puzzling the authors was that an indicator like *Sales Volume of an enterprise*, which is often used as an exposure base in Cyberinsurance contracts has been ranked as a less important indicator with a mean ranking of only 4,61.

Table 13.2 Ranking of first-party loss exposure indicators.

Rank	Indicator	Mean	SD
1	Critical dependency of business processes on IT	8,83	1,13
2	Low failure tolerance with regard to IT	8,78	1,20
3	Processing sensitive data with high confidentiality requirements	8,38	1,31
4	Existence of worth-protecting know how, patents and otherwise valuable information	8,29	1,70
5	High demands on the availability of data and systems in the organization	8,19	1,63
6	Online execution of Business Processes	7,83	1,62
7	Environmental and physical risks at the location of the data center	7,79	1,70
8	Link-ups of external partners to the enterprise IT	7,67	1,11
9	High level of automation in the production of goods and services	7,50	1,71
10	Data recoverability in data loss scenarios	7,26	2,54
11	Just-in-time supply/delivery relationships with partners	7,25	1,73
12	Above average growth of an enterprise	7,11	1,74
13	Industrial Sector	7,07	1,92
14	Labor turnover rate (in general)	6,90	2,27
15	Availability of qualified workforce	6,85	1,68
16	Use of mobile devices in the organization	6,72	2,29
17	Outsourcing of IT processes including coordination and control	6,68	1,94
18	IT-personnel / overall number of employees ratio	6,57	1,95
19	Demand on the professional qualification of employees (in general)	6,42	1,81
20	Number of PC-Workplaces in the enterprise	6,34	1,48
21	Global Activity of an enterprise	6,14	1,75
22	B2C (End Customer) Activities	6,00	1,93
23	Private Internet use of employees in the organization	5,93	2,21
24	Centralized IT-Infrastructure	5,86	2,08
25	Number of employees (overall)	5,62	1,89
26	Separate IT budget existent	5,62	2,06
27	Operation of standardized IT solutions	5,48	2,02
28	Number of Customers	5,21	1,94
29	Geographical distance between day-to-day business and IT production	4,89	2,18
30	Sales Volume of an enterprise	4,61	2,11
31	Age of the organization	3,69	2,06

The third-party loss exposure indicators have been collected using the classification of the layer model as outlined in Section 13.4.1. As can be seen in Section 13.9.2 in the Appendix 13.9, the classification we have employed to further classify the third-party exposure indicators according to the layers of information management did never yield more than 10 indicators per class. Therefore in the result the indicators for third-party loss exposure have been compiled in a unique list (cf. Table 13.3).

The highest ranked indicator for the third-party loss exposure was the degree of *Control and Coordination of outsourced customer processes by the outsourcing provider* with a mean of 8,21, followed by *Access to central systems and information of customers* with a mean of 7,96. Interesting to note, that also in the case of

Table 13.3 Ranking of third-party loss exposure indicators.

Rank	Indicator	Mean	SD
1	Control and Coordination of outsourced customer processes by the outsourcing provider	8,21	1,57
2	Access to central systems and information of customers	7,96	1,73
3	Quality of Patch-Management for the offered Information Systems	7,52	1,67
4	High service level requirements	7,48	1,86
5	Traceability of provider actions and interventions at customers sites	7,39	1,75
6	Extensive test procedures during development and deployment	7,16	2,05
7	Adoption of standardized methods and practices in software engineering	7,08	1,81
8	Availability of qualified workforce at the provider	6,92	1,80
9	Clarity and detailing of Service Level Agreements	6,84	2,36
10	Outsourcing of Information Systems Development (Off-shoring)	6,81	1,94
11	Remote maintenance of systems installed on customers sites	6,76	1,75
12	Existence of Update- and Version-management of the offered information systems	6,75	1,73
13	Sole Seller for customers	6,73	2,68
14	Offer of Backup Services (Data backup)	6,71	1,86
15	Offered IT solutions could cause bodily injury	6,65	2,58
16	Quality and standardization of project management	6,54	2,11
17	Definition of liability provisions and contractual sanctions in Service Level Agreements	6,46	2,42
18	High portion of employees occupied with further technical development and innovation	6,25	1,96
19	High portion of internal activity in offered services	6,25	1,88
20	Customers have the possibility to switch to alternative providers	6,08	2,32
21	Industry classification of customers	5,92	2,34
22	Offering of training concepts and courses for customers	5,88	1,71
23	Certification of employees/organization by manufacturers or for specific products	5,67	2,47
24	Longevity of customer and contractual relationships	5,48	1,78
25	Sales Volume per customer	5,39	2,58
26	Number of employees of the provider	5,04	1,88
27	Sales Volume of provider	5,04	2,30
28	Provider offers a Central Contact Point or Service Desk	4,92	2,00
29	Customer structure (Number and quality of customers)	4,87	2,05

the third-party loss exposure indicators the *Sales Volume of provider* ranked third-lowest on the relative importance scale with a mean of only $5,04$.

The indicators for the quality of the IT risk management have generally a higher mean ranking than the first-party and third-party loss exposure indicators. The top 13 first-party exposure indicators have a mean ranking higher than 7. In the case of the third-party loss exposure indicators only 7 indicators have a mean ranking higher than 7. The ranking of the indicators for the quality of the IT risk management has yielded 25 indicators with a mean ranking higher than 7.

Table 13.4 Ranking of quality indicators for the IT risk management.

Rank	Indicator	Mean	SD
1	Existing Role of a Risk Officer or Security Officer	8,78	1,26
2	Business Continuity concept and contingency and emergency plans available	8,46	1,35
3	Continual improvement process in Risk Management (PDCA Cycle)	8,41	1,37
4	Existence of an institutionalized risk management in the organization	8,37	1,48
5	Policies concerning handling of confidential information	8,15	1,58
6	Security Policies for employees	8,04	1,35
7	Acceptance Testing required prior to release of new technologies and product versions	8,04	1,37
8	Investments in further education and training of employees to increase security awareness	8,04	1,15
9	Existence of a proper IT Risk Reporting	8,04	1,66
10	Periodical internal and/or external audits	8,00	1,54
11	Documentation of IT-Infrastructure	7,96	1,09
12	Existence of Protection requirements analysis	7,85	1,76
13	Password Policy available	7,85	1,59
14	Accounted budget for IT Security present	7,85	1,36
15	Existence of an IT-Governance function	7,74	1,70
16	Redundancies in the technical infrastructure	7,65	1,61
17	Security Manual available	7,63	1,68
18	Systematic Problem and Solution Management in IT	7,59	1,85
19	Policies and Guidelines for and control of external service providers	7,52	1,57
20	International standards and best practices orientation (in general and with regard to IT)	7,48	1,73
21	Physical protection measures	7,38	1,79
22	Ratio of IT employees dedicated to security	7,33	1,85
23	Physical control and registration of visitors at the entrance area	7,31	1,75
24	Maintenance of a proper Loss database	7,22	1,62
25	Measurement of performance indicators and and operating figures for assessing IT processes	7,00	1,70
26	IT-Management reports directly to the board level or is part of the board of directors	6,96	2,72
27	Information Security Certification(s)	6,96	2,13
28	IT-Service-management approach based on Standards	6,81	1,95
29	Quality Assurance of published content (Legal Review)	6,78	1,88
30	Employee Background Check	6,58	2,06
31	Quality Management System(s) Certification(s)	6,48	1,99
32	Definition of Life cycles for the IT-Infrastructure	6,44	1,87
33	Size of the enterprise	5,59	2,19
34	Age of the organization	4,96	2,14

13.6 Limitations

In this Section we discuss some limitations of this research. First of all, as with any Delphi-type study, the results are based on a sample with a limited number of subjects. While we tried to compose the sample with professionals who have a broad and long experience managing security or risks inside organizations or as external consultants and experts, we cannot claim any kind of representativeness of our sample. We used a convenience sample based on our direct relations with industry. Also the selection of additional participants was not random since we employed a snowball strategy to identify additional candidates.

Another potential weakness is the cultural background of the experts. Since the sample of interviewees is taken solely from the DACH region, there is potential bias in the findings, due to the lack of cultural diversity. Schmidt et al. conducted an international Delhi study to identify software project risk with panelist from Hong Kong, Finland and the United States [35]. They identified differences in the relative importance of risk factor across the various cultures. Thus in this paper, the results might be biased since the whole sample has a common cultural background.

Another significant limitation of this research is the fact that a great part of the study was conducted in German, necessitating the translation of the indicators in English for this paper.

A limitation of the research process as outlined in Section 13.4 is the fact that we have not conducted any type no validation of the consolidated list of factors as it would be in a Delphi type of study. Due to the restricted time budgets of the participating experts, it would not have been realizable to introduce one more step to check the results of the consolidation process. This might introduce a potential bias since the authors of this study have interpreted the statements and consolidated them.

Another potential point of criticism is the decision not to reduce the resulting list of indicators with the experts who participated in the interviews. Instead the reduction was done in a workshop with three actuaries who are experienced in risk assessments in the context of Cyberinsurance. These three actuaries were not interviewed and taking part in the initial interviews. This raises the possibility of varying interpretations of the indicators. In addition it introduces a potential bias since the initially interviewed experts had no influence on the selection of indicators. To reduce the risk of misinterpretations we have used the concept maps including the statements and examples as an additional aid during the selection workshop with the three actuaries.

The premium-rating models and indicators contained in the actuarial tables of the underwriters are a business secret of the insurance companies. How they do calculate rates can not be transparently said. Therefore the indicators that were identified in this chapter might already be in use and not contribute to an improvement of the state of the art in practice. Despite this limitation we believe that the results of this research represent an interesting resource for practitioners and there may be some additional factors, that might be worth incorporating into the existing models. Clas-

sification systems evolve over time [15, 40] and hence the rating variables used for classification will also continue to improve.

Regarding the theoretical value of these lists of indicators, we have just marked a first step in the direction of developing a rating model for cyber risks. We have not conducted any type of evaluation regarding the validity of the identified indicators. In addition, there might be interesting relations and an interplay between the various indicators and their influence on the actual risk exposure. These are interesting questions that might stimulate further research.

13.7 Related Work

To the best knowledge of the authors there are no related works focusing on indicators for premium-rating in the context of Cyberinsurance. There are however different works from other fields which are providing risk factors.

This related work on risk factors in information systems and information technology provides additional valuable input for developing premium-rating models. Some of these works are focusing on software development risk such as Jiang et al. [23] of software project risk such as Schmidt et al. [35]. Sherer and Alter have conducted a review of different risk models used in the information systems literature [37].

The only exposure model that the authors came across was the *Risk Exposure Model for Digital Assets* published in Turban et al. Their exposure model for digital assets contains five general factors [41]:

- Asset's value to the company
- Attractiveness of the asset to a criminal
- Legal liability attached to the asset's loss or theft
- Operational, marketing, and financial consequences
- Likelihood of a successful attack against the asset

This risk exposure model is focusing on digital assets and therefore provides also valuable input for rating cyber risks. In contrast we are focusing on a risk exposure model for organizations and therefore focus on a different level of abstraction.

13.8 Conclusions and Outlook

The results presented in this paper provide lists of indicators which could serve as potential candidates for rating variables for Cyberinsurance. The indicators have been consolidated from semi-structured expert interviews with 36 participants from the DACH region. After a reduction of the indicators to a set of 94 indicators which are measurable and objectively answerable, the indicators were again presented to

the initial 36 experts. In the ranking step 29 experts have ranked the indicators according to their relative importance.

These indicators could be used to build new or refine existing risk classification systems and premium-rating models. Due to the lack of concrete scenarios with quantified losses, it was out of scope of this research to validate which of these potential rating variables actually reflect the risk exposure.

Further research would also investigate the relations and the interplay between the listed indicators. For some indicators, which are rather abstract and difficult to objectively assess, it would be interesting to research better indicators which could act as proxies for them.

Another important question is regarding the relation and the interplay between these indicators. We have already identified some relations during the course of the interviews and in the combined concept map of all statements. However, we have not systematically analyzed the hierarchical structure of the statements.

The list of 94 ranked indicators and the initial list of 198 indicators in the Appendix provide a starting point for developing a model and a framework for risk rating in the context of Cyberinsurance. A task that is surely left to do is to organize the presented indicators in meaningful categories. The authors believe that some of the identified indicators might only be relevant for certain types of coverages. Such a categorization of exposure indicators would also provide an excellent baseline for developing a theoretical exposure model for organizations.

Another task that is left to future work is the important issue of operationalization of these indicators. While some of the indicators are binary and can be easily answered using yes or no, most of the indicators are not binary. Some indicators might be measured using a qualitative range of values to reflect the degree to which they apply in a certain organization. We are currently investigating the operationalization of these indicators.

In a previous publication we have analyzed publicly announced security incidents to identify different types of losses related to security incidents [19]. Matching the identified indicators presented in this paper with damages and losses resulting from security incidents will provide further valuable insights.

Acknowledgments

This research is part of the results of the research project BITSEC between the Research Group Quality Engineering, Swiss Re Germany and Arctis Softwaretechnology which focused on the external assessment of IT related risks in the context of cyber-insurance. The research cooperation was funded by the Austrian Research Agency FFG.

13.9 Appendix

13.9.1 First-party loss exposure indicators

Table 13.5: Exposure indicators for first-party losses

Indicator	Count	Selected
Critical dependency of business processes on IT	51	√
Existence of worth-protecting know how, patents and otherwise valuable information	20	√
Industrial Sector	16	√
Environmental and physical risks at the location of the data center	12	√
Demand on the professional qualification of employees (in general)	10	√
Number of employees (overall)	8	√
Availability of qualified workforce	7	√
Above average growth of an enterprise	6	√
Link-ups of external partners to the enterprise IT	6	√
High demands on the availability of data and systems in the organization	5	√
Low failure tolerance with regard to IT	5	√
Processing sensitive data with high confidentiality requirements	5	√
Separate IT budget existent	5	√
Centralized IT-Infrastructure	4	√
High level of automation in the production of goods and services	4	√
Labor turnover rate (in general)	4	√
Number of PC-Workplaces in the enterprise	4	√
Online execution of Business Processes	4	√
Outsourcing of IT processes including coordination and control	4	√
Geographical distance between day-to-day business and IT production	3	√
Just-in-time supply/delivery relationships with partners	3	√
Number of Customers	3	√
Operation of standardized IT solutions	3	√
Private Internet use of employees in the organization	3	√
Sales Volume of an enterprise	3	√
Data recoverability in data loss scenarios	2	√

Continued on next page

Table 13.5 – continued from previous page		
Global Activity of an enterprise	2	√
IT-personnel / overall number of employees ratio	2	√
Use of mobile devices in the organization	2	√
Age of the organization	1	√
B2C (End Customer) Activities	1	√
Enterprise subject to strict legal regulations	11	
High-profile enterprise	11	
Market leader	4	
Age of the IT infrastructure	3	
Owner Management	3	
Short-term optimization	3	
High competitive pressure	2	
High number of transactions per day	2	
Highly dynamic business environment	2	
Highly dynamic IT landscape	2	
Homogeneous IT landscape	2	
Obligations to supply and exchange data	2	
Sufficient safety stock	2	
Systems based on open standards	2	
Current high risk IT projects	1	
Early adopter and use of recent technologies	1	
Enterprise operates in a high technology sector	1	
Enterprise operates in a niche market with high market share	1	
High number of heterogeneous applications	1	
Highly complex production processes	1	
Highly interlinkage of processed data	1	
Incidents can strongly affect customers	1	
Operating system in use	1	
Potential to damage the economy	1	
Production of goods and services in front of customer	1	
Products or services subject to strict legal regulations	1	
Reputation in the market	1	
Revenues per employee	1	
Stock turnover ratio	1	
Strict contractual obligations toward customers	1	
Technical state of the art infrastructure	1	
Volume of stored critical data	1	

13.9.2 Third-party loss exposure indicators

Table 13.6: Exposure indicators for third-party losses (in general)

Indicator	**Count**	**Selected**
High service level requirements	12	√
Industry classification of customers	12	√
High portion of internal activity in offered services	11	√
Clarity and detailing of Service Level Agreements	10	√
Availability of qualified workforce at the provider	9	√
Sales Volume of provider	8	√
Certification of employees/organization by manufacturers or for specific products	6	√
Longevity of customer and contractual relationships	6	√
Traceability of provider actions and interventions at customers sites	5	√
Customer structure (Number and quality of customers)	4	√
Offered IT solutions could cause bodily injury	4	√
Provider offers a Central Contact Point or Service Desk	4	√
Sole Seller for customers	3	√
Definition of liability provisions and contractual sanctions in Service Level Agreements	1	√
High portion of employees occupied with further technical development and innovation	1	√
Number of employees of the provider	1	√
Sales Volume per customer	1	√
Assumption of risk management tasks for customers	18	
Standardized solutions	15	
Location of offered services in the OSI model	11	
Sufficient financial stability	11	
Provider references	9	
Offerer of stand-alone or black-box systems	8	
Market adoption of offered products and services	5	
Partnership with manufacturers	5	
Frequency of occurrence on vulnerability lists	4	
System commission and integration competencies	4	
Provider focuses on core competence fields	3	
Regional activity	3	
State of the art tools	3	
Global activity	2	
High-profile provider	2	

Continued on next page

Table 13.6 – continued from previous page		
In-house IT know-how	2	
Low cost strategy	2	
Offerer of brand-new products	2	
Proactive description of error scenarios and risks	2	
Reputation in the market	2	
Service provider business model	2	
Strategy alignment between customer and provider	2	
Established provider	1	
Individual arrangement of rules with customers	1	
Products and services require specialized know-how on the customer side	1	
Provider is market leader	1	
Provider is technological leader	1	
Provider uses cyber-insurance	1	
Regular reporting to the customer	1	
Service controllability	1	

Table 13.7: Exposure indicators for third party losses (Information Use)

Indicator	**Count**	**Selected**
Control and Coordination of outsourced customer processes by the outsourcing provider	3	√
Access to central systems and information of customers	1	√
Stability of outsourced processes	2	
Capacity management	1	
Established case-law in business sector	1	
Local distance to the customer	1	
Provision of dedicated resources for customers	1	

Table 13.8: Exposure indicators for third party losses (Information Systems)

Indicator	**Count**	**Selected**
Adoption of standardized methods and practices in software engineering	8	√
Quality and standardization of project management	5	√
Continued on next page		

Table 13.8 – continued from previous page		
Offering of training concepts and courses for customers	3	√
Existence of Update- and Version-management of the offered information systems	2	√
Extensive test procedures during development and deployment	2	√
Outsourcing of Information Systems Development (Offshoring)	1	√
Quality of Patch-Management for the offered Information Systems	1	√
Solutions corresponding to customer requirements	3	
Offered information systems equipped with security features	1	
Professional competencies for offered industry solutions	1	
Software architecture	1	
Supported platforms	1	

Table 13.9: Exposure indicators for third party losses (IT Infrastructure)

Indicator	**Count**	**Selected**
Customers have the possibility to switch to alternative providers	14	√
Remote maintenance of systems installed on customers sites	4	√
Offer of Backup Services (Data backup)	2	√
Few providers of critical core services on the market	1	
High market maturity	1	
Infrastructure of business location	1	
Products are subject to certification obligation	1	
Products used in adverse physical environments	1	
Provider has a depot of spare parts and components	1	

13.9.3 Indicators for the quality of IT risk management

Table 13.10: Indicators for the quality of IT risk management

Indicator	Count	Selected
Existence of an institutionalized risk management in the organization	36	√
Business Continuity concept and contingency and emergency plans available	24	√
Existence of a proper IT Risk Reporting	23	√
International standards and best practices orientation (in general and with regard to IT)	19	√
Investments in further education and training of employees to increase security awareness	19	√
Information Security Certification(s)	17	√
Existing Role of a Risk Officer or Security Officer	15	√
IT-Management reports directly to the board level or is part of the board of directors	15	√
Security Policies for employees	14	√
IT-Service-management approach based on Standards (e.g. ITIL)	12	√
Policies concerning the handling of confidential information	12	√
Continual improvement process in Risk Management (PDCA Cycle)	10	√
Maintenance of a proper Loss database (Incident reporting)	10	√
Redundancies in the technical infrastructure	10	√
Documentation of IT-Infrastructure	6	√
Periodical internal and/or external audits	6	√
Accounted budget for IT Security present	4	√
Acceptance Testing required prior to release of new technologies and product versions	3	√
Existence of an IT-Governance function in the organization	3	√
Existence of Protection requirements analysis	3	√
Physical control and registration of visitors at the entrance area	3	√
Physical protection measures	3	√
Quality Management System(s) Certification(s)	3	√
Size of the enterprise	3	√
Systematic Problem and Solution Management in the IT area	3	√
Age of the organization	2	√
Employee Background Check	2	√

Continued on next page

Table 13.10 – continued from previous page		
Measurement of performance indicators and and operating figures for assessing IT processes	2	√
Password Policy available	2	√
Policies and Guidelines for and control of external service providers	2	√
Definition of Life cycles for the IT-Infrastructure	1	√
Quality Assurance of published content (Legal Review)	1	√
Ratio of IT employees dedicated to security	1	√
Security Manual available	1	√
High degree of organization	23	
Contracts contain liability exclusions or limits	15	
Comprehensive decision making for selecting product and service providers	6	
Provider subject to legal form obligations	6	
Business oriented management of IT risks	5	
Tidiness and cleanliness of IT premises	5	
Existing logical and physical security architecture	4	
Presence at risk forums and interest groups	4	
State of the art of technical security controls	4	
High security and quality requirements of customers	3	
Internal control system	3	
Asset-Management	2	
Corporate Governance Guideline	2	
Presence of risk provisions	2	
Pursuance and external communication of innovative IT projects	2	
Safeguards for organizational security	2	
Adequate backup facilities	1	
Availability of controlling instruments	1	
Clear specification of service level agreements with providers	1	
Contractually guaranteed alternatives in case of failure	1	
Crisis public relation	1	
Defined corporate communications interfaces	1	
Employee suggestion system	1	
License-Management	1	
Methodical approach to IT investment appraisal	1	
Portfolio-management of IT projects	1	

References

1. AAA (American Academy of Actuaries Committee – Committee on Risk Classification): Risk Classification Statement of Principles (2008)
2. AICPCU (American Institute for CPCU/Insurance Institute of America): Foundations of Risk Management, Insurance, and Professionalism (Course Leader Handbook) CPCU 510 Appendix A (2006)
3. Baer, W.S.: Rewarding IT security in the marketplace. In: TPRC. (2003)
4. Betterley, R.S.: Cyberrisk Market Survey 2008 (June 2008) The Betterley Report.
5. Böhme, R.: Cyber-insurance revisited. In: Proceedings of the 4th Workshop on the Economics of Information Security (WEIS). Cambridge, MA (2005)
6. Böhme, R., Nowey, T.: 15 economic security metrics. In: Eusgeld, I., Freiling, F., Reussner, R. (eds.) Dependability Metrics, *LNCS*, vol. 4909, pp. 176–187. Springer, Berlin Heidelberg (2008)
7. Bouska, A.S.: In: Proceedings of the Casualty Actuarial Society Casualty Actuarial Society **LXXVI, Part 1**(145), 1–23 (1989)
8. BSI (British Standards Institution): BS 7799-3:2006 Information security management systems – Part 3: Guidelines for information security risk management (2006)
9. B̈uchel, M., Favre, R., Wiest, R.: Law, insurance and the Internet: the new perils of cyberspace. Technical report, Swiss Re Publishing (2000)
10. Cashell, B., Jackson, W., Jickling, M., Webel, B.: The economic impact of cyber-attacks. Congressional Research Service Documents, CRS RL32331 (2004)
11. Cummings, J.: S&P rolls out ERM review (2008). `http://businessfinancemag.com/article/sp-rolls-out-erm-review-0513`
12. Daley, B.: Using concept maps in qualitative research. In: Concept Maps: Theory, Methodology, Technology: Proceedings of the First International Conference on Concept Mapping, pp. 191–197. (2004)
13. Deloitte Touche Tohmatsu: Protecting what matters: The 6th annual global security survey (2009)
14. Ernst & Young: Moving beyond compliance: Ernst & Young's 2008 global information security survey (2008)
15. Finger, R.: Risk classification, chapter 6. In: Foundations of Casualty Actuarial Science, pp. 287–342. Casualty Actuarial Society (2001)
16. Gordon, L.A., Loeb, M.P., Sohail, T.: A framework for using insurance for cyber-risk management. Communications of the ACM **46**(3), 81–85 (2003)
17. Herath, H., Herath, T.: Cyber-insurance: copula pricing framework and implications for risk management. In: Proceedings of the 6th Workshop on the Economics of Information Security (WEIS). Pittsburgh, PA (2007)
18. Imriyas, K., Pheng, L.S., Teo, E.A.L.: A framework for computing workers' compensation insurance premiums in construction. Construction Management and Economics **25**(6), 563–584 (2007)
19. Innerhofer-Oberperfler, F., Breu, R.: An empirically derived loss taxonomy based on publicly known security incidents. In: Proceedings of the Fourth International Conference on Availability, Reliability and Security. Fukuoka, Japan (2009)
20. ISO (International Organization for Standardization): ISO/IEC 13335-1:2004 Information technology – Security techniques – Management of information and communications technology security – Part 1: Concepts and models for information and communications technology security management (2004)
21. ISO (International Organization for Standardization): ISO/IEC 73:2002 Risk management – Vocabulary – Guidelines for use in standards (2002)
22. Jackson, K., Trochim, W.: Concept mapping as an alternative approach for the analysis of open-ended survey responses. Organizational Research Methods **5**(4), 307 (2002)
23. Jiang, J., Klein, G., Ellis, T.: A measure of software development risk. Project Management Journal **33**(3), 20–41 (2002)
24. Kesan, J.P., Majuca, R.P., Yurcik, W.J.: Cyberinsurance as a market-based solution to the problem of cybersecurity. In: Proceedings of the 4th Workshop on the Economics of Information Security (WEIS). Cambridge, MA (2005)

25. Kotulic, A.G., Clark, J.G.: Why there aren't more information security research studies. Information & Management **41**(5) (2004) 597–607
26. Kovacs, P., Markham, M., Sweeting, R.: Cyber-incident risk in Canada and the role of insurance. ICLR Research Paper Series 38, ICLR (Institute for Catastrophic Loss Reduction) (2004)
27. Krcmar, H.: Informationsmanagement, 4., überarb. und erw. Aufl. Springer (2005)
28. Mattiacci, G.D.: The economics of pure economic loss and the internalisation of multiple externalities. In: Pure Economic Loss, vol. 9 of Tort and Insurance Law, 167–190. Springer, New York (2004)
29. Mukhopadhyay, A., Chatterjee, S., Saha, D., Mahanti, A., Sadhukhan, S.K.: e-Risk management with insurance: a framework using copula aided Bayesian belief networks. In: HICSS. IEEE Computer Society (2006)
30. Myers, M., Newman, M.: The qualitative interview in IS research: Examining the craft. Information and Organization **17**(1), 2–26 (2007)
31. Novak, J.D., Cañas, A.J.: The theory underlying concept maps and how to construct them. Technical Report Technical Report IHMC CmapTools 2006-01, Florida Institute for Human and Machine Cognition (2006)
32. Official Journal of the European Communities: Council Directive 2004/113/EC of 13 December 2004 implementing the principle of equal treatment between men and women in the access to and supply of goods and services (2004)
33. Ogut, H., Raghunathan, S., Menon, N.: Information security risk management through self-protection and insurance (2005)
34. Power, M.: The invention of operational risk. Review of International Political Economy **12**(4), 577–599 (2005)
35. Schmidt, R., Lyytinen, K., Keil, M., Cule, P.: Identifying software project risks: an international delphi study. Journal of Management Information Systems **17**(4), 5–36 (2001)
36. Schneier, B.: The insurance takeover. Information Security (2001)
37. Sherer, S., Alter, S.: Information system risks and risk factors: are they mostly about information systems? Communications of the Association for Information Systems **29**(64), 29 (2004)
38. Tipton, H., Krause, M.: Information Security Management Handbook. Auerbach Publishers (2007)
39. Trochim, W., Kane, M.: Concept mapping: an introduction to structured conceptualization in health care. International Journal for Quality in Health Care **17**(3), 187–191 (2005)
40. Trowbridge, C.: Fundamental concepts of actuarial science. Actuarial Education and Research Fund (1989)
41. Turban, E., Leidner, D., McLean, E., Wetherbe, J.: Information Technology for Management: Transforming Organizations in the Digital Economy. John Wiley & Sons (2008)
42. Wiegers, W.A.: The use of age, sex, and marital status as rating variables in automobile insurance. The University of Toronto Law Journal **39**(2), 149–210 (1989)
43. Wollnik, M.: Ein Referenzmodell des Informationsmanagements. Information Management **3**(3), 34–43 (1988)
44. Yurcik, W., Doss, D.: CyberInsurance: a market solution to the Internet security market failure. In: Proceedings of the 1st Workshop on the Economics of Information Security (WEIS). Berkeley, CA (2002)
45. Zimmermann, H.: OSI reference model – the ISO model of architecture for open systems interconnection. IEEE Transactions on Communications **28**(4), 425–432 (1980)

Chapter 14
The Risk of Risk Analysis And its Relation to the Economics of Insider Threats

Christian W. Probst and Jeffrey Hunker

Abstract Insider threats to organizational information security are widely viewed as an important concern, but little is understood as to the pattern of their occurrence. We outline an argument for explaining what originally surprised us: that many practitioners report that their organizations take basic steps to prevent insider attacks, but do not attempt to address more serious attacks. We suggest that an understanding of the true cost of additional policies to control insider threats, and the dynamic nature of potential insider threats together help explain why this observed behavior is economically rational. This conclusion also suggests that further work needs to be done to understand how better to change underlying motivations of insiders, rather than simply focus on controlling and monitoring their behavior.

14.1 Introduction

The *insider threat* or *insider problem* has received considerable attention, and is cited as the most serious security problem in many studies.[1] It is also considered the most difficult problem to deal with, because an insider has information and capabilities not known to other, external attackers. Examples for insider threats are manifold (*e.g.*, [2, 8, 10]), but usually only those resulting in significant harm are noticed by the public.[2]

Christian W. Probst
Technical University of Denmark, e-mail: probst@imm.dtu.dk

Jeffrey Hunker
Jeffrey Hunker Associates, e-mail: hunker@jeffreyhunker.com

[1] For example, in a 2007 Computer Security Institute survey about computer crime and security, 59 percent of respondents perceived that they had experienced insider abuse of network resources [9].

[2] Since 1995 only 119 cases of insider threats prosecuted under US Federal law have been identified [11].

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_14,

For example, in January 2008 Societé Generale suffered a $7 billion equities derivative loss due to the activities of a trader who had moved from the back office of the bank to become an apprentice trader in the dealing room. As the newspaper Liberation noted,"the stars of finance must be very cross that a simple base trader has succeeded in sinking a bank. The fraud is terrible for the credibility of the bank in the equities derivative sector, a business in which Societé Generale has become a global leader" [13, 17]. Of course only vague details of the case were revealed, and we will probably never know the exact details of the case, but the little we know hints at insider actions being responsible for the considerable damage.

Or for example Christina Binney, a senior employee of a small company, Banner Therapy, who without violating a specific company policy took home for the weekend the company's hard drive. She was subsequently fired for this action, the company claiming that her action put the company's very existence at jeopardy [5].

In a third highly public example, the US District of Columbia is pursuing a fraud case against a middle manager who used her influence to exclude her unit, dealing with real estate tax refunds, from a new Integrated Tax System. This exclusion allowed her to create bogus tax records that were not checked against actual real estate records [14].

In contrast to these high profile cases, lesser damages caused by insiders usually are covered up even if discovered. This goes in line with reports by professionals in organizations concerned about insider threats: their organization is aware of insider threats but takes only limited steps to prevent them, including threats posing the most serious impact. After the event, however, it is often considered crucial to have sufficient proof and documentation to be able to deal with these cases [20].

In the light of such severe consequences one should expect that preventing these threats would be one of the topmost priority for organizations. However, as many senior managers state, their organization is aware of the threat, but does little to prevent it.

We find this observation to be surprising, to say the least, and in case it is true, which recent events like the ones mentioned above indicate, the question is why organizations choose to be so vulnerable? The answer would be simple if the vulnerability were a matter of sloppiness by the organization. However, it seams that what we are talking about reflects what is presented by senior managers as a distinct choice.

In this way insider threats fundamentally differ from external threats. Organizations rarely choose to leave open vulnerabilities in their systems that might be exploited by outsiders to destroy or significantly damage the organization. If organizations do leave open such vulnerabilities, the reason is either limited resources (in which case one needs to examine the substance of the organization's risk analysis), or sloppiness.

In this paper we discuss the question why organizations, given the importance of insider threats, choose policies that allow insider threats to occur even in the face of adequate resources? Is this decision based on the sense that organizations (or their security personnel) figuratively throw up their hands in the face of a threat that, while recognized, seems impossible to adequately address? Little public data exist to help

answer the question of whether such behavior is economically or organizationally rational.

We develop an answer to this question by examining the relation between an organization's risk analysis, the assessment of trust in an insider, and how both of them (should) develop over time. We argue that as insiders over time gain more knowledge and thereby become a bigger risk, the organization only has two choices how to react. Either, the organization chooses implicitly or explicitly to have more trust in insiders as they pose a potentially bigger risk, or the organization needs to apply and enforce an ever-increasing number of policies to regulate the insider's actions.

In the rest of this paper we lay out a series of observations (based on anecdotes and extensive consultations with both researchers and practitioners), from which we derive a framework for understanding this observed behavior, and its implications for strategies for dealing with insider threats. We develop a combined view of the economics of the different components in this framework—the organization, the insider, and elements of mitigation all have a combination of goal function, risk function, and/or cost function associated with them. This obviously results in a multi-dimensional optimization problem, whose complexity eventually explains that our standard tool for assessing threats, risk analysis, breaks down in the face of one of the most vicious threats.

Our main conclusion will be that "complex" insider threats emerge as insiders with malicious intentions adapt their behavior to circumvent control systems. They often succeed because they have intimate knowledge of the control system, and especially of its blind spots. This adaptation of behavior makes it almost impossible to detect and prevent insider threats, and leads to a high uncertainty about possible threats, and in turn renders preemptive actions prohibitively costly. This benefit/cost ratio ultimately is the reason for organizations to refrain from defending this kind of insider threats. We discuss some possible measures how to prevent these complex threats from occurring. For a discussion of risk and uncertainty see Knight's seminal work [16].

While most of this paper considers insider threats, many of our results are applicable just as well in any risk/threat scenario, which involves trust. For a discussion of models for explaining insider threats see, for example, [21, 22].

14.2 Insiders, Outsiders, and Their Threats

While there is no commonly accepted definition of either an "insider" or an "insider threat" recent work [12, 20] points to a trust-based definition of an insider:

> *"An insider is a person that has been legitimately empowered with the right to access, represent or decide about one or more assets of the organization's structure."*

The rationale behind this definition is that it removes any specific IT bias from the definition, and focuses on organizational assets rather than a narrow approach

based on system credentials. The insider has been legitimately empowered to do some things that affect the organization, and he is trusted to use this empowerment wisely in a way that will benefit the organization, or at least not harm it. Beyond this definition [20] identifies factors of a "good" insider:

- Knowledge, intent, motivation
- Possessing the power to act as agent for the business
- Knowledge of underlying business IT platforms
- Knowledge/control over IT security controls
- Ability to incur liability, in pecuniary terms or in brand damage or other intangible terms.

All of these are affected both by time and position within the organization.

As mentioned in the introduction, insiders obviously have a special role for an organization. While an organization in general will try to do whatever possible to prevent threats from the outside, it often can or will not do so with threats on the inside. In the next section we present a series of observations, clarifying the relation between trust and risk, and their role for internal threats.

In contrast to insiders, outsiders usually are easily identified, as is the amount of access they should have to an organization's data and assets. The clear separation of concerns between outsiders and an organization eases controlling interactions with outsiders by means of access control and policies. It should be noted that above definition of insiders elegantly solves the problem of outsiders having special rights on an organization's assets—since they have been granted access, they are correctly treated as insiders.

Before further investigating the role of an insider in an organization, we first define what we mean by "insider threats". Insider threats emanate from individuals who are insiders according to our definition, and whose actions place the organization at risk. These actions can be maliciously motivated, the result of accident or error, or made because the individual is deceived. The insider threat can be caused by an insider acting alone, or in concert with other insiders, outsiders, or various combinations of the two.

Thus, insider threats encompass a wide variety of different types of actions that can have a correspondingly wide range of impacts on the organization. While work on developing complete taxonomies of different insider threats is underway [6, 12], a simple categorization, sufficient for our purposes, is to differentiate by motive and complexity of trust relationship:

- For *motivation* we distinguish between accidental and intentional actions; and
- for *complexity of trust relationship* we distinguish between simple and complex ones.

Based on this categorization, we consider the following scenarios of insider threats.

14.2.1 Insider Threats That Do Not Represent a Violation of Trust

- *Accidents or stupidity*: People will be stupid and we cannot anticipate stupidity or accidents very easily. There is considerable work on ways of anticipating and preventing such instances, and much of it draws on work in fields (like nuclear plant operation for example) where the lessons are nonetheless applicable to the insider threat issue [18].
- *Fulfillment of duty*: Organization's policies tend to get in the way of performing a task. Insiders may decide to disobey a policy, and thereby on the one hand be able to fulfill their duty, on the other possibly causing an insider threat, which they might or might not be aware of. With Binney and Banner Therapy, Binney apparently was unaware that she was potentially threatening the organization's survival. Considering the trust-based definition of insiders given above, they are trusted to judge whether or not the situation justifies breaking the rules [1, 24].

14.2.2 Insider Threats That Do Represent a Violation of Trust

- **"Simple" insider threat**: The typical example for this is the disgruntled employee, who might be at risk of being fired and causes damage to the system, or steals some files they have access to; or a person who is paid to steal data. Most of these are cases where the system facilitates the damage, *i.e.*, the same damage could have been caused in a pen and paper system, or the threat involves violation of trust that is not easily picked up on, *e.g.*, an employee reading printouts in a printer room they have access to, or the recently fired employee who still has (through administrative oversight) access to the organization's computers. The key property of these cases is that the damage done to the organization while potentially considerable, also reflects a violation of a simple trust relationship.

"Simple" insider threats depending on violations of trust can be thought of as follows: the losses caused are not too high, and can therefore justifiably be ignored; or the potential harm is considerable but the threat depends on trust relationships being violated in a simple fashion, meaning that they could easily have been prevented. In either of these cases the organizational response is appropriate—we either absorb the cost as part of doing business, or revise our security policies so as to avoid a repeat of the insider threat again.

- **High profile (or charismatic) insider threat**: This is the type of insider threat that usually is reported on in the press—the one everyone is fascinated by. Examples include the aforementioned French trader [13, 17], the D.C. real estate tax fraud [14], or the Danish case of Stein Bagger [23], who used his position to build up a complex system of fraud and deception.

These high-profile insider threats with devastating consequences can represent extremely clever schemes. What is more, the insiders causing them usually have

more information than the typical insider. As the head of the D.C. tax office commented after the real estate tax fraud was discovered, "Our system has got a plethora of internal controls on it. On top of that, we have manual controls. But you're always vulnerable to an enterprising employee who knows how the controls work." [14].

We hypothesize that these intentional malicious insider threats with large impacts on organizations occur more frequently than appears in the public eye, but the risk of their occurring is accepted by the organization as an unavoidable risk of doing business.

This hypothesis is based on anecdotal evidence from discussions with private sector and government managers, from the public record, and on the observation that the high level of interest in preventing insider threats by many financial institutions suggests that the problem is viewed as being very serious.

In the rest of the paper we argue that charismatic insider threats fundamentally challenge the basis for risk analysis. In its simplest form, risk analysis depends on:

- Policies[3] directed towards a risk (and their costs)
- Losses due to risks, and
- Probabilities of risks taking place.

We conclude that risk analysis focused on blocking or detecting high-level insiders from carrying out their threats is of only limited value. Alternative ways to increase the confidence that the trusted insider does not become a threat depend on human factors (basically keeping insiders happy); the effectiveness of these policies appears to be little understood in the insider-threat literature.

14.3 Building up Trust and Risk

Trust is a central ingredient of our private and public life, be it as a person or as an organization [7], whenever we have to consider a risk. In this section we discuss in detail the relation between risk, trust, organizations, and insiders. In doing so we will repeatedly get back to a mock-up insider story, which illustrates the process of an organization hiring a new employee, and how he thrives and prospers, turning into an insider and eventually representing a serious threat to the organization. In Figure 14.1 we plot the relation between time and the risk that the insider poses to the organization, and the trust relation between the organization and the insider. Before looking at trust and risk, however, we first lay out the beginning of the example.

Example: Organization X wants to hire a new employee. They interview a flock of applicants, eventually picking one.

At this point the organization has a basic understanding of their future employee, but they do not necessarily have a reason to trust him. This moment is marked by

[3] We define policies to mean the set of technical, organizational, and behavioral actions or rules that an organization has created to prevent, control or encourage actions that affect their information systems. Of course, not all policies are necessarily followed in practice [19], a point we will discuss further.

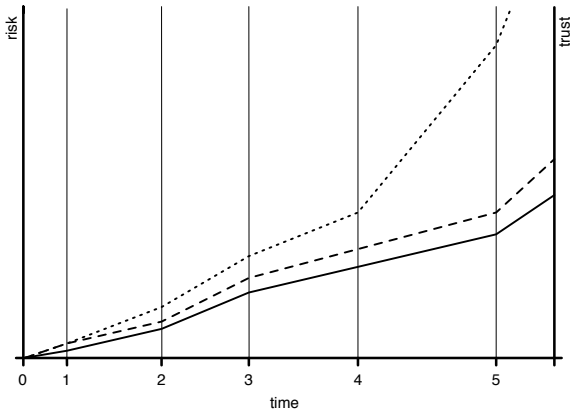


Fig. 14.1 Plot of the **trust** that an organization has in an insider (solid line) against the risk that the insider poses to the organization. The dashed line represents the organization's **acceptable risk**, using tools such as policies and auditing to minimize the distance between risk and trust. The dotted line represents the **effective risk** that the insider emanates. The marks on the time line represent events during the insider's employment in the organization (see text for discussion). Note that the effective risk could very well be smaller than the acceptable risk.

point "0" on the time line in Figure 14.1—the new employee is not hired yet, the company has neither trust into him, nor does he constitute a risk.

In order to increase, establish, or justify their initial trust, and assess the potential risk the future employee might pose, they will usually (or at least should) run a background check. Independent of whether or not a background check is performed, once the applicant is hired, the organization on the one hand establishes a simple trust relation to him, on the other hand he poses a certain risk. Both risk and threat correlate to the position in the organization he starts at, as well as the assets and data he get access to. This is identified by point "1" on the time line in Figure 14.1.

For the sake of this section we assume that the insider to be is hired at a rather low-level entrance level. Whatever we describe in the following could just as well occur when joining the organization at a senior level, which in our classification of insider threats, would represent a complex trust relationship.

14.3.1 Simple Trust, Low Risk

We now are at point "1" on the time line in Figure 14.1. The company has established a simple trust relation to the insider, but as just mentioned the new employee also emanates a certain risk for the organization, part of which might be acceptable.

How does an organization deal with this situation? To mitigate the risk, and to justify the trust, only simple mechanisms are needed. Based on the established trust, the insider can be granted access to certain parts of the organization's assets. However, the insider poses a (small) risk to the organization, and should therefore not be

able to freely act in the organization. A usual mechanism is to control the insider's access to the organization's assets by means of security clearance, and access rights to certain data and locations. This establishes with help of simple means an easy to control limitation of the risk that the insider can pose.

In this phase the insider's knowledge of the organization and its assets is fairly limited, and so is the amount of damage he can cause. Over time, this knowledge will increase, and will result in the need to adjust the risk analysis. At the same time, the organization and the employee develop a hopefully mutual, more complex trust relationship, which to a certain degree justifies accepting more risks.

14.3.2 Medium Trust, Elevated Risk

Example: After some time the insider changes positions and joins the internal auditing unit, where he works as part of a team that audits the organization's transactions.

This obviously represents a substantial increase in trust into the employee, and it also means that the employee now represents a significantly higher risk for the organization, since he gets access to potentially secret data of internal transactions.

On the time line in Figure 14.1 we are now at point "2"—the trust in the employee has increased, as has the risk that he poses. When considering the complexity of the trust and the risk relation, it has increased considerably, too. This increase is due to the insider's more detailed knowledge about the organization, both with respect to inner workings and with respect to internal data. As before the organization may want to limit the difference between risk and trust, by means of a combination of policies, monitoring, and auditing.

The overall situation stays the same as before—the organization has some trust in the employee, and is willing to accept a certain risk beyond that. As before the organization may want to limit this risk as well as the potential additional risk posed by the employee, and in this case a typical solution is a set of policies that among others might result in two or more members of the auditing unit being required to access the auditing data, thus spreading the risk over several employees. In contrast to the previous situation, the mix of mitigating factors now is getting more diverse, and potentially more restrictive.

14.3.3 Complex Trust, Even More Complex Risk

Example: After having worked in the auditing department for some time, the insider has been promoted again (point "3"), and we meet him some time later, as he joins the trading unit (point "4"), having already established himself in the organization.

At this point the organization has built up a fairly high amount of trust into the employee. Due to potentially diverse positions the insider has worked in, and consequently due to potentially manifold knowledge the insider has on internal workings

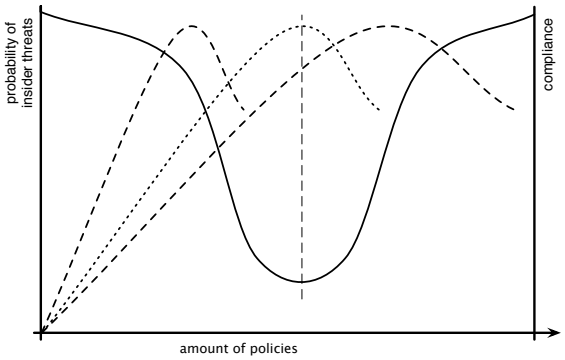


Fig. 14.2 Plot of the number of policies against the likelihood of insider attacks to occurring (solid line), and the likelihood that employees will comply with the policies (dashed/dotted lines) [3, 4]. Organizations want to be at the "sweet spot", where maximum compliance coincides with minimum number of policies (the dotted plot). For compliance, the x axis could also be interpreted as "time passed since a certain policy was introduced", assuming that it takes some time to establish the policy's efficiency, which will eventually degrade again.

and assets, the trust relationship now is fairly complex. The interplay of different areas of the organization that the insider has experienced is hard to clearly describe, and even harder to measure.

As a consequence of the trust relationship getting more complex, the risk assessment of the insider will rise in lockstep, as before. However, Figure 14.1 illustrates that we assume the effective risk to grow significantly larger than the acceptable risk. This is motivated exactly by the fact that the insider has developed a more precise model and knowledge of the organization, its inner workings, and assets.

Example: While the insider might no longer have direct access to the auditing system, he still knows the details of how the system works and when it is triggered.

For the organization this can have dramatic consequences. From a "local" viewpoint, whatever policies are applied for employees in the trading unit should work just fine for the insider, since they are tuned to cover exactly the transactions and behavior that is expected from a member of this unit.

From a more "global" viewpoint, this mitigation of course is completely inadequate, since it does not take into account previous knowledge of the employee. While this problem might be resolvable for transfers inside of the organization, imagine the effort necessary to identify, assess, and mitigate the risk when hiring somebody from outside into the trading unit.

Simple trust relationships are relatively straightforward in the ability to control or monitor the risk in our interactions; more complex trust relationships on the other hand pose difficult problems in terms of how to ensure or monitor some degree of trust. While this kind of trust relationships pervade our whole existence, it largely depends on situational factors how much we rely on them in making decisions.

In any kind of relationships we therefore face a number of problems related to trust and risk. First of all we need to establish trust in another actor. Based on this

trust, we may be able to accept a certain risk when interacting with this actor (the dashed line in Figure 14.1). However, since we are not able to completely validate our assessment, there always exists the possibility that the actor poses a (significantly) larger risk than what we can accept (dotted line in Figure 14.1).

Combining our conclusions, we summarize that:

- the compliance of insiders to policies for control and monitoring will peak and then decline—at exactly which point depends on organizational factors that require more research;
- as policies for control and monitoring increase, as expected the probability of insider threats falls;
- at some crucial inflection point, however, two events occur: first, compliance with "too many" policies starts to fall, while insiders continue to gain knowledge that makes them potential high-level insider threats. Thus the combination of these two factors (which need not be simultaneous) means that the risk of insider threats starts to increase again. Furthermore, since the high-level insider is more fully knowledgeable about the organization, their potential for damage as an insider threat is high.

A note seems in place regarding Figure 14.1. We implicitly assume that the factors considered, knowledge and authentication, both evolve over time. One might argue that for many actors in an organization the risk does not increase over time, or the trust/risk relationship does not become more complex. However, even though an employee "only" gets to know the system better, he also understands better how to perform actions that he wants to not to be observed, or where to leave "markers" to document that he did something [22].

14.4 Policies and Compliance

We can think of policies in two basic forms:

- those that control or monitor behavior to attempt to enforce the trust relationship (*e.g.*, through access control or monitoring of behavior); and
- those that motivate insiders to "act in the appropriate way" — in other words to act in a way that ensures that they do not become insider threats.

In this section we will consider the impact and economics only of the first sort – those that seek to control behavior. As trust relationships grow more complex we observe distinct differences in the economics and effectiveness of these sorts of policies.

To account for the difference between trust, acceptable risk, and potential risk as described in the previous section, we use policies to control the admissible actions, and the accessible assets. The goal of these mitigating factors clearly is to minimize the likelihood of a big differential between acceptable and actual risk or threat.

All restrictive policies seek to control or monitor behavior. The costs of these policies, especially the hidden costs of policies interfering with the normal work

flow of the organization, can be high. This cost, however real, may be difficult to measure. Gaps and conflicts in policies can create confusion among insiders in terms of "what is right" or "how do I get my job done?" While ideally security should support people in doing their jobs, several examples are known of technological security approaches that, because they interfered with the work flow, were not accepted and in fact actively subverted (e.g. an iris reader with an "unacceptable" delay before allowing access resulted in staff finding other ways of gaining access; motion detectors designed to automatically log off users were disabled by covering them with plastic cups). Compliance with security policies is hard. Making compliance easy for insiders is absolutely necessary for any successful effort to constrain insider threats. Yet none of these instances lend themselves to clear-cut cost measurements, but intuitively they cost the organization if not in money then in factors like staff time or motivation.

14.4.1 Enforcing Simple Trust Relationships

Control of simple trust relationships lends itself to access control and monitoring policies with commonly acceptable cost/benefit ratios. Typical questions faced when enforcing simple trust relationships are

- Who should have access to what information?
- Under what circumstances, and how defined?

We would posit that, although restrictive policies have organizational costs, some of these measures appear to have acceptable cost/benefit ratios. Basic access control measures (passwords or tokens, required and automatic cryptographic use, selective file access) and monitoring (to a point) appear beneficial in preventing or discouraging a large set of insider threat activities that could create a potentially large loss to the organization. The deciding factor in all of these cases is how much monitoring and access control is acceptable (both ethically and legally) and at what point does it stop being beneficial, compared with the costs (both monetary and otherwise) to the organization.

It also appears that it is commonly understood that there is a "reasonable" probability that these measures will prevent certain common types of insider threats. None of this is supported, to our knowledge, by anything other than anecdotal evidence.

The impact of these policies is aggregative up to a point—in other words, certain sets of policies work together to create a greater benefit compared to cost than they would individually. For instance, passwords together with physical limitations on data copying (blocking certain ports, for example) together with selective monitoring together may provide much greater benefit than that provided by each policy separately. Part of the reason for this, simply, is that policies controlling simple trust relationships oftentimes affect a large number of insiders (passwords may be required for all insiders, for example), and that to a point combinations of these policies reinforce each other.

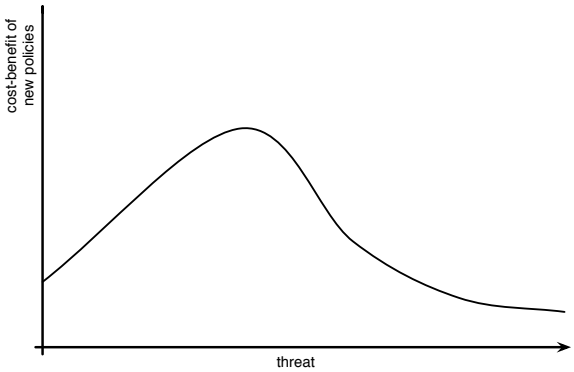


Fig. 14.3 Plot of cost-benefit ratio of new policies against an increasing threat.

The marginal effectiveness of each additional policy declines, all other things being equal. Thus, up to a point we conclude that the probability adjusted benefit-cost value for restrictive policies aimed at insider threats is positive, and may even be increasing, up to a point. In Figure 14.3 we illustrate this argument. Up to a point we are able to predict that new policies will benefit the organization, based on a reasonable risk analysis—our actions to reduce insider threats do more good than harm to the organization. Beyond that point, however, the added policies harm the organization, either because employees do not comply, or because they disturb the work flow too much [3, 4].

An important caveat is worth repeating—none of the factors going into this evaluation have, to our knowledge, any sound basis in data; nonetheless our description above captures (albeit in somewhat different language) a set of sentiments commonly expressed by practitioners dealing with insider threats.

14.4.2 Managing Complex Trust-Risk Relationship

Policies for controlling complex trust relationships face a number of challenges not faced to the same extent when controlling simple trust relationships.

The questions in terms of controlling complex trust relationships include those mentioned for simple relationships, plus:

- When does complex behavior signal that an insider threat is taking place, as opposed to, say, creative activity?
- The effectiveness of a particular policy is unclear—are we putting in place policies that deal with potential threats that never will materialize?

Just like simple relationships, policies for controlling complex trust relationships face a number of challenges.

The aforementioned cost of policies interfering with the natural work flow of the organization increases, we posit, for large complex systems with equivalently complex trust relationships. In such cases prevention and detection require a significant effort. Not only may expanded monitoring (e.g., anecdotally many professionals object to the notion that their use of the computer is being monitored) affect trust within an organization, it becomes increasingly nuanced in what to look for in complex trust relationships. For example an inordinate amount of system searching may indicate that an unauthorized person has access to that account (a masquerader)—or a forgetful mind. The cost of false positives may be significantly higher for senior managers than for data entry clerks—or even IT administrators.

Solutions may themselves be complex, and have limited applicability across the organization. For example a set of actions to ensure that senior executives do not steal vital information in order to create their own company or move on to a competitor requires, at a minimum, heightened monitoring of system activity. But if the data is commonly used, and commonly used by the staff of senior executives, then the problem of actually detecting data theft might become immensely intrusive both to the ability of the staff and senior executives to do their work, and to morale and other human factors. A threat which may be of immense impact if it happens, but of totally unknown likelihood, and affecting only a very small number of insider s directly, probably only has high cost solutions to preventing it – if it has any at all.

Attempting to control complex trust relationships increases the risk that those actions will severely damage the organization.

As noted above, restrictive policies (monitoring, access control) all carry the risk of increasing the cost to the organization by interfering with people's ability to do their job. We find it intuitive that attempting to control complex trust relationships carries with it an especially high cost—in fact one that may not be acceptable to the organization.

Thus, organizations attempting to manage the risk of high-level insider threats face a number of special challenges:

- The probability of a high-level insider threat event is difficult (impossible?) to predict or even imagine in advance.
- The longer an individual is in the organization (or some other descriptor that captures this notion of increasing trust), the greater their knowledge of the valuable information assets or services and how to circumvent the policies in place.
- So a trusted person is also in the position to do the most damage to the organization.
- The cost of more information security policies is poorly understood, but in general the anticipated cost is if anything less than the real cost (in other words, a well meaning set of policies runs the risk of damaging the organization severely and in unanticipated ways, but it is unlikely that the real cost is far less than what was anticipated).

The difficulties in dealing with insider threats are increased by the complexity of organizational and insider threat goals, which we now discuss.

14.4.3 Simple vs. Complex

The boundary between simple and complex insider threats is blurry. By "simple" insider threats we mean those that are obvious, like the linear dependencies in [18]; while they might potentially cause severe damage, they can easily be identified and monitored. This might for example be the confidential document where every insider with access rights might pose a threat. When considering policies this would typically involve actors, roles and assets that are mentioned explicitly in policy rules.

Complex insider threats, on the other hand, got their name from Perrow's complex dependencies [18]. Here it is often unclear how they built up over time as a combination of different factors discussed above. These threats may develop "under cover", and eventually be triggered by apparently unrelated events, which exactly makes them so hard to predict.

14.5 Organizational and Insider Goals

Goals shape what is important to both the organization and the insider; goals also shape what options are chosen both by the organization and by the insider.

14.5.1 Organizations

Organizations have many, potentially conflicting goals that also influence how they choose to deal with insider threats. Most important they of course try to maximize their gain function, most often in the form of maximizing the organization's profit. This is supported by trying to minimize the risk of both outside and inside attacks. Factors in reaching these goals are trying to ensure (maximize) compliance with the organization's policies as described in Section 14.4, to try to maximize the employee's loyalty with the organization, and to find the right number of policies.

Poorly articulated and conflicting goals make it more difficult to determine both what is of value to the organization, and what trust relationships in the organization are most critical.

One key question is whether organizations who have suffered insider threats now act differently than they did in the past. And, what they are prepared to pay to avoid another occurrence? In other words, do organizations "learn" over time or by experience so as to forge clearer links between their goals and the most important values and trusts? Anecdotally, past insider threats seem to raise awareness of the threat, but it is unclear whether this also leads to more effective measures. To preview our conclusions, for insider threats that violate highly complex trust relationships the specific threat may be strictly unique.

Organizations that have faced high-level insider threats before probably do act differently. However, the only truly effective responses are not controls—how can

the next high-level insider threat be anticipated? The effective responses are to first help build the organizational culture where insiders do not want to become threats, and second, to consider ways in which damage can be mitigated after the fact.

14.5.2 Insiders

Insiders have complex, poorly articulated goals, too—they want to, e.g., maximize the damage to the company/CEO/... or their personal gain, at the same time trying to minimize the risk of being detected. Just like organizations, insiders often have muddled goals, and organizations cannot completely predict the many forms that insider threats might take. If, as we argue, the high-level insider also has a strong incentive to be creative in their threat, then predicting in advance the form of the charismatic threat becomes even more difficult—indeed, we might conclude, almost impossible.

14.6 The Risk of Risk Analysis

In particular more complex trust relationships pose a set of difficult questions when performing risk analysis.

As noted above, we observe that complex trust relationships generally are associated with more complex behaviors. Thus, understanding the nature of the threat itself in any actionable way, the potential losses accruing, and the probability of such instances happening (even if they can be imagined beforehand) are all difficult.

Major, complex insider threats appear to be rare, and largely unique in their construction and execution. Of course, successfully executed, their impact on the organization can be very large, and they should therefore be accounted for in the risk analysis. There does, however, not seem to be an adequate way of systematically deciding that "this potential complex threat is more likely that that threats", nor any generally accepted perception across the community such as exists for less complex insider threats.

Since a priori it is difficult to predict the form that a high-level insider threat will take, organizations cannot adequately anticipate beforehand the possibly high costs that insider threats could have to their systems and enterprises. We hear frequently from corporate managers that they did not appreciate the value of what was lost through the insider threat until after the event. More formally, with poorly articulated goals making it difficult at best to estimate the value of organizational resources, and possibly highly complex insider threats affecting many different organizational resources, of course organizations have great difficulty in anticipating the costs of some insider threats.

Estimating losses from high-level threats is also challenging, since it frequently does not show up until some time after the event started. A currency speculator may

appear for years to be a major profit center for his banking group—until one day he is discovered to be an insider threat. Or, the loss is virtual until it hits. It could even be that maybe the risk is constant, but due to the risky behavior going on for some time, the disastrous effect is getting bigger and bigger.

As described above, at the same time risk builds up in the background. Thus, complex trust relationships develop over time. In some cases this simply may be due to greater familiarity over time with the workings of the information system or in their daily work; for example over time an employee may learn or be able to guess the passwords of fellow workers. In other cases the trust relationship that extends over time is more complex. The important observation, however, is that over time more complex trust relationships grow between insiders and others in the organization.

It is not a problem, until high-level or charismatic insiders go bad and use that knowledge to maximize their goal. We think it therefore is crucial for mitigation to make this risk explicit in an organization's risk assessment. But just like insiders often are able to do harm because they know the system and can play it, the same holds if they are aware of what the risk function looks like.

Thus, in parallel the consequences of violating those trust relationships can become more costly to the organization. As a gross generalization, certainly not always true, insiders have the potential to cause more damage to an organization the longer they have been an insider, simply as a function of the greater trust relationships that may have been established.

14.6.1 Plotting the Value Function

As stated above, we consider two different situations; either the organization can anticipate a type of threat, or it can not even imagine it.

For the first case it seems that the value function of the organization will be convex—in other words up to some point we can anticipate the most common (or imagine that we can anticipate...) types of threats; the policies put in place are not too costly; we perceive their effectiveness as being high; and we believe that the probability of these types of threats to be high enough to worry about.

Thus, up to a point the value function of the organization looks as described before (Figure 14.3).

This assumes almost perfect information—we can anticipate a certain type of threat, though we do not know who will emanate the threat; we can estimate its probability of taking place; we know the cost of putting in place policies to address this threat; we know how effective these policies will be, *i.e.*, the probability that they will prevent or detect an insider threat; we know what the cost to the organization will be.

It seems logical that in this case there will be some very serious (but not totally absolutely catastrophic) insider threats for which the cost will exceed the benefit of putting in place the required policies adjusted for their likelihood of being success-

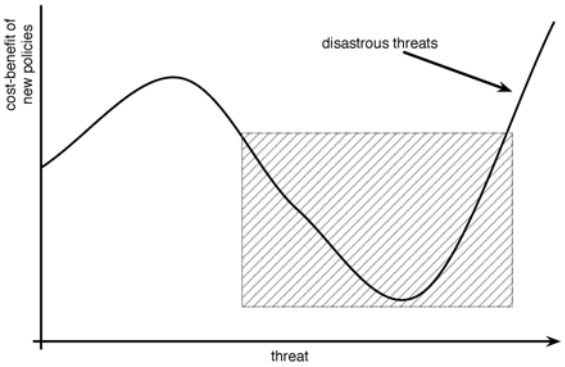


Fig. 14.4 Extension of the cost-benefit ratio plot against threat. Once the threat reaches disastrous levels, it may very well be beneficial to instantiate policies trying to prevent these events. Organizations face the problem that in a certain area they are unable to predict the effect of policies on optimizing their overall gain function—this is exactly the area where senior managers throw their hands in the air and choose to ignore the threat.

ful. This complex nature of threats causes a decrease in the cost-benefit ratio as illustrated in Figure 14.4, weighting cost of policies against their probability weighted value, since the added policies have a negative effect on compliance, or work flow, or a combination thereof.

As the impact of threats increases, there can occur some absolutely disastrous threats, whose outcome is absolutely unacceptable for the organization. For these, the probability-weighted value is positive again, as it is inevitable that these threats are mitigated—in other words, it is appropriate to take action.

But in a more "real" circumstance we observe that once the organization gets to a certain point—beyond the "handful" of normal actions that we would take (whatever those are; access control, monitoring, periodic background checks)—what we have entered is totally unknown territory even if we can anticipate the nature of the inside threat. In this area we do not know how our employees react to more policies, how our gain function evolves, how the risk of attacks evolves, and so on. This unknown territory is marked by the box in Figure 14.4.

This unknown territory is defined by:

- At some point we start to get on shaky ground in terms of estimating the true cost to the organization of the policies— *i.e.*, insiders get irritated with the working environment, or we start to increase the risk that in some sort of unexpected situation needed data is not accessible;
- Additional policies may also increase the likelihood of false positives, or also the cost of false positives goes up— *e.g.*, as we monitor senior executives; and
- As we move further out on the trust curve, we get less confident that the threat we are trying to solve is real (or is it just imaginary?). But the cost of the additional policies is real.

Thus the organization's real value function looks like that in Figure 14.4, with the boxed area replaced by a huge question mark. This is exactly the area where senior managers state that they throw the hands up in the air, being aware of the threats, but also being aware of not knowing how their organization will behave. And, while they are at it, they often ignore the high-risk threats as well, taking them into account, because, *e.g.*, the risk may be high, but so is the gain. Besides the Binney case, all examples mentioned above fall into this category. Kerviel was earning his bank huge amounts of money before going bad, so it might have been convenient to ignore the risk, and in the case of the tax fraud, the insider's suggestion not to implement a certain auditing system was followed since the budget had already been overspent.

14.6.2 The Benefit of Obscurity

It should be noted that a risk analysis itself, once performed, poses a significant risk to the organization; this is especially true if we consider higher management as potential insiders. Since they certainly have a complex risk/trust relationship to their organization, it seems at least mandated to do so.

Once a detailed risk analysis has been performed, it may be hard to keep secret, *especially* from upper management. Ironically, the very risk analysis that is performed to identify and *limit* the effect of insider threats (or threats in general), does actually *increase* their potential effect if the result gets in the wrong hands. The same information, being confidential, is much less harmful in relation to outsiders, and the threat they pose will therefore not increase.

We argue therefore that for the result of a detailed risk analysis Kerckhoffs' principles [15] should *not* be applied, since its content can cause disastrous damage and should therefor be accessible only to a very limited group of actors. This, however, may lead to a circular dependency, since the risk analysis may be needed to identify who should be allowed to access its results.

It should be noted that this approach of "security by obscurity" might also seem advisable for selected other documents, which could be described as the spinal cord of a company. However, it might be infeasible to identify who can or cannot be trusted to access these documents. Eventually one has to trust actors to behave well.

14.7 Strategies to Change Motivation Rather than Prevent Bad Insider Actions

This points to organizations behaving economically rationally for all but high-level threats by picking a small number of insider threats that can be managed, and dealing with the rest through mitigation after the fact. There may be some threats posing such a great risk to the organization that the cost to the organization of the nec-

essary policies may be justifiable. However, most high-level threats are, by nature, unpredictable.

For high-level insider threats, two other types of policies may be most useful:

- Mitigation of the impact of the insider threat. Are there ways of increasing the successfulness of mitigation? Are there types of insider threats for which mitigation is just not going to be an acceptable path? We suspect that the potential damage from a high-level insider threat may be too great to think of mitigation as a relief (for example, the case of Aldrich Ames, the insider who spied on behalf of enemies of the United States, does not appear to lend itself to mitigation).
- However, investment in the other sorts of policies—changing behavior so that people trust their organization and do not *want* to cause harm—makes the most sense. Even though these sorts of "positive" policies are even less well understood in terms of their effectiveness/impact than the technically based "control" measures which we show break down at a certain point, anecdotes suggest that friendly, supportive, organizational cultures, where insiders do not have the incentive to become a threat, are possible to construct. Even difficult situations, like a large number of firings, can be done in a way that preserves a positive atmosphere.

14.8 Conclusion

We conclude, therefore, based on this logic, that risk analysis for insider threats is useful up to a point, but that the whole risk analysis approach as a means of selecting what actions to take breaks down as we get into the territory of dealing with highly complex trust relationships—insiders who are highly knowledgeable about the information, its value, and the protections in place. We can imagine all sorts of threats, but do not know which ones to take seriously. Maybe too as we get into highly specialized threats the types of policies we would take to counter each threat become less universal, and more specialized.

We see the net effect of risk analysis breaking down in all sorts of organizations. This article began by noting that organizations act as though they tolerate some serious insider activity—in other words, that in addressing the insider threat there is an even worse perceived risk of severely damaging the organization.

We also observe organizations figuratively throw up their hands in the face of a threat that, while recognized, seems impossible to adequately address. Consider for example a complex organization like a hospital. Even *defining* a trust relationships strikes us as being very difficult, time consuming, and prone to errors. Having defined (somehow) the trust relationships at risk of an insider threat, the organization is still faced with the task of developing policies to counter the threat. Is it any wonder then that some organizations throw up their hands in the face of this challenge?

14.8.1 Probability of Policies Being Successful in Blocking High-Level Insider Threats

To further our conclusion, we note that all of this is that policies have a probability of being successful (that they actually work). So the expected loss function is the *probability of an insider threat to occur*, times the *probable damage of a certain amount or type*, times the *probability that the policies imposed will be unsuccessful* in blocking that threat. We believe that for more complex trust relationship based insider threats the very effectiveness of the policies deployed to counter the threat may be less effective. This goes in line with observations that organizations with increased surveillance and auditing often state that the number of detected cases stays constant, as was recently reported by several public agency and private company officials [20].

So for high-level insider threats it is very expensive to put in place all of the policies to block these threats, with increasingly low probability that the policies will actually be successful (because the more policies you add the less successful cumulatively they will become). The loss function is very high at one end, with low probability throughout, but when they do occur it's a big loss.

To summarize: our chief tool for assessing threats (risk analysis) and for deciding what threats to deal with, and how, breaks down for what might be the worst sorts of threats. This finally explains why organizations behave as they do, and that, even though surprising, their behavior is economically rational even in the face of high-level threats—by picking a small number of insider threats that can be managed, and dealing with the rest through mitigation after the fact (even if mitigation is not likely to be very successful). For high-level threats it may be that in a few cases (where the event can be anticipated in advance, and the costs to the organization are very high) the organizational cost and disruption of imposing control policies may be worth it. But this probably describes the exception rather than the rule.

The appropriate insider threat control strategy depends on an organization's perceived loss function from insider threats. Different organizations presumably have differently shaped loss functions: US intelligence organizations probably have a big bump at the far right, making them very sensitive to high-level insider threat. Banks are probably like intelligence organization, though the evidence is mixed on this.

We conclude as well that it becomes economically rational at some point in the threat function to invest heavily in policies to change behavior in a positive fashion, even if these policies are not well understood in terms of their impact or effectiveness.

References

1. Adams, A., Sasse, M.A.: Users are not the enemy. Commun. ACM **42**(12), 40–46 (1999). DOI http://doi.acm.org/10.1145/322796.322806

2. Anderson, R.H.: Research and Development Initiatives Focused on Preventing, Detecting, and Responding to Insider Misuse of Critical Defense Information Systems: Results of a Three-Day Workshop. RAND Corporation, Santa Monica, CA, U.S.A. (1999)
3. Beautement, A., Coles, R., Griffin, J., Monahan, B., Pym, D., Sasse, M., Wonham, M.: Modelling the human and technological costs and benefits of usb memory stick security. In: Proceedings of the Workshop on Economics in Information Security (2008)
4. Beautement, A., Sasse, M., Wonham, M.: The compliance budget: Managing security behaviour in organisations. In: New Security Paradigms Workshop (2008)
5. Binney v. Banner Therapy Products, 631 S.E. 2d 848, 850. North Carolina Court of Appeals (2006)
6. Bishop, M., Engle, S., Peisert, S., Whalen, T., Gates, C.: Case studies of an insider framework. In: Proceedings of the 42nd Hawaii International Conference on System Sciences (HICSS) (2009)
7. Cofta, P.: Trust, Complexity and Control: Confidence in a Convergent World. John Wiley and Sons (2007)
8. Cole, E., Ring, S.: Insider Threat: Protecting the Enterprise from Sabotage, Spying, and Theft. Elsevier (2006)
9. Computer Crime and Security Survey. Computer Security Institute (2007)
10. Contos, B.T.: Enemy at the Water Cooler. Elsevier (2007)
11. Hunker, J., Bulford, C.: Federal Prosecution of Insider Threats Demonstrates Need for Reform; Analysis based on data base of Federal prosecutions since 1995. Manuscript under review (2009)
12. Hunker, J., Predd, J., Pfleeger, S.L., Bulford, C.: Insiders behaving badly: A taxonomy of bad actors and their actions. Manuscript under review (2008)
13. Jérôme Kerviel. Available from `http://en.wikipedia.org/wiki/Jerome\_Kerviel`, last visited February 27, 2009
14. Keating, D.: Tax suspects guidance on software left d.c. at risk. Washington Post (2008)
15. Kerckhoffs, A.: La cryptographie militaire. Journal des sciences militaires **IX** (1883)
16. Knight, F.H.: Risk, Uncertainty, and Profit. Hart, Schaffner & Marx; Houghton Mifflin Co. (1921). Library of Economics and Liberty [Online] available from http://www.econlib.org/library/Knight/knRUP.html; accessed 24 May 2009.
17. Michelson, M.: Bank scandal a blow to french pride. In *International Herald Tribune* (2008)
18. Perrow, C.: Normal Accidents: Living with High-risk Technologies. Princeton University Press (1999)
19. Predd, J., Pfleeger, S.L., Hunker, J., Bulford, C.: Insiders behaving badly. IEEE Security and Privacy **6**(4), 66–70 (2008). DOI http://doi.ieeecomputersociety.org/10.1109/MSP.2008.87
20. Probst, C.W., Hunker, J., Bishop, M., Gollmann, D.: Countering insider threats. Dagstuhl Seminar Proceedings (2008). URL `http://drops.dagstuhl.de/opus/volltexte/2008/1793`
21. Schudel, G., Wood, B.: Modeling behavior of the cyber-terrorist
22. Schultz, E.E.: A framework for understanding and predicting insider attacks. In: Proceedings of CompSec (2002)
23. Stein Bagger. Available from `http://en.wikipedia.org/wiki/Stein\_Bagger`, last visited February 27, 2009
24. Weirich, D., Sasse, M.A.: Pretty good persuasion: a first step towards effective password security in the real world. In: NSPW '01: Proceedings of the 2001 workshop on New security paradigms, pp. 137–143. ACM, New York, NY, USA (2001). DOI http://doi.acm.org/10.1145/508171.508195

Chapter 15
Competition, Speculative Risks, and IT Security Outsourcing

Asunur Cezar, Huseyin Cavusoglu and Srinivasan Raghunathan

Abstract Information security management is becoming a more critical and, simultaneously, a challenging function for many firms. Even though many security managers are skeptical about outsourcing of IT security, others have cited reasons that are used for outsourcing of traditional IT functions for why security outsourcing is likely to increase. Our research offers a novel explanation, based on competitive externalities associated with IT security, for firms' decisions to outsource IT security. We show that if competitive externalities are ignored, then a firm will outsource security if and only if the MSSP offers a quality (or a cost) advantage over in-house operations, which is consistent with the traditional explanation for security outsourcing. However, a higher quality is neither a prerequisite nor a guarantee for a firm to outsource security. The competitive risk environment and the nature of the security function outsourced, in addition to quality, determine firms' outsourcing decisions. If the reward from the competitor's breach is higher than the loss from own breach, then even if the likelihood of a breach is higher under the MSSP the expected benefit from the competitive demand externality may offset the loss from the higher likelihood of breaches, resulting in one or both firms outsourcing security. The incentive to outsource security monitoring is higher than that of infrastructure management because the MSSP can reduce the likelihood of breach on both firms and thus enhance the demand externality effect. The incentive to outsource security monitoring (infrastructure management) is higher (lower) if either the likelihood of breach on both firms is lower (higher) when security is outsourced or the benefit (relative to loss) from the externality is higher (lower). The benefit from the demand

Asunur Cezar
Middle East Technical University, Ankara, Turkey, e-mail: asunur@metu.edu.tr

Huseyin Cavusoglu
School of Management, The University of Texas at Dallas, Richardson, TX 75083, e-mail: huseyin@utdallas.edu

Srinivasan Raghunathan
School of Management, The University of Texas at Dallas, Richardson, TX 75083, e-mail: sraghu@utdallas.edu

T. Moore et al. (eds.), *Economics of Information Security and Privacy,*
DOI 10.1007/978-1-4419-6967-5_15,

externality arising out of a security breach is higher when more of the customers that leave the breached firm switch to the non-breached firm.

15.1 Introduction

Information security management is emerging as a critical business function, partly because of firms' increasing reliance on the Internet to conduct business and increasing regulatory requirements. Simultaneously, information security management is becoming more complex and challenging. Some of the reasons for this include changes in attack patterns over time (increased frequency, severity and sophistication of attacks); complex information technology (IT) environments consisting of multitudes of hardware, operating systems, application software, and distributed networks, each with its own vulnerabilities; shortage of security professionals with the required expertise; diverse security solutions from vendors; limited IT budgets; and demanding audit and regulatory requirements (e.g., Sarbanes Oxley (SOX), California Senate Bill No. 1386, Gramm-Leach-Bliley Act (GLBA), Health Insurance Portability and Accounting Act (HIPPA), Payment Card Industry Data Security Standard (PCI DSS), Basel II, among others). Outsourcing to Managed Security Service Providers (MSSP) has emerged as one of the key strategies to deal with the complexities of IT security management. The MSSP industry is relatively new, but analysts project significant growth in the MSSP industry. According to IDC, the value of U.S. managed security services market was approximately $1.3 billion in 2007, an increase of 19.6% over 2006; this figure is expected to reach $2.8 billion by 2012 [21]. Yankee Group [29] estimated that the global spending on managed security services was approximately $4 billion in 2006. They projected managed security services market to grow at a compound rate of 14 percent from 2006 through 2010. Frost and Sullivan [15] projects that managed security services will exceed $6 billion by 2011. According to Gartner, in 2006, 60% of Fortune 500 enterprises had used an MSSP, and about 20% of enterprise firewalls were under remote monitoring or management [23]. The range of services outsourced includes perimeter protection which includes managed services for firewalls, IDSs, VPNs, and other security infrastructure management, security event monitoring, incident management including emergency response and forensic analysis, and security consulting that includes vulnerability assessment, penetration testing, network architecture review, and compliance gap analysis.

Even though many security managers are skeptical about outsourcing of IT security [28], [14], mainly due to the fear of losing control over sensitive information, industry analysts have cited cost savings, better protection, leveraging of expertise, economies of scale, compliance with laws, and liability transfer as the primary drivers for the outsourcing of information security functions [42], [10], [33]. Schneier [34] noted that information security is part of IT infrastructure and "infrastructure is always outsourced". These reasons for IT security outsourcing suggest that practitioners and industry experts do not view IT security as different from tra-

ditional IT functions such as systems development, maintenance, help desk support, and data center operations, which are routinely outsourced. However, while IT security and traditional IT functions share many common risks, some of these risks are more significant in the IT security context. For instance, when the security of a firm is breached, the firm not only incurs losses related to recovery from the breach and from possible business disruptions, but may also lose customers to non-breached competitors if the breach is publicly revealed. Thus, the non-breached firm may stand to gain from the breached competitor. The recovery and business disruption-related costs are found in traditional IT failures. However, the cost (benefit) associated with loss (gain) of customers is not significant in traditional IT failures [1]. The risks involving traditional IT failures are in general non-speculative, which are those exogenous events from which only a loss can occur. However, because of competitive externalities, IT security may involve speculative risks, which are events from which either a profit or a loss can occur [39].

One reason that IT security environment exhibits speculative risks is the competition between firms induced by security-sensitive customers that may switch from a firm that does not protect their information to another firm that does. The relative magnitudes of non-speculative and speculative components of information security breach risk are evident from the results of a recent Ponemon Institute study [30]. The study found that the average total cost of a data breach, which included direct cost (such as free or discounted services offered, notification letters, phone calls and emails, and legal and auditing fees), lost productivity costs (such as the lost employee or contractor time diverted from other tasks to security breach related tasks), and customer opportunity costs which cover turnover of existing customers and increased difficulty in acquiring new customers was $197 for each breached customer record in 2007, an increase of 8% and 43% since 2006 and 2005, respectively. In the financial services industry, the cost per breached record was even higher at $239. The cost of lost business (due to customer churn) averaged 65% of the total cost (versus 54% in 2006) or $128 per breached record, and this figure increased at more than 30 percent, averaging $128 per breached record. The customer churn rate averaged 2.67% in 2007, an increase from 2.01% in 2006. A second study by Ponemon Institute [31] reported that customers have not become accustomed to new data breaches, but, on the contrary, they are increasingly prone to terminate their business relationships due to security breaches. Further, results from 2007 CSI Survey [32] suggest that the burden of security breaches are often transferred to customers, thus enhancing customers' incentives to switch to another firm that is not breached. The empirical data related to risks associated with security breaches clearly suggest that analyzing IT security outsourcing decisions solely on the basis of non-competitive and non-speculative factors is incomplete.

Using a simple game theoretical model of two firms deciding to either perform in-house security management or outsource their IT security functions, we show that if competitive externalities are ignored, then a firm will outsource security if

[1] The literature on traditional IT outsourcing does not suggest this risk as one of the reasons for firms' decisions to outsource, implying that this risk is not a significant factor in traditional IT failures.

and only if the MSSP offers a quality (or a cost) advantage over in-house operations, which is consistent with the traditional explanation for security outsourcing. However, a higher quality is neither a prerequisite nor a guarantee for a firm to outsource security. The competitive risk environment and the nature of the security function outsourced, in addition to quality, determine firms' outsourcing decisions. If the reward from the competitor's breach is higher than the loss from own breach, then even if the likelihood of a breach is higher under the MSSP the expected benefit from the competitive demand externality may offset the loss from the higher likelihood of breaches, resulting in one or both firms outsourcing security. The incentive to outsource security monitoring is higher than that of infrastructure management because the MSSP can reduce the likelihood of breach on both firms and thus enhance the demand externality effect. The incentive to outsource security monitoring (infrastructure management) is higher (lower) if either the likelihood of breach on both firms is lower (higher) when security is outsourced or the benefit (relative to loss) from the externality is higher (lower). The benefit from the demand externality arising out of a security breach is higher when more of the customers that leave the breached firm switch to the non-breached firm.

The rest of the chapter is organized as follows. In the next section, we review the vast research on general IT outsourcing and the limited research on IT security outsourcing. In section 3, we describe the model. In section 4, we present our analysis and discuss firms' sourcing decisions. Finally, we discuss the implications of our results and provide directions for further research in section 5.

15.2 Literature Review

The literature on outsourcing traditional IT functions is extensive, but the literature specifically on IT security outsourcing is limited. Dibbern et al. [9] provides a comprehensive review of the IT outsourcing literature. Prior work has utilized transactional cost theory, agency theory, core-competency argument, and vendor-client relationship management to understand and explain why firms outsource IT, the benefits and risks associated with IT outsourcing, the IT functions outsourced, and factors that affect IT outsourcing outcomes. The bulk of this work relied on data collected through surveys.

Early research focused on cost savings as the primary motivation for outsourcing. Loh and Venkatraman [24] found that the degree of IT outsourcing was positively related to business and IT cost structures and negatively related to IT performance. On the other hand, McLellan et al. [27] did not find any evidence for the hypothesis that firms with weak financial performance were more likely to outsource. Ang and Straub [2] and Sobol and Apte [37] concluded that firm size was negatively associated with the degree of outsourcing. Caldwell [4] reported that one third of outsourcing contracts targeted at cost reductions failed to match the expectations.

Another stream of research explored the diffusion of IT outsourcing. Loh and Venkatraman [25] investigated whether the source of the influence of diffusion was

internal (imitative behavior), external (external channels of communication such as media, etc.) or mixed (both). They concluded that the internal influence model explained the diffusion of IT outsourcing deals better than other models and that the internal influence was stronger after the Kodak's well-publicized outsourcing announcement. Reexamining the study by Loh and Venkatraman [25] with expanded data set, Hu et al. [20] found that mixed influence was the dominant factor, and did not find support for the Kodak effect. Ang and Cummings [1] found that when the source of influence was federal regulators, banks responded more to institutional demands and less to strategic economic contingencies, and when the source of influence was peers, banks responded more to strategic economic contingencies. Slaughter and Ang [36] found that firms were more likely to outsource jobs having volatile demand and requiring scarce skills. They explained their results using economies of scale and mitigation of technological risk arguments. DiRomualdo and Gurbaxani [13] found that three strategic intents of IS outsourcing – IS improvement, business impact and commercial exploitation – impacted the degree of outsourcing and type of sourcing relationship.

Although IT security outsourcing is a widely discussed topic among the practitioner community, academic literature in IT security outsourcing is limited. Rowe [33] suggested firms may enjoy benefits from network effects when they outsource IT security to the same MSSP. He argued that when more firms outsource to the same MSSP, the MSSP will be able to provide a better service to all customers because of access to a larger set of data and being able to analyze more network configurations. However, the MSSP could also become a more valuable target to attackers, increasing the likelihood of attack.

Very few papers in the information systems literature have developed economic models to understand either traditional IT or IT security outsourcing. Whang [41] analyzed a multi-period software development contract between a firm and an outside developer and derived an optimal contact which replicates the equilibrium outcome of a benchmark in-house development. More recently, Sen at al. [35] analyzed the impact of demand heterogeneity and variance in user preferences on the pricing and the allocation of resources for service-oriented models of information technology. Dey et al. [8] analyzed different types of software outsourcing contracts under information asymmetry and incentive divergence and showed that by improvements on outsourcing process and control mechanisms, contract performance could be improved. In the IT security context, Ding et al. [10] examined the characteristics of optimal MSSP contracts under moral hazard and reputation effects and found that an optimal contract should be performance based even in the existence of a strong reputation effect. In a subsequent work, Ding et al. [12] showed that outsourcing decision is relatively insensitive to variation in service quality but highly sensitive to bankruptcy risk. Ding et al. [11] showed that when transaction cost uncertainty or transaction costs are high, MSSPs are forced to charge a lower price to balance these costs. Gupta and Zhdanov [18] analyzed the growth of MSSP network under a for-profit MSSP monopoly and under a consortium-based market structure.

The other literature on outsourcing has been in the manufacturing/production area [6] and has primarily focused on principal-agent models to identify the con-

ditions under which firms prefer outsourcing, the type of job that will be outsourced [38], and investment levels [40].

The economic models considered in prior work on outsourcing typically relied on a principal-agent model with a single principal (the firm) and a single agent (the MSSP). However, our model incorporates the competition between two firms. Therefore, we are able to identify how competitive externalities influence firms' IT security outsourcing decisions.

15.3 Model Description

We consider an industry that has two competing firms, labeled as firm 1 and firm 2. Each firm offers a single product or service. The demand for the product of a firm is affected by whether one or both firms suffer from a security breach, in addition to its and the competing product's prices. The likelihood of a security breach on one or both firms depends on whether they manage their security in-house or they outsource their security. The specific assumptions of our model along with their justifications follow:

Assumption 1: The demand for the product of firm i is given by the following.

$$q_i = a - b_1 p_i + b_2 p_j + B_i \;\; i, j \in \{1,2\},\; i \neq j \tag{15.1}$$

where a, b_1, $b_2 > 0$ and $b_2 < b_1$. The linear competitive demand model given by (15.1) is standard in the literature [26], [16]. In (15.1), a represents the base demand to a firm when both firms set prices to zero and there is no security breach, b_1 denotes a firm's own price effect, b_2 denotes the cross-price effect, and B_i captures the change in firm i's demand when firm i, firm j, or both i and j are breached.

Assumption 2:

$$B_i = \begin{cases} -\Delta,\ B_j = \alpha\Delta & \text{if } i \text{ is breached and } j \text{ is not breached} \\ -\Delta,\ B_j = -\Delta & \text{if both } i \text{and } j \text{ are breached} \\ 0,\ B_j = 0, & \text{if neither firm is breached} \end{cases} \tag{15.2}$$

If firm i is breached and firm j is not, firm i's demand decreases by Δ and firm j gets a fraction $\alpha \leq 1$ of Δ, and therefore, firm j's demand increases by $\alpha\Delta$, and the industry's demand decreases by $(1-\alpha)\Delta$. Parameter α can be interpreted as a measure of the degree of spillover of demand to the non-breached competitor. The degree of spillover is likely to be dependent on factors such as the type (essential vs. non-essential) of product or service provided by the firms, substitutability of the products or services, and switching costs. For example, a publicized breach event in banking, health or pharmaceutical industry may cause more switching than a breach event in the manufacturing industry. When both firms are breached, each firm's demand decreases by Δ, resulting in a total decrease of 2Δ for the industry. The value of Δ is likely to be affected by factors related to the nature of breach,

such as the sensitivity of customer information compromised in the breach as well as the product type. We assume that the decrease in demand due to a breach can not be larger than the primary demand each firm faces, i.e., $\Delta \leq \alpha$.

Assumption 3: Firms can manage security through in-house operations or by outsourcing it to a MSSP. There is a single MSSP. While the assumption of single MSSP is not critical to our analysis [2], consolidation trends in MSSP industry and comments by security experts suggest that MSSP industry is likely to have few large players [22], [7], [3].

Assumption 4: The joint probability distribution for the breach events at the two firms when firm 1 decides X and firm 2 decides Y, where $X, Y \in \{$*outsource(O)*, *in-house(I)*$\}$, is given by the following probability matrix.

Table 15.1 Joint probability distribution of breach events.

		Firm 2	
		Breached	Non-breached
Firm 1	Breached	P^{XY}	$\theta^X - P^{XY}$
	Non-breached	$\theta^Y - P^{XY}$	$1 - \theta^X - \theta^Y + P^{XY}$

The marginal probability of a security breach for a firm when it outsources and when it manages in-house is θ^O and θ^I, respectively. A lower marginal probability implies a higher level of protection or a higher quality of security services. The quality of security services is likely to depend on the technology and expertise used by the firm managing the security services. We denote the environment in which $\theta^O < \theta^I$ as the *High Quality Outsourcing* environment, and that in which $\theta^O > \theta^I$ as the *Low Quality Outsourcing* environment.

The probability that both firms are breached is P^{XY}. A higher value for P^{XY} implies that the breach events in the two firms are more correlated. Whether the degree of correlation will be higher when both firms outsource than when one or both firms do not outsource will depend critically on the function outsourced. If the MSSP specializes in the management of security infrastructure that includes firewall, IDS, and other security technologies, then the MSSP is likely to use same or similar technologies and expertise to manage the security of both firms in order to take advantage of economies of scale. In this case, the correlation between breach events in two firms is likely to be higher when both firms outsource than when one or both do not, i.e., $P^{OO} > P^{OI}, P^{II}$. If the MSSP specializes in monitoring services, then it is likely to focus on observing and analyzing the breach event on firms and use information pertaining to breach on one firm and protect the other firm from a similar breach, if it is not already breached. That is, MSSP facilitates information-sharing relationship between firms [33]. In this case, the joint probability of breach in two firms is likely to be lower when both firms outsource than either one or both do not, i.e., $P^{OO} < P^{OI}, P^{II}$. In the light of above arguments, we characterize an MSSP environment as belonging to one of the four regions given in Fig.15.1. The vertical

[2] We discuss the impact of relaxing the single MSSP assumption in Section 5.

axis denotes the difference in the quality of MSSP and that of in-house management, $\theta^O - \theta^I$. The horizontal axis denotes the difference in the joint probability of breach events in two firms when both firms outsource IT security and that when only one firm outsources IT security.

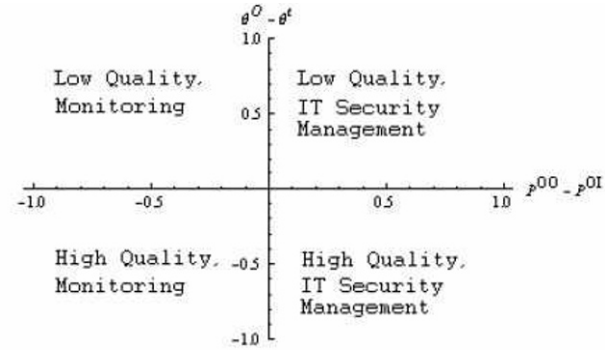


Fig. 15.1 MSSP operating regions.

Assumption 5: The firms have a fixed budget for managing information security, and they spend this budget on security whether they outsource or manage in-house. This assumption is consistent with current industry practices in information security. Further, we assume that the budget is the same and is normalized to zero in order to eliminate the impact of cost differences in firm's outsourcing decisions [3] and to focus on the impact of competition, industry, and breach characteristics.

Assumption 6: The marginal cost of production is fixed and is normalized to zero. This assumption does not affect the results qualitatively.

Assumption 7: We consider a one shot, 3-stage non-cooperative game. The sequence of events is the following. In stage 1, each firm decides whether to outsource security or manage it in-house. In stage 2, nature moves and the breach events occur. After observing breach events at stage 2, firms set their prices simultaneously at stage 3. The assumption that outsourcing decisions are made before price decisions indicates that outsourcing decisions are more strategic and long-term compared to price decisions for firms. This is consistent with the empirical observation that IT outsourcing contracts tend to be long-term [19]. The assumption also indicates that price decisions are flexible in the sense that prices can be changed relatively easily and frequently.

Assumption 8: All model parameters are common knowledge. This assumption allows us to analyze the strategic interaction between firms caused by competition, MSSP, and breach characteristics, which is the focus of this paper.

[3] Note that for the same budget, MSSP could offer a higher (or a lower) quality than in-house operations. So, θ^O and θ^I could be viewed as cost-adjusted quality measures.

15.4 Model Analysis

We use backward induction to solve for the Nash equilibrium for the sourcing game. At stage three, after observing the breach events (if any), both firms choose their prices simultaneously by maximizing their individual payoffs. The payoff for firm i, π_i, is given by the following.

$$\pi_i = p_i q_i = p_i(a - b_1 p_i + b_2 p_j + B_i) \tag{15.3}$$

Solving simultaneously the first-order conditions for the maximization problems of both firms, we obtain the following optimal price for firm i in stage 3 of the game. Details of this and other derivations as well as proofs of propositions in this paper are provided in the Appendix.

$$p_i^* = \frac{a(2b_1 + b_2) + 2b_1 B_i + b_2 B_j}{4b_1^2 - b_2^2} \tag{15.4}$$

The values for B_i and B_j depend on the breach scenario (viz., zero, one, or two breached firms) is realized in stage 2. Substituting (15.2) in (15.4), we obtain the following optimal prices in stage 3.

$$p_i^* = \begin{cases} \frac{a}{2b_1 - b_2} & \text{if neither firm is breached} \\ \frac{2b_1(a+\alpha\Delta) + b_2(a-\Delta)}{4b_1^2 - b_2^2} & \text{if firm } i \text{ is not breached and firm } j \text{ is breached} \\ \frac{2b_1(a-\Delta) + b_2(a+\alpha\Delta)}{4b_1^2 - b_2^2} & \text{if firm } i \text{ is breached and firm } j \text{ is not breached} \\ \frac{a-\Delta}{2b_1 - b_2} & \text{if both firms are breached} \end{cases} \tag{15.5}$$

We make the following observations regarding the optimal prices in stage 3. The price charged by the breached firm is lower than that charged by firms when there is no breach and that charged by the non-breached firm, but is higher than that charged when both firms are breached. These observations are intuitive and can be explained by the demand effects of the breach events. We also observe that the non-breached firm's price (when its competitor is breached) may be higher or lower than the price charged when there are no breaches; it is higher when $\alpha > \frac{b_2}{2b_1}$ and is lower otherwise. This shows that if the spill-over demand, relative to the degree of price competition, is not sufficiently large, then the non-breached firm is unable to take advantage of the increase in its demand and charge a higher price because, at high levels of price competition, the non-breached firm is forced to reduce its price in response to the lower price charged by the breached firm.

Substituting (15.5) in (15.3), we find that $\pi_i = b_1(p_i)^2$ under any breach scenario. Therefore, the breached firm always sees a reduction in its profit. However, the non-breached firm may see its profit increase or decrease depending on whether $\alpha > \frac{b_2}{2b_1}$. Breach has a direct effect and an indirect effect on the non-breached firm. The direct effect is that it enjoys a higher primary demand, ceteris paribus, because of the spillover of consumers from the breached firm. The indirect effect is that the changes in demands of the two firms force the firms to change their prices, which may or may not favor the non-breached firm. Depending on which effect dominates, a non-breached firm may be rewarded or penalized by a breach on the competitor.

In stage 1 of the game, each firm simultaneously makes its sourcing decision by maximizing its expected payoff in stage 3 of the game. The expected payoff for a firm depends on the outsourcing decisions of both firms. The expected payoffs for firm i and firm j in stage 1 are shown in Fig.15.2. The first (second) element in the ordered pair within each cell is the expected payoff to firm 1(firm 2). We define the following variables for ease of exposition.

$$L = \frac{\Delta b_1(2b_1 - \alpha b_2)((4a - 2\Delta)b_1 + (2a + \alpha\Delta)b_2)}{(4b_1^2 - b_2^2)^2} \tag{15.6}$$

$$V = \frac{\Delta b_1(2\alpha b_1 - b_2)(2(2a + \alpha\Delta)b_1 + (2a - \Delta)b_2)}{(4b_1^2 - b_2^2)^2} \tag{15.7}$$

$$L^b = \frac{\Delta b_1(2a - \Delta)}{(2b_1 - b_2)^2} \tag{15.8}$$

We can show that L and V, respectively denote the decrease in profit to the breached firm and the increase in profit to the non-breached firm when only one firm is breached, and L^b denotes the loss of profit to each firm when both firms are breached. Note that L and L^b are always positive, but V is positive when $\alpha > \frac{b_2}{2b_1}$ and negative when $\alpha < \frac{b_2}{2b_1}$.

Firm 1 \ Firm 2	O	I
O	$\begin{pmatrix} -P^{OO}L^b + (\theta^O - P^{OO})(V - L), \\ -P^{OO}L^b + (\theta^O - P^{OO})(V - L) \end{pmatrix}$	$\begin{pmatrix} -P^{OI}L^b + (\theta^I - P^{OI})V - (\theta^O - P^{OI})L, \\ -P^{OI}L^b + (\theta^O - P^{OI})V - (\theta^I - P^{OI})L \end{pmatrix}$
I	$\begin{pmatrix} -P^{OI}L^b + (\theta^O - P^{OI})V - (\theta^I - P^{OI})L, \\ -P^{OI}L^b + (\theta^I - P^{OI})V - (\theta^O - P^{OI})L \end{pmatrix}$	$\begin{pmatrix} -P^{II}L^b + (\theta^I - P^{II})(V - L), \\ -P^{II}L^b + (\theta^I - P^{II})(V - L) \end{pmatrix}$

Fig. 15.2 Normal Form of the game in Stage 1.

We define the variable $R = \frac{V + L^b}{L}$ which replaces all cost and benefit terms in the above figure and allows us to analyze the total impact of these terms using a single variable. The numerator denotes, given that the competitor is breached, how

much higher the firm's profit is if it is not breached than if it is breached. Similarly, the denominator shows the same profit difference given that the competitor is not breached. This is a ratio of the (net) value a firm obtains from being non-breached when the competitor is breached to that when the competitor is not breached, i.e., a measure of the relative value non-breached firm gets from the competitor's breach. Note that R can be less than 1 or greater than 1. R describes, in a restrictive sense, the competitive risk associated with security breaches. That is, it measures the relative benefit to loss a firm realizes if only one of the two competing firms is breached. Note that if both firms are breached or no firm is breached, neither firm has a competitive advantage over the other. If $R > 1$, then a firm realizes a positive expected payoff given that only one firm is breached. Following the terminology used in [39], we label this environment as "speculative risk" environment. We label the environment in which $R < 1$ as "non-speculative risk" environment, and in this case, a firm realizes a negative expected payoff given only one firm is breached.

The following result characterizes the Nash equilibrium outcome for the sourcing game.

Lemma 15.1. *The Nash equilibrium outcome for the sourcing game is given by the following:*

$$\begin{cases} \textit{(outsource, outsource), if } (\theta^O - \theta^I) < (P^{OI} - P^{OO})(R-1) \\ \textit{(in-house, in-house), if } (\theta^O - \theta^I) > max((P^{II} - P^{OI})(R-1), (P^{OI} - P^{OO})(R-1)) \\ \textit{Mixed strategy with probability of outsourcing } \frac{(P^{II}-P^{OI})(R-1)-(\theta^O-\theta^I)}{(2P^{OI}-P^{OO}-P^{II})(R-1)}\textit{, otherwise} \end{cases}$$

Lemma 1 shows that a firm's outsourcing decision depends critically on three factors: the quality of the MSSP relative to that of in-house security management, the security function (viz., security monitoring or infrastructure management) outsourced, and the ratio R. Fig.15.3 illustrates the regions where the different decisions are optimal for the firms for a speculative risk environment. In the region below line AB, both firms outsource. In the shaded region above line AB, both firms manage in house, and in the non-shaded region, each firm outsources with a probability as given in Lemma 1. It is evident from the figure that even when the MSSP does not offer a higher quality than in-house management, both firms may outsource (see the shaded region below line AB in quadrant II).

Because our interest is in deriving insights about how the risk environment, the MSSP, industry and breach characteristics affect the firms' outsourcing decisions, we next analyze each of these impacts separately.

15.4.1 Impact of Competitive Risk Environment on Firm's Outsourcing Decisions

We show the following result.

Proposition 15.1. *(i) If R=1, then both firms outsource iff the MSSP provides a higher quality than in-house management and both firms manage in-house otherwise.*

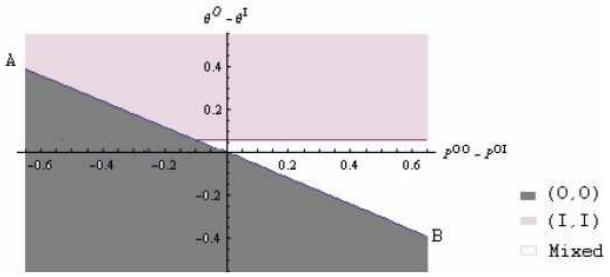


Fig. 15.3 Sourcing regions for $P^{OI} = 0.35, ^{II} = 0.45, R = 1.6$.

(ii) An increase in R leads to an increase in the likelihood of both firms outsourcing if security monitoring is outsourced and a decrease in the likelihood of both firms outsourcing if security infrastructure management is outsourced.

Proposition 15.1 provides important insight into the role played by the risk environment in firms' decisions to outsource security. The environment is speculation-risk-neutral (i.e., $R = 1$) when either the firms are not competitors or firms do not expect a net benefit from the competitor's breach. In this environment, firms will outsource security if and only if the MSSP offers a quality advantage over in-house management. It is worthwhile to note that security researchers and practitioners frequently cite quality or cost advantages of MSSPs as the primary reason for firms to outsource security ([42], pp. 199-200). While this conventional explanation is consistent with our result, it is only partial. Specifically, the explanation based solely on quality or cost advantage does not provide any insights into the role of strategic factors or the security function outsourced on the firms' decisions.

We find that when firms do compete with each other, based on either price or breach events, strategic considerations, in particular the nature and extent of risk, influence firms' decisions. When firms face speculative or non-speculative risk from security breaches, they may outsource even when the MSSP does not offer a quality advantage. For example, as it is seen in Fig.15.3, even when $\theta^O > \theta^I$, both firms outsource monitoring function in the shaded region below line AB in quadrant II. Furthermore, Proposition 15.1 shows that the security function also plays an important role in firms' outsourcing decisions. For instance, if the MSSP does not offer a higher overall quality than in-house management, then both firms will likely manage security infrastructure management in house.

Proposition 15.1(*ii*) shows the impact of extent of competitive risk on firms' decisions and is illustrated visually using Fig.15.4a and Fig.15.4b. An increase in R, shown by the clockwise movement of the line AB, increases the likelihood of outsourcing the security monitoring function by both firms and decreases the likelihood of outsourcing infrastructure management by both firms. R is higher when the expected payoff to a firm given that one is breached and the other is not breached is higher, which implies that a scenario in which one is breached and the other is not

becomes more profitable to a firm at higher values of R. Therefore, the outsourcing decision that increases the likelihood of this scenario becomes more attractive to firms. Because outsourcing security monitoring increases the likelihood of only one firm being breached, firms have more incentives to outsource security monitoring. Using the same logic, we can explain why firms are less likely to outsource infrastructure management when R increases. In essence, as the risk associated with the security environment becomes more speculative, firms are more (less) likely to outsource monitoring (infrastructure management).

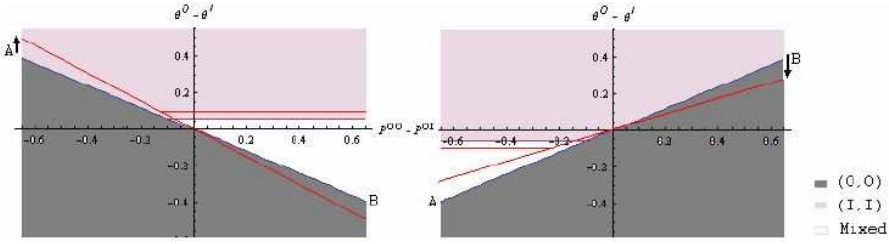


Fig. 15.4 Impact of reward-to-risk ratio for $P^{0I} = 0.35$, $P^{II} = 0.45$ a)$R = 1.6$ (left) b)$R = 0.4$ (right).

15.4.2 Impact of MSSP Characteristics on Firms' Outsourcing Decisions

In our model, the MSSP is characterized by two parameters: the marginal probability of breach for the outsourcing firm (θ^O), which measures the overall quality of service offered by the MSSP and the joint probability of breach events in the two firms (P^{OO}), which measures, in some sense, the relative degree of security monitoring vis-à-vis infrastructure management services offered by the MSSP. A higher value for P^{OO} often suggests that the MSSP focuses more on infrastructure management and less on security monitoring.
We show the following result.

Proposition 15.2. *(i) An improvement in the MSSP quality leads to more outsourcing by both firms irrespective of the security function outsourced.*
(ii) An increase in P^{OO} decreases the likelihood of both firms' outsourcing if $R > 1$, and increases the likelihood of both firms' outsourcing, otherwise.

Proposition 15.2(i) is intuitive because an improvement in MSSP quality improves a firm's payoff from outsourcing. Further, if the firms compete with each other, then an improvement in the MSSP quality hurts the payoff of the firm that manages security in-house. Therefore, both firms have a higher incentive to use the MSSP, and firms that manage security in-house may shift to the outsourcing strategy if the MSSP quality increases. In Fig.15.3, an improvement in the MSSP quality can

be shown as a downward movement on the vertical axis, which represents a movement towards the outsourcing region.

Fig.15.5a and Fig.15.5b illustrate Proposition 15.2(ii). An increase in P^{OO} is shown as a horizontal movement to the right, from the initial position at (x_1, y_1) to the final position at (x_2, y_1) after the increase in P^{OO}. In Fig.15.5a, in which $R > 1$, (x_1, y_1) lies in the region where both firms outsource and (x_2, y_1) lies in the region where both firms manage in-house. We find the opposite in Fig.15.5b, in which $R < 1$. The intuition underlying Proposition 15.2(ii) can be explained as follows. When $R > 1$, the positive expected benefit in the scenario in which only one firm is breached induces firms to prefer an environment in which only one of them is breached to that in which both firms are breached. Therefore, if the joint probability of breach increases, then firms' incentives to outsource decrease. When $R < 1$, the expected benefit in the scenario in which only one firm is breached is negative. So, firms prefer an environment in which both are breached to the environment in which only one of them is breached. Thus, in this case, an increase in the joint probability of breach events in two firms increases both firms' incentives to outsource.

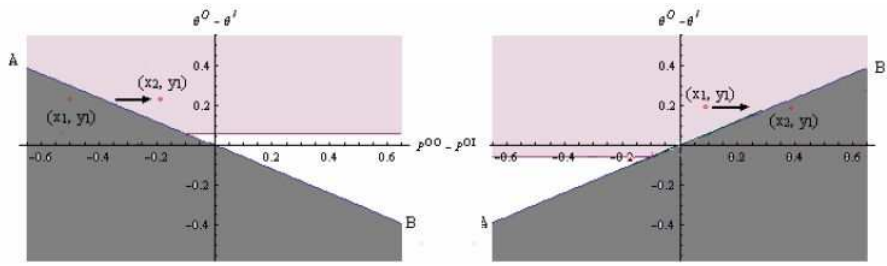


Fig. 15.5 Impact of MSSP specialization for $P^{0I} = 0.35$, $P^{II} = 0.45$ a)$R = 1.6$ (left) b)$R = 0.4$ (right).

It is worthwhile to compare Proposition 15.2 with prior results about firms' reluctance to share security information. Prior research showed that positive externality effects of information sharing may discourage firms to share security information with one another [17]. Leakage of sensitive information has also been cited as a reason for firms' reluctance to share security information [16] which may lead to the loss of the market value of the firm due to negative publicity [5]. Though we do not model information sharing explicitly, security experts have argued that security monitoring of two firms by the same MSSP is an indirect way of sharing breach information through an intermediary [33], and it is the information that the MSSP obtains by analyzing the breach on one firm that enables the MSSP to prevent a breach on the other firm, and thereby reducing the probability of both firms being breached. In sharp contrast to prior results, we find that firms may have a greater incentive to share security information by outsourcing the monitoring function if the MSSP can use that information to decrease the joint probability of breaches more. This result holds in the speculative risk environment because, as we stated earlier,

firms prefer an environment in which only one of them is breached to that in which both firms are breached in this environment. The difference between our results and that of prior research can be attributed to the observation that prior research considered the improvements in the efficiency of security investments in terms of a reduction in the breach probability enabled by information sharing whereas we consider the use of information to reduce the joint probability of breaches.

15.4.3 Impact of Breach Characteristics on Firms' Outsourcing Decisions

The parameters that characterize a breach in our model are the extent of spillover α, and the breach severity Δ. We note from (15.6)-(15.8) and from Lemma 1 that the breach parameters affect firms' decisions only through their impact on the value of R. Therefore, the security risk environment is determined partly by breach characteristics [4], and once the impact of these parameters on the risk environment is known, the impact on firms' decisions can be determined using Proposition 15.1(*ii*). We have the following result regarding the impact of breach characteristics on firms' outsourcing decisions.

Proposition 15.3. *(i) An increase in spillover increases the likelihood of both firms' outsourcing security monitoring and decreases the likelihood of both firms' outsourcing infrastructure management.*
(ii) An increase in breach severity increases the likelihood of both firms' outsourcing security monitoring and decreases the likelihood of both firms' outsourcing infrastructure management if $\alpha > \frac{2b_2}{2b_1 - b_2}$ *and decreases the likelihood of both firms' outsourcing security monitoring and increases the likelihood of both firms' outsourcing infrastructure, otherwise.*

Proposition 15.3(*i*) is a surprising result because one would expect that when the competition induced by spillover effects of security breaches becomes more intense, firms will prefer an environment in which the likelihood of only one firm being breached is smaller to one in which this likelihood is larger so as to mitigate the spillover effect. However, Proposition 15.3(*i*) implies the opposite. The reason for the counter-intuitive result can be attributed to the following. Using 15.5, we find that the price charged by the non-breached firm as well as by the breached firm increases in spillover (α). Therefore, the loss to breached firm (L) decreases in spillover and the benefit to non-breached firm (V) increases in spillover [5]. Since there is no switching when both firms are breached, α has no impact on L^b. Hence, an increase in the demand spillover caused by security breaches makes the risk environment more speculative, which favors outsourcing of security monitoring.

The impact of breach severity on the risk environment can also be explained using how an increase in Δ affects the reward and risk from a security breach. Con-

[4] Other demand parameters such as a, b_1, and b_2, also affect the risk environment.

[5] Note that a firm's profit is directly proportional to the square of price it charges.

sider the scenario in which one firm is breached. A larger value for Δ implies that the breached firm will face a larger reduction in demand and therefore a larger reduction in profit if it is breached. Therefore, the risk from getting breached is higher when Δ is larger. Now consider the scenario in which no firm is breached. If the competitor is breached, then the demand to a non-breached firm is higher when Δ is larger. However, the marginal increase in demand to a non-breached firm is smaller than the marginal decrease in the demand for the breached firm. In this scenario, the breached firm will set its price more aggressively, causing an even smaller increase in the reward to the non-breached firm. In order to make the risk environment more speculative (i.e., increase R) when Δ increases, the spillover rate has to be sufficiently large so that the increase in reward because of the competitor breach offsets the increase in risk.

In summary, our analysis shows that firms have stronger incentives to outsource security if the MSSP offers a higher quality in terms of preventing breaches compared to in-house management. However, a higher quality is neither a prerequisite nor a guarantee for a firm to outsource security. The competitive risk environment and the nature of the security function outsourced, in addition to quality, determine firms' outsourcing decisions. If the reward from the competitor's breach is higher than the loss from own breach, then even if the likelihood of a breach is higher under the MSSP, the expected benefit from the competitive demand externality may offset the loss from the higher likelihood of breaches, resulting in one or both firms outsourcing security. The incentive to outsource security monitoring is higher than that of infrastructure management because the MSSP can reduce the likelihood of breach on both firms and thus enhance the demand externality effect. The incentive to outsource security monitoring (infrastructure management) is higher (lower) if either the likelihood of breach on both firms is lower (higher) when security is outsourced or the benefit (relative to loss) from the externality is higher (lower). The benefit from the demand externality arising out of a security breach is higher when more of the customers that leave the breached firm switch to the non-breached firm.

15.5 Conclusion

The risks associated with IT security are fundamentally different from those associated with traditional IT functions. However, the reasons cited by both academics and security experts for why firms outsource IT security are the same as those cited for outsourcing of traditional IT functions. We believe that the IT security outsourcing decision is a strategic one in which a firm considers the ramifications of the competitor's action on its payoff and vice versa. Ignoring such strategic considerations and solely using criteria related to cost or quality measures in IT security outsourcing decision making process may result in sub-optimal decisions. To this end, while analyzing firms' decision to outsource IT security, we consider the information security risk not only as a form of non-speculative risk but also a form of speculative risk and analyze the impact of competitive externalities on firms' incen-

tives to outsource IT security. Thus we offer a novel explanation for firms' decision to outsource IT security based on such externalities.

We show that a firm's outsourcing decision depends critically on the interaction of the quality of the MSSP relative to that of in-house security management, MSSP specialization, and the risk environment. Consistent with the traditional explanation given for firms' outsourcing decision, we also found that if outsourcing leads to a lower probability of breach, then firms outsource security if competition is not an issue. However, because of the competitive externalities, firms may prefer outsourcing even if it does not reduce the breach probability. Nevertheless, an improvement in MSSP quality leads to more outsourcing. If firms operate in a speculative risk environment, then they outsource more if MSSP is specialized in monitoring and less if MSSP is specialized in management of security infrastructure. However, when firms operate in a non-speculative risk environment, then they outsource more if MSSP is specialized in management of security infrastructure and less if MSSP is specialized in monitoring. The risk environment becomes more speculative with increases in spillover and in breach severity if spillover is higher than a threshold.

We made a number of simplifying assumptions to make the analysis tractable. However, the qualitative nature of or results will likely hold even when we relax many of these assumptions. We discuss the impact of relaxing some of the more critical assumptions in the following paragraphs. One, we assumed that there is a single MSSP. Existence of multiple MSSPs complicates the analysis in two ways. The MSSPs may specialize in different security functions, and the two firms may outsource to different MSSPs. If the MSSPs offer the same function, then the analysis for the cases in which neither firm outsources, only one of the firms outsources, and both firms outsource to the same MSSP remains the same as in this paper. Even when the firms outsource to different MSSPs, if the firms outsource infrastructure management and the MSSPs apply similar procedures and best practices, then our analysis and results will hold. On the other hand, if the firms outsource security monitoring, then it is likely that the probability of breach events is not likely to be as low as when there is a single MSSP unless the MSSPs share their information about breach events. The modeling and analysis of the case when the MSSPS offer different functions, and the firms outsource to different MSSPs is challenging and requires further research. Two, we assumed identical firms, ex ante. A model with heterogeneous firms will offer insights into how firm-specific factors such as firm size affect outsourcing decisions. Three, we assumed MSSP parameters and the firms' investment in security as exogenous. However, some of these parameters can be dependent on each other, and endogenizing these parameters could be possible extensions to the model.

Appendix

Derivation of optimal prices

$\pi_i = p_i q_i = p_i(a - b_1 p_i + b_2 p_j + B_i)$

$\frac{\partial \pi_i}{\partial p_i} = a - 2b_1 p_i + b_2 p_j + B_i = 0 \Rightarrow p_i(p_j) = \frac{a + b_2 p_j + B_i}{2b_1}$. This implies that $p_i = \frac{q_i}{b_1} \Rightarrow \pi_i = (p_i^*)^2 b_1$. Solving the reaction functions simultaneously, we obtain the following optimal price, $p_i(p_j) = \frac{a(2b_1+b_2)+2b_1B_i+b_2B_j}{4b_1^2-b_2^2}$.

Lemma 1

Proof. When firm i outsources, firm j outsources if its payoff under outsourcing is higher than its payoff under in-house management, i.e. $-P^{OO}L^b(\theta^O - P^{OO})(V - L) > -P^{OI}L^b + (\theta^O - P^{OI})V - (\theta^I - P^{OI})L$. Replacing $R = \frac{V+L^b}{L}$, we get the (outsource, outsource) Nash equilibrium, $(\theta^O - \theta^I) < (P^{OI} - P^{OO})(R-1)$. Similarly, we obtain (in-house, in-house) Nash equilibrium, when (outsource, outsource) is not Nash equilibrium and $(\theta^O - \theta^I) > (R-1)(P^{II} - P^{OI})$ holds, that is $(\theta^O - \theta^I) > Max((R-1)(P^{II} - P^{OI}), (R-1)(P^{OI} - P^{OO}))$. When there is no pure strategy, firms outsource with the mixing probability calculated by payoff-equating method.

Proposition 1

Proof. (*i*) When there is no speculative risk, i.e., $V = L \Rightarrow R = 1, (\theta^O - \theta^I) < (P^{OI} - P^{OO})(R-1)$ holds when $\theta^O < \theta^I$.
(*ii*) Follows from the fact that the RHS of above inequality increases in R if $P^{OI} > P^{OO}$ and decreases otherwise.

Proposition 2

Proof. (*i*) Follows from the fact that the LHS of the inequality, $(\theta^O - \theta^I) < (P^{OI} - P^{OO})(R-1)$, is increasing in θ^O.
(*ii*) The RHS of above inequality is decreasing in P^{OO} when $R > 1$ and is increasing in P^{OO} when $R < 1$.

Proposition 3

Proof. (*i*) R is increasing in α and Proposition 15.1(*ii*) provides the proof.
(*ii*) The proof follows from Proposition 15.1(*ii*) and
$\frac{\partial R}{\partial \Delta} = \frac{4\Delta b_1(1+\alpha)(2b_1+b_2)(2b_1\alpha-(2+\alpha)b_2)}{(2b_1-\alpha b_2)((4a-2\Delta)b_1+(2a+\alpha\Delta)b_2)^2} < 0$ iff $\alpha < \frac{2b_2}{2b_1-b_2}$.

References

1. Ang, S., Cummings, L.L.: Strategic response to institutional influences on information systems outsourcing. Organization Science **8**(3), 235–256 (1997)
2. Ang, S., Straub, D. W.: Production and transaction economies and IS outsourcing: a study of the U.S. banking industry. MIS Quarterly **22**(4), 535–552 (1998)
3. Berthillier, A.: Managed security services for network service provider. Juniper Networks Inc. Solution Brief (2005)
4. Caldwell, T.: Downturn gives a lift to outsourcing. IT services and solutions, Management consultants association, p. 2 (2002)
5. Cavusoglu, H., Mishra, B., Raghunathan, S.: The effect of Internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce **9**(1), 70–104 (2004)
6. Chalos, P., Sung, J.: Outsourcing decisions and managerial incentives. Decision Sciences **29**(4), 901–919 (1998)

7. Conry-Murray, A.: Bye-bye independent managed security providers. Network Computing **17**(24), 18 (2008)
8. Dey, D., Fan, M., Zhang, C.: Design and analysis of contracts for software outsourcing. Forthcoming in Information Systems Research (2008)
9. Dibbern, J., Goles, T., Hirschheim, R.: Information systems outsourcing: a survey and analysis of the literature. The DATA BASE for Advances in Information Systems **35**(4) (2005)
10. Ding, W., Yurcik, W., Yin, X.: Outsourcing Internet security: economic analysis of incentives for managed security service providers. In: Proceedings of the Workshop on Internet and Network Economics (WINE). Hong Kong (2005)
11. Ding, W., Yurcik, W.: Outsourcing Internet security: the effect of transaction costs on managed service providers. In: Proceedings of the International Conference on Telecommunication Systems, Modeling and Analysis. Dallas, TX (2005)
12. Ding, W., Yurcik, W.: Economics of Internet security outsourcing: simulation results based on the Schneier model. In: Proceedings of the Workshop on the Economics of Securing the Information Infrastructure. Washington DC (2006)
13. DiRomualdo, A., Gurbaxani, V.: Strategic intent for IT outsourcing. Sloan Management Review, pp. 67–80 (1998)
14. Ernst & Young: Moving beyond compliance: Ernst & Young's 2008 global information security survey (2008)
`http://www.ey.com/Global/assets.nsf/UK/Global_Information_Security_Survey_2008/$file/EY_Global_Information_Security_Survey_2008.pdf`
15. Frost & Sullivan: World managed security service provider markets #7426–74. (2003)
16. Gal-Or, E. Ghose, A.: The economic incentives for sharing security information. Information Systems Research **16**(2), 186–208 (2005)
17. Gordon, L. A., Loeb, M., Lucyshyn. W.: Sharing information on computer systems security: An economic analysis. Journal of Accounting and Public Policy **22**(6), 461–485 (2003)
18. Gupta, A., Zhdanov, D.: Growth and Sustainability of Managed Security Services Networks: An Economic Perspective. In: Proceedings of the 6th Workshop on the Economics of Information Security (WEIS). Pittsburgh, PA (2007)
19. Gurbaxani, V.: Information systems outsourcing contracts: theory and evidence. In: U. Apte, U. Karmarkar (eds.) Managing in the Information Economy: Current Research. Kluwer (2006)
20. Hu, Q., Saunders, C., Gebelt, M.: Diffusion of information systems outsourcing: a reevaluation of influence sources. Information Systems Research **8**, 288–301 (1997)
21. IDC: U.S. Managed Security Services 2008–2012 Forecast and Analysis (IDC #213551). (2008)
22. Kavanagh, K. M., Pescatore, J.: Magic Quadrant for MSSPs. North America, 2H05, Gartner (2005)
23. Kavanagh, K. M., Pescatore, J.: Magic Quadrant for MSSPs. North America, 1H07, Gartner (2007)
24. Loh, L., Venkatraman, N.: Determinants of information technology outsourcing: a cross-sectional analysis. Journal of Management Information Systems **9**, 7–24 (1992)
25. Loh, L., Venkatraman, N.: Diffusion of information technology outsourcing: influence sources and the Kodak effect. Information Systems Research **3**(4), 334–358 (1992)
26. McGuire, T.M., Staelin, R.P.: An industry equilibrium analysis of down–stream vertical integration. Marketing Science **2**, 161–192 (1983)
27. McLellan, K.L., Marcolin, B.L. and Beamish, P.W.: Financial and strategic motivations behind IS outsourcing. Journal of Information Technology **10**, 299–321 (1995)
28. Messmer, E.: Outsourcing security tasks brings controversy. Network World (2008). `http://www.networkworld.com/news/2008/032008--outsourcing--security.html`
29. Palumbo, S.: The managed security services opportunity. Security Solutions & Services, Yankee Group (2006)
30. Ponemon Institute: Annual study: U.S. cost of a data breach understanding financial impact, customer turnover, and preventitive solutions (2007)
31. Ponemon Institute: Consumer survey on data breach notification (2007)
32. Richardson, R.: CSI computer crime and security survey (2007)

33. Rowe, B.: Will outsourcing IT security lead to a higher social level of security? In: Proceedings of the 6th Workshop on Economics of Information Security (WEIS). Pittsburgh, PA (2007)
34. Schneier, B., Ranum, M.: Face-off: Is security market consolidation a plague or progress. Information Security (2008)
35. Sen, S., Raghu, T. S., Vinze, A.: Demand heterogeneity in IT infrastructure services: modeling and evaluation of a dynamic approach to defining service levels. Information Systems Research **20**(2), 258–276 (2009)
36. Slaughter, S. A., Ang, S.: Employment outsourcing in information systems. Communications of the ACM **39**(7), 47–54 (1996)
37. Sobol, M. G., Apte, U. M.: Domestic and global outsourcing practices of America's most effective IS users. Journal of Information Technology **10**, 269–280 (1995)
38. Sridhar, S. S., Balachandran, B. V.: Incomplete information, task assignment, and managerial control systems. Management Science **43**(6), 764–778 (1997)
39. Tarantino, A.: Governance, Risk, and Compliance Handbook: Technology, Finance, Environmental, and International Guidance and Best Practices. John Wiley & Sons (2008)
40. Van Mieghem, J.A.: Coordinating investment, production, and subcontracting. Management Science **45**(7), 954–971 (1999)
41. Whang, S.: Contracting for software development. Management Science **38**, 307–324 (1992)
42. Wylder, J.: Strategic information security. Auerbach Publications (2004)