Rainer Böhme *Editor*

The Economics of Information Security and Privacy

Springer

The Economics of Information Security and Privacy

Rainer Böhme
Editor

The Economics of Information Security and Privacy

Springer

Editor
Rainer Böhme
Westfälische Wilhelms-Universität Münster
Münster, Germany

ISBN 978-3-642-39497-3 ISBN 978-3-642-39498-0 (eBook)
DOI 10.1007/978-3-642-39498-0
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013955004

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In the late 1990s, researchers began to grasp that the roots of many information security failures can be better explained with the language of economics than by pointing to instances of technical flaws. The first *Workshop on the Economics of Information Security*, WEIS in shorthand, took place in Berkeley, California, in 2002. This series of annual events laid the foundations of a thriving new interdisciplinary research field combining economic and engineering insights, measurement approaches, and methodologies to ask fundamental questions concerning the viability of a free and open information society and to provide answers where possible. While economics and information security is no doubt the nucleus of an academic movement that quickly drew the attention of think tanks, industry, and governments, WEIS expanded its scope to include even more perspectives. Surrounding areas include management of information security, privacy, and (more recently) cybercrime, all studied from an interdisciplinary angle by combining methods from microeconomics, econometrics, qualitative social sciences, and (also more recently) behavioral sciences as well as experimental economics.

This volume contains a selection of 13 revised contributions to the 11th WEIS, which took place in Berlin, Germany, on 25–26 June 2012. It is structured in four parts, reflecting the main areas in the scope of WEIS: Management of Information Security, Economics of Information Security, Economics of Privacy, and Economics of Cybercrime. Each individual contribution documents, discusses, and advances the state of the art concerning its specific research questions. Beyond this, our intention is that this volume in its entirety draws a comprehensive picture of current research questions, the breadth of methodological approaches, and their relevance for designing (or engineering) a secure and livable information society. We hope that the readers find the topics as fascinating as the authors and that the book stimulates and structures discussions among and between academic researchers, practitioners, media representatives, and policy makers. In the best case, the results of such discussions feed into research contributions submitted to and presented at future editions of WEIS.

Let me take this opportunity to thank all the people who contributed to make WEIS 2012 in Berlin a success. Gert G. Wagner and Nicola Jentzsch, both from the economic research institute DIW Berlin, served as general chairs and hosted the conference in Berlin, supported by Denis Huschka and his team, Petra Holthöfer, Claudia Kreutz, Claudia Oellers, Sören Schumann, and Jörg Wernitz. They gave us a warm welcome in Berlin's historical downtown with an outstanding Berlinish social event. In preparation of the conference, the Program Committee did the hard work of screening and discussing all submitted papers, of which the best were selected for presentation. Many distinguished researchers in the various fields of WEIS volunteered to serve on the Program Committee, including Alessandro Acquisti (Carnegie Mellon University), Ross Anderson (University of Cambridge), Rainer Böhme (University of Münster, Program Chair), L. Jean Camp (Indiana University), Jonathan Cave (RAND Europe), Huseyin Cavusoglu (University of Texas at Dallas), Nicolas Christin (Carnegie Mellon University), Michel van Eeten (Delft University of Technology), Benjamin Edelman (Harvard Business School), Allan Friedman (Brookings Institution), Jeremy Epstein (SRI International), Neil Gandal (Tel Aviv University), Dan Geer (In-Q-Tel), Lawrence Gordon (University of Maryland), Jens Grossklags (Penn State University), Thorsten Holz (Ruhr-University Bochum), Jean-Pierre Hubaux (EPFL Lausanne), Nicola Jentzsch (DIW Berlin), M. Eric Johnson (Dartmouth Tuck School of Business), Kanta Matsuura (University of Tokyo), Martin Loeb (University of Maryland), Tyler Moore (Southern Methodist University), Andrew Odlyzko (University of Minnesota), David Pym (University of Aberdeen), Brent Rowe (RTI International), Stuart Schechter (Microsoft Research), Bruce Schneier (BT Counterpane), Richard Sullivan (Federal Reserve Bank of Kansas City), Rahul Telang (Carnegie Mellon University), Catherine Tucker (MIT), Liad Wagman (Illinois Institute of Technology), and Rick Wash (Michigan State University). Nevena Vratonjic (EPFL Lausanne) contributed reviews on behalf of one committee member. The Program Chair received additional support from his colleagues at the University of Münster. Lars Greiving designed an appealing website (using artwork by Claudia Kreutz), Pascal Schöttle gave organizational support wherever needed, Ursula Kortemeyer administrated the student travel grants, and Benjamin Johnson helped out with proofreading essential conference material. A special thanks is due to Malte Möser, who served as an outstanding editorial assistant for the compilation of this volume. Without his hard work in corresponding with all contributors and meticulously keeping track of all changes, this volume would not be in its current shape. Ronan Nugent of Springer was our always helpful contact point with the publisher. The last thank you goes to the WEIS Steering Committee for entrusting us to bring this conference to Germany, the first time in continental Europe.

Finally, we could not have run the conference at this scale without the generous support of our sponsors. First and foremost, Volkswagen Foundation supported young researchers with a number of travel grants and helped us to offer a heavily subsidized registration fee for student participants. Google Inc. served as a platinum

and dinner sponsor which supported the unforgettable social event. Siemens Enterprise Communications GmbH served as our gold sponsor, and additional financial support from Facebook Inc. is gratefully acknowledged.

Münster, Germany Rainer Böhme
October 2012

Contents

Contributors

Ross Anderson Computer Laboratory, University of Cambridge, Cambridge, UK

Daniel Bachlechner Department of Information Systems, School of Management, University of Innsbruck, Innsbruck, Austria

Chris Barton Security Research and Operations, Cloudmark, Inc., Reading, UK

Jörg Becker European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany

Alastair R. Beresford Computer Laboratory, University of Cambridge, Cambridge, UK

Rainer Böhme European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany
Department of Information Systems, University of Münster, Münster, Germany

Matthias Brecht University of Regensburg, Regensburg, Germany

Dominic Breuker European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany

Sacha Brostoff University College London, London, UK

L. Jean Camp School of Informatics and Computing, Indiana University, Bloomington, IN, USA

Richard Clayton Computer Laboratory, University of Cambridge, Cambridge, UK

Lukas Demetz Department of Information Systems, School of Management, University of Innsbruck, Innsbruck, Austria

Serge Egelman University of California, Berkeley, CA, USA

Adrienne Porter Felt Google Inc., Mountain View, CA, USA

Vaibhav Garg Indiana University Bloomington, IN, USA

Jens Grossklags College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA

Tobias Heide European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany

Justus Holler European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany

Jean-Pierre Hubaux School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland

Bongkot Jenjarrussakul Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo, Japan

Charlene Jennett University College London, London, UK

Chris Kanich University of Illinois at Chicago, Chicago, IL, USA

Timothy Kelley School of Informatics and Computing, Indiana University, Bloomington, IN, USA

Ayako Komatsu Security Economics Laboratory, IT Security Center, Information-Technology Promotion Agency, Bunkyo-ku, Tokyo, Japan

Kat Krol Department of Computer Science and Security Science Doctoral Research Training Centre (SECReT), University College London, London, UK

Michael Levi School of Social Sciences, Cardiff University, Cardiff, UK

Miguel Malheiros University College London, London, UK

Mohammad Hossein Manshaei Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran

Kanta Matsuura Institute of Industrial Science, The University of Tokyo, Meguro-ku, Tokyo, Japan

Tyler Moore Department of Computer Science and Engineering, Southern Methodist University, Dallas, TX, USA

Stephan Neuhaus Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

Thomas Nowey Krones AG, Neutraubling, Germany

Bernhard Plattner Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland

Sören Preibusch Computer Laboratory, University of Cambridge, Cambridge, UK

Hans Peter Rauer European Research Center for Information Systems (ERCIS), University of Münster, Münster, Germany

M. Angela Sasse University College London, London, UK

Stefan Savage Department of Computer Science and Engineering, University of California, San Diego, CA, USA

Toshihiko Takemura The Research Institute for Socionetwork Strategies, Kansai University, Suita, Osaka, Japan

Hideyuki Tanaka Graduate School of Interdisciplinary Information Studies, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

Nevena Vratonjic School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland

Michel J. G. van Eeten Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands

David Wagner University of California, Berkeley, CA, USA

Part I
Management of Information Security

Chapter 1
A Closer Look at Information Security Costs

Matthias Brecht and Thomas Nowey

Abstract Economic aspects of information security are of growing interest to researchers and to decision-makers in IT-dependent companies. From a business-perspective, cost-benefit justifications for information security investments are in focus. While previous research has mostly focused on economic models for security investments, or on how to quantify the benefits of information security, this chapter aims to take a closer look at the costs of information security. After providing the reader with basic knowledge and motivation for the topic, we identify and describe the problems and difficulties in quantifying an enterprise's cost for information security in a comprehensive and comparable way. Of these issues, the lack of a common model of costs of information security is the most prominent one. This chapter also discusses four approaches to categorize and determine the costs of information security in an enterprise. Starting with the classic approach frequently used in surveys, we continue by describing three alternative approaches. To support research on the costs of information security we propose two metrics. We conclude with input for future research, especially for an empirical analysis of the topic.

1.1 Introduction

Applying methods from microeconomics and business administration to the field of information security (IS) has become popular. One important aim of that kind of research is to get a quantitative perspective on information security. In general the advantages of quantification are its accuracy, objectivity, and comparability.

M. Brecht (✉)
University of Regensburg, Regensburg, Germany
e-mail: infosec@brecht.me

T. Nowey
Krones AG, Neutraubling, Germany
e-mail: thomas.nowey@krones.com

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_1,

In addition, quantification is the basis for calculations and statistical analyses. Nowadays also security investments have to be compared with other investments and cost-benefit analyses of security investments, to answer Kevin J. Soo Hoo's question of how much security is enough [17].

So far most research has focused on economic models for cost-benefit evaluation and on decision rules. When it comes to data the focus was mainly on the provision of data for the quantification of IS risks respectively the benefits of information security. In the following we want to take a look at the other side of the balance sheet – the costs of information security.

Being able to accurately determine security costs is a prerequisite for any cost-benefit calculation. Another important field of application for a cost model for information security is benchmarking between different companies. This could foster, for example, the comparison of the percentage of the IT budget, or the absolute budget that is spent on information security. To the best of our knowledge, today there no suitable or applicable model available for the costs of information security in commercial enterprises. A cost model could be used by, for example, Chief Information Security Officers (CISOs), budget planners, financially responsible staff or managers as a common basis for communication and for decisions.

The remainder of this chapter is structured as follows. Section 1.2 provides a brief overview of related work. In Sect. 1.3 we present challenges in quantifying the costs of information security. In Sect. 1.4 we analyze existing approaches to categorize security costs and introduce two new approaches to the topic. In Sect. 1.5 a conclusion of this work as a whole and promising topics for future work are presented.

1.2 Background and Related Work

Since researchers have shown the importance of economic aspects for information security research (cf. [2]) the topic has been further developed in various directions reaching from behavioral theory to risk management. In the context of this chapter it is especially important to consider approaches with a cost focus that have an application in business administration and risk management.

1.2.1 Cost-Benefit Evaluation of Information Security

After a time in which the often cited fear, uncertainty, and doubt (FUD) strategy was used to sell investments in security (cf. [4]) practitioners as well as researchers are now looking for methods that allow for quantitatively founded cost-benefit evaluations of information security measures. To identify costs or benefits of security measures it is necessary to determine both the expected damage before and after a security measure has been taken and the costs for this measure.

Scholtz points out that information security professionals need to articulate the value of their activities in business terms [35]. He states that especially during bad economic times, only security initiatives that are able to demonstrate clear business value will be funded. However, due to a lack of comparable historical data, security staff must continuously work to evolve alternative mechanisms to capture and articulate the business value of measures. At a project level, a possible approach could be to analyze the expected risk reduction, quantifiable financial return or other expected improvements.

Soo Hoo [17] states that information security management (ISM) needs analytic, decision-focused and quantitative techniques to address many of the failings of previous modeling paradigms and to answer the question: How much is enough? During the last decade this changed and IT security management is now increasingly based on economic principles. This also means that a balance between costs and benefits of IT security is necessary (cf. [26, 27, 32]) and that investments in security have to be geared towards the principle of economic efficiency. This includes that in the case of an economic revision they have to withstand the then applied measures. Starting with the idea of a Return-on-Security Investment (ROSI) several concepts have been developed to support the decision for or against an information measure. One way to do this is to apply the concept of Net Present Value (NPV). Faisst et al. have developed a NPV formula for information security investments [9]:

$$NPV = -I_0 + \sum_{t=1}^{T} \frac{\Delta E(L_t) + \Delta OCC_t - C_t}{(1 + i_{\text{calc}})^t} \tag{1.1}$$

where

I_0 = initial investment for security measure

$\Delta E(L_t)$ = reduction in expected loss in t

ΔOCC_t = reduction in opportunity costs in t

C_t = costs of security measure in t

i_{calc} = discount rate

The presented model returns a positive or negative value and thus advises an enterprise to make a security investment or not. The costs of a security measure are treated as a single value without advising how it could be determined. Besides that most of the ROSI approaches are aimed at single security measures and do not consider information security management.

Gordon and Loeb provide security related cost-benefit guidelines for companies [15]. The authors present analyses and answers to questions like how to determine the right amount of money to spend and where to invest, but also explains the role of risks in the allocation of resources, strategies to minimize the impact of incidents, and an approach to articulate business values to ensure future funding. In addition, NIST's risk management guide emphasizes the importance of

risk management in today's ISM and guides through the different phases of risk assessment and cost-benefit analyses [29].

While Longstaff et al. propose a hierarchical model to assess the security risks of IT [25], Gordon and Loeb developed an economic model to determine the optimal level of investment in information security [14]. In addition, Soo Hoo provides a decision-analytic framework to evaluate different IT security policies [17]. From the point of view of our research these suggestions all have one major shortcoming: they treat security investments as a black box (see [6]).

1.2.2 Costs of Cyber-Crime

In the approaches mentioned in Sect. 1.2.1 the term costs of information security always refers to the necessary investments for information security measures. Contrary to that, sometimes also the costs caused by a lack of information security – mostly referred to as costs of cyber-crime – are denoted as the costs of information security.

Although our research is not focused on cyber-crime, it is worth taking a look at this area, since the difficulties with quantification are similar. They become visible in several cyber-crime surveys. Florêncio and Herley show the discrepancy between several studies that tried to determine the costs related to cyber-crime [11]. The Federal Trade Commission (FTC) estimated the losses due to identity theft in 2004 at $47 billion [7], in 2006 at $15.6 billion [8], and in 2008 at $54 billion. The huge drop in 2006 seems odd and leads to the assumption that these estimates are extremely noisy. In addition, during the last 2 years alone, claims can be found that show or predict losses or damages due to cyber-crime from $560 million to $1 trillion (cf. [1, 20, 28, 36]).

This reveals the difficulties in finding consistent and comparable values and measures for information security related costs. Florêncio and Herley point out that with the existing lack of consistency there remains a large room for interpretation [11].

1.2.3 Surveys on Costs of Information Security

As shown above in Sect. 1.2.2 surveys on cyber-crime show a great variety in estimated costs or losses. A similar phenomenon can be recognized when looking at surveys on costs of information security.

Every year, several different information security spending surveys are performed, where at least partly surprising results can be found:

- Sullivan finds that almost 70 % of respondents spend less than 7 % of the overall IT budget on information security, an increase of companies that outsource the entire information security function by 80, and 55 % of respondents in medical practices use either part-time or external staff to handle security [37].

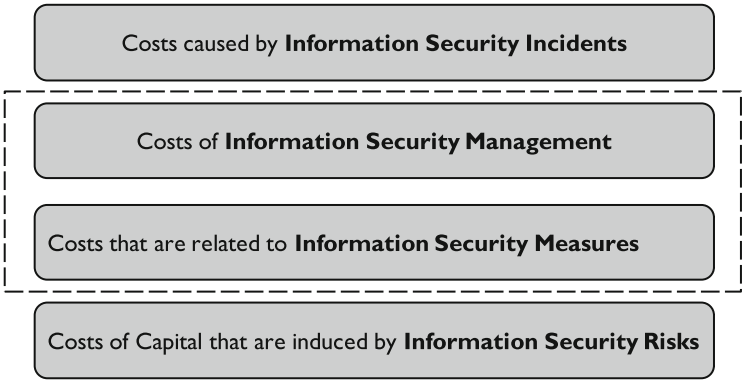

Fig. 1.1 Four aspects of costs of information security

- Penn states that the security portion of the IT budget is expected to rise by 12.6 % in 2009, up from 7.2 % in 2007 and 11.7 % in 2008 [31].
- Weigelt found an increase in the IT security spending of the Bush administration in 2009 of $646.8 million. Agencies as a whole would spend 10.3 % of their IT budget on information security. Spendings for cyber-security were planned to be increased by 73 % since 2004 [38].

Standards or universally accepted conventions, for example, for structures of IT costs, are to the best of our knowledge not available.

1.2.4 Definitions

Especially due to the growing importance of economic values for any enterprise, it is increasingly important to measure the efficiency and effectivity of security measures (cf. [21]). This makes it necessary to identify costs and benefits. Kütz defines costs in the context of ISM as the evaluated use of resources in monetary terms [22]. The same author states that the benefits of security management are often expressed as the avoided damage. Before a security measure is introduced an objective estimation of these two values is used to decide whether it will be deployed or not. Afterwards, the decision that was made should be revisable and comprehensible.

For the purpose of this chapter it is necessary to define the term costs of information security more precisely. We identify four main interpretations of the term (see Fig. 1.1):

1. The costs that are caused by information security incidents.
2. The costs of managing information security.
3. The costs that are related to information security measures.
4. The costs of capital that are induced by information security risks.

This chapter will focus on the two middle layers. Therefore for the remainder of this chapter we define the scope of costs of information security as follows:

> The term costs of information security refers to costs that are associated with all kinds of measures or activities – including technical as well as organizational aspects – within an organization that are aimed at reducing information security risks for its information assets.

1.2.5 Costs of Quality

Well-established standards for information security like ISO/IEC 27001 require a management system approach for information security. In this regard there is an analogy to the field of quality management. Some companies even integrate both fields in integrated management systems. Since quality management has a long history it should be worth taking a look at approaches for quality costs.

Schiffauerova and Thomson give an overview over the field of Cost of Quality (CoQ) [34]. They point out that most companies have implemented the so-called prevention-appraisal-failure (P-A-F) model stemming from [10] that divides CoQ into three subcategories: the cost for failure prevention, the cost for failure approval plus the cost induced by failures. Besides the P-A-F-model there are numerous approaches for categorizing CoQ. The choice of the appropriate model depends on factors like purpose, situation, environment and individual needs [34].

According to [34] companies that adopt CoQ concepts improve their quality while reducing the cost for quality. Yet the authors also point out one main difficulty in assessing CoQ: Since most CoQ measurement methods are activity oriented they do not mesh traditional cost accounting that is more expense oriented. Consequently there is no standard procedure for determining and categorizing data on quality costs.

1.3 The Challenges in Quantifying Security Costs

It is widely agreed that quantifying the benefits of information security measures is hard (see for example [30]). In this section we present multiple issues that underline that also the quantitative determination of costs related to information security has to overcome some serious challenges.

1.3.1 Information Security as a Cross-functional Task

During the past decade we have seen a development from IT-security to information security. Information security covers all activities to protect the confidentiality, integrity and availability of an organization's information assets. Therefore it is

widely agreed that information security is more than just technical measures. Providing information security is a cross-divisional task that encompasses technical (e.g., hardware, software) as well as organizational (e.g., employee trainings, processes) aspects. With information security awareness becoming more and more important virtually every employee in a company can do her bit to provide information security.

This cross-divisional nature is a huge challenge for categorizing and analyzing costs of information security. On the one hand costs of information security cannot be easily mapped to one single category of traditional cost accounting. On the other hand it is not easily possible to define what part of the costs of a measure are directly accountable to information security.

Even in the case of investments whose costs are closely related to information security, such as very strict programming guidelines or the operation of a firewall, it may be hard to determine that part of the investment that may be accounted to information security. Programming guidelines are used to improve the security of a company's products, but in a case where employees write hundreds or thousands of lines of code (LOC) a day no one can tell what amount of time is really invested in security. Also, a hardware firewall is definitely a security product, but if it can also act as an e-mail gateway, are really all related costs accountable to security?

The aforementioned examples show one of the problems that occur during the management of information security; it is hard to determine if a task is an information security task or if it is another task with some part concerning security.

Another problem is revealed when we look at the focus of information security management. Information security management is a systematic, long-term, cross-divisional task with the aim of protecting a company's assets and goals (cf. [16]). Since expenses for security are in general made to meet the goal of long-term security, security products are usually very complex. Often, the initial costs (e.g., purchase price of the product, the product's introduction, operation, but also security-related adaptations of business processes) may only account for a fraction of the overall costs. This does not only mean that the decision for or against security investments needs to be considered carefully, but also makes the determination of the actual costs of a security measure much harder.

1.3.2 Divergent Goals of Cost Quantification

Organizations have different reasons and goals for quantifying costs in general, and costs of information security in particular. Different perspectives on costs of information security are needed to approach the various possible information needs. Table 1.1 provides the reader with a brief overview of the different goals of cost quantification. Thus, a flexible cost model is required that can satisfy different demands by enabling various perspectives on costs of information security in an enterprise.

Table 1.1 Goals of cost quantification

Goal	Explanation/implications
Budgeting	Providing guidelines on how much may be spent, categorization to provide internal comparability, oriented towards general controlling and accounting guidelines
Cost accounting	Usually, no special way of dealing with security, main goal is to meet compliance regarding financial aspects
Benchmarking	Comparability with other organizations, identification of differences, point out different strategies or starting points
Risk management	Preparation for controlling decisions, determine advantageousness of security investments/measures
Cost-benefit analysis of investments/projects	Economic assessment of certain measures/projects, return on investment analyses, the overall costs of a measure or project need to be identified
Surveys/research	Identification of trends, tendency towards higher/lower security spending, determination of preferences (technical/organizational measures)

1.3.3 Hidden Costs: For Example, Security-Related Outsourcing

A major challenge in analyzing security costs is what we call hidden costs, i.e., costs that are at first glance not directly related to a security risk management decision, but indirectly caused by it. A good example is the field of outsourcing, where various hidden costs have been identified in the past.

According to Schaffry, the investments in security services will continue to grow steadily over the next years [33]. One of the main reasons the author identifies is the increasing popularity of Managed Security Services (MSS). MSS change the market significantly. MSS describe the outsourcing of operation and management of an enterprise's security solutions in order to save money, but this outsourcing relationship has to be managed and its results have to be reviewed and verified. For an overview of how to manage an IT outsourcing relationship see [23].

Barthélemy identified four often forgotten and hidden costs of IT outsourcing [3]. These are categorized into costs that occur during the search for a suitable vendor and the contracting phase, during the transition phase of switching the in-house delivery of a service to another company, or during the transition from the outsourcing partner to another outsourcer or the reintegration of the formerly outsourced service. In addition, also costs for the management of the outsourcing relationship (e.g., monitoring, bargaining, and renegotiation) must not be forgotten.

1.3.4 Difficulties in Finding the Right Baseline

An additional challenge, especially for benchmarking of costs of information security, is finding the right baseline. As one can see from the facts mentioned in

Sect. 1.3.1, the nature of costs of information security makes it unlikely that this kind of cost can be seen as a subset of IT costs and thus be put in relation to the overall IT budget of a company. This procedure may work for IT security costs, but in the case of information security costs one will always be able to cover only parts of the overall costs. The main reason for this is that IT security can be seen as a subset of information security, regarding only an organization's IT.

Some results show that following the increase of the number of security threats and incidents, also the budget spent on information security in relation to the overall IT budget increases. Other results show that especially in industries with highly sensitive information, the importance of information security may not have been fully understood. In addition, it seems that in bad economic times other topics that may help companies to increase their profits or strengthen their market position have higher priority (cf. [5]). In general, most of the results shown above use the overall IT budget of a company as the main reference. This is – at least partly – problematic since the costs that are accumulated to the IT budget may be significantly different for several companies or industries. For example, the costs for telephony and mobile telephony are part of the IT costs in some companies while in others they are not.

1.4 Towards a Model for Categorizing Costs of Information Security

The challenges identified above lead to the conclusion that a common understanding of costs of information security is required. Therefore, a common way of categorizing and structuring costs in a repeatable and comparable way is required. Building on that basis, it becomes possible to identify cost-drivers and to analyze different security management approaches. In the following we present different approaches to structure costs of information security. We will refer to such a categorization by the term cost model. We will focus on enabling benchmarking between organizations, but will also take a look at other areas of application.

1.4.1 Approach 1: The Balance Sheet Oriented Approach

Benchmarking initiatives are frequently driven from an organization's controlling or accounting department. Thus, it is quite common to use structures for classifying costs that are oriented towards the chart of accounts. A typical example for that type is the model developed by the consulting firm Gartner[1] for Total Cost of

[1] http://www.gartner.com

Table 1.2 Cost categories for information security (used by Gartner [13])

Cost category	Description
Personnel costs	Includes all personnel costs supporting information security functions
Hardware	Dedicated security hardware (e.g., security gateways, disaster recovery hardware)
Software	License costs of software dedicated to managing security systems (e.g., IAM, endpoint security suites)
Outsourcing/managed security services (MSS)	Costs of monitoring/managing security devices, systems and processes or other costs related to MSS

Ownership (TCO)[2] analyses of Information Systems (cf. [12]). Equally, for the field of information security, Gartner chose an approach that could be called balance sheet oriented.

To cover the costs spent on information security, Gartner uses a scheme which distinguishes between the four different cost categories presented in Table 1.2. For the year 2011, they found a distribution of the information security budget into 21 % hardware, 29 % software, 40 % personnel and 10 % outsourcing.

In general, this approach is a first step towards the classification of information security costs. The classification in hardware, software, personnel and outsourcing may also be good for IT-related budget planning.

As previously shown in Sect. 1.2, the classification of security costs into hardware or software is problematic, if at all possible. Accordingly, Gartner's approach leads to a situation in which comparability between several industries or even single companies – due to a lack of transparency in the procedural method – cannot be provided. However, this approach allows companies to easily determine their costs in those categories because of existing accounting/budgeting processes. On the other hand, a more detailed analysis of the results of this approach is not possible since too much information about the creation of reference data is missing or unclear. Thus, a comparability between several companies may hardly be possible. In addition, this approach focuses more on IT-security than on information security.

1.4.2 Approach 2: The Security Measure Life Cycle Approach

Particularly when it comes to investment decisions on information security measures, decision-makers have the goal to capture the TCO of a measure. This leads to an approach that not only covers the costs of purchase for a security measure,

[2]TCO is a financial approach to help managers or consumers to estimate the overall costs of a product over its whole life cycle. It can also be used to determine the economic value of an investment and contains both acquisition and operation costs.

but also other costs within its life cycle. An example of such a categorization is sketched in [30]:

- Costs of purchase,
- Costs of setup,
- Costs of operation, and
- Costs of change.

This approach is well-suited for cost-benefit analyses of single measures since it covers all aspects of costs that are connected to the implementation of a security measure. Thus, the values can be compared to the potential benefits (mainly risk reduction) of a security measure. Likewise, the approach can be used for the selection of different security measures. However, it is hardly possible to apply this approach to a company's information security management as a whole since it lacks a process perspective and is mainly focused on IT. In addition, this approach is not suitable for benchmarking between several companies as no reference values are calculated or presented.

1.4.3 Approach 3: IT-Security Process Oriented Approach

The categorization proposed by Humpert-Vrielink and Vrielink is depicted in Fig. 1.2 [18]. It gives a comprehensive picture of costs associated with IT-security activities. Even though the focus is on single security measures, costs for operation also cover some high-level aspects like change of processes. The approach covers all of the four cost aspects of Sect. 1.4.2 in one category (cost of tools), making it more comprehensive. Nevertheless it is still focused on IT-security.

The categories leave room for interpretation, e.g., some may have the opinion that the complete part of costs for operation should be accounted to the costs for the tool section. This, and the fact that the categories are not compatible with standard cost accounting models, complicate the collection of data.

Another fact that may lead to controversy is mentioning costs of risk as actual costs of information security. Locher states that higher risks, meaning higher uncertainty, leads to higher interest rates for fresh capital [24]. The reduction or elimination of risks, which is a major goal of the management of information security, would lead to lower interest rates and so decrease these costs. In any case, information security measures should rather be seen as a means to reduce risks than as a trigger for costs.

All of the models mentioned so far try to categorize costs of information security, but tackle their goal in a completely different way. This is additional proof of the need for a universally accepted and applicable cost model to provide comparability and a universal understanding of the topic.

Since providing comparability between several companies or even industries is one of the main goals of the cost model we hope to develop, we will propose

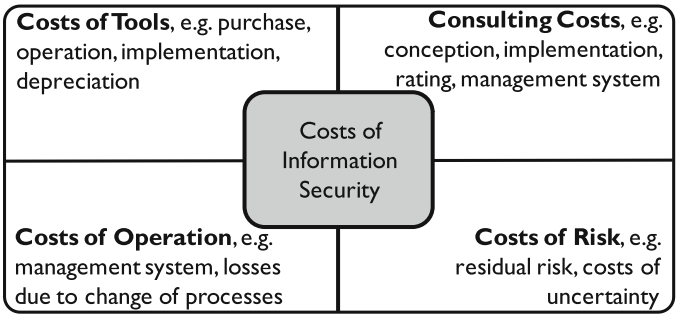

Fig. 1.2 Dimensions of costs of information security (cf. [18])

our own, more specific approaches to classify and determine the costs for information security in an enterprise in the following sections.

1.4.4 Introducing Determinability and Security-Cost Ratio

We introduce two new metrics that should help to describe and determine the cost ratio of a security measure or investment.

When it comes to benchmarking between organizations, we are facing the challenge that some of the cost categories may be very unambiguous and easy to determine, while others leave much room for interpretation. Companies should be aware of this issue when interpreting deviations from the benchmark. Therefore, *determinability* describes how difficult the determination of the related costs is in practice. The value of determination is indicated on a scale from "easy" to "hard", with intermediate steps of "easy–medium", "medium", and "medium–hard". For example, the determinability of the control "human resource management" is defined as "medium–hard". With empirical data available we could also analyze the statistical spread.

A challenge that is closely related to the latter is the decision of what proportion of a cost category should be attributed to information security. *Information Security Cost Ratio* describes the real percentage of the costs that may be accounted to information security. This value is indicated on a scale from "low" to "very high". Intermediate steps are "medium", "high–very high" and "very high".

1.4.5 Approach 4: The ISO/IEC 27001 Oriented Approach

The international standard ISO/IEC 27001 has a high acceptance and distribution in companies around the world. This section will provide an approach for categorizing

the costs of information security based on the ISO/IEC 27001 standard (cf. [19]). The different controls that are relevant for information security are shown and described in Table 1.3.

This approach should provide organizations with possibilities to determine costs of the controls and control areas, and guide companies in their decision on how much to invest in which control.

For the support of business decisions like this, distinguishing between hardware or software as suggested in Sect. 1.4.1 would not provide any help for an organization (cf. Sect. 1.2.1).

Besides the explanation of the cost aspects, Table 1.3 also gives an indication for the values of determinability and the information security cost ratio. Some examples are especially noticeable: Within human resource management, only parts of certain processes are information security motivated. The determination of the mentioned ratios is definitely not easy, but requires a detailed analysis of these processes. The information security cost ratio of "information security incident management" was set to be "very high". This is the case because this control can be seen as very closely related to, and almost exclusively motivated by information security management. However, if an organization uses its conventional incident management processes also for information security incidents, the information security cost ratio may be significantly lower.

Managing information security in an organization can be seen as an information security task. In addition, the technical aspects of this control are usually exclusively information security motivated. Thus, its costs can mostly be seen as costs of information security only.

Due to a lack of related literature, the chosen values derive from the practical experience of the authors. These values (especially the information security cost ratio) need to be researched in detail and should be subject to future research towards a cost model of costs of information security (cf. Sect. 1.5.2).

This discussion reveals that the costs of information security can originate from various different departments or directions. For several controls only parts of the overall costs can be accounted to information security. In addition, one might argue that, for example, the control human resource management and thus its costs cannot be seen as part of the IT budget of a company. The same may be true for physical security.

In general, the mentioned controls and related measures could be further classified. For example, regarding their main aspects like

- Organization,
- People,
- Technology or
- Processes.

It goes without saying that it is possible to implement and operate information security measures that are not mentioned in Appendix A of the ISO27001 standard. However, for the purpose of benchmarking, the participants need to agree on a set of common controls.

Table 1.3 Overview of costs of information security (according to Appendix A of ISO/IEC 27001 [19])

Control	Description	Determinability	Information security cost ratio
A.5 Security policy	Controls to provide management direction and support for information security in accordance with business requirements and relevant laws	Easy	Very high
A.6 Organization of information security	Controls for the organization of an enterprise's information security	Medium	Very high
A.7 Asset management	Controls for the management of an enterprise's assets (e.g., (de)classification)	Medium	Medium
A.8 Human resources security	Controls for the reduction of risk of human error, fraud, theft or misuse of facilities (e.g., training)	Medium–hard	Low
A.9 Physical and environment security	Controls to achieve security due to the prevention of unauthorized access, damage and interference to business premises or information	Easy–medium	Medium
A.10 Communications and operations management	Controls in order to ensure the correct and secure operation of information processing facilities	Medium	Medium
A.11 Access control	Controls to ensure only authorized access to information	Medium	High
A.12 Information systems acquisition, development and maintenance	Controls for security motivated/related costs for purchasing, development or maintenance	Medium	Medium
A.13 Information security incident management	Controls to help dealing with information security incidents	Medium–hard	Very high
A.14 Business continuity management	Controls needed in order to ensure business continuity or disaster recovery	Hard	Medium
A.15 Compliance	Controls to avoid violation of law, other obligation or any other security requirements	Medium–hard	Medium
4–8 ISMS	The mandatory parts of the ISMS as described in Chaps. 4–8 of ISO/IEC 27001, including for example risk management, internal audits, etc.	Easy–medium	Very high

So far we have only covered the controls that are part of Appendix A of ISO/IEC 27001. The ISMS itself – being obligatory and containing activities like risk management and internal audits – can be found in Chaps. 4–8 of the standard. Although harder to categorize, this part of the standard should also be part of a cost model, e.g., as one block called ISMS-costs.

Since the ISO/IEC 27001 standard is well accepted, widely distributed and standardized, this approach is well-suited for benchmarking and research. In addition, due to the fact that management aspects as well as technical and organizational measures are covered, this approach can also be used to provide an organization's security management with an overall view. However, since single security measures cannot be examined in detail, this approach is not suitable for cost-benefit analyses.

1.4.6 Approach 5: The ISMS-Layers Approach

Our second proposed approach takes the perspective of information security management. To achieve a top-down determination and classification of costs for information security, we suggest an approach that basically consists of four layers, plus basic prerequisites. In contrast to other approaches, including the one distinguishing personnel, software, hardware and outsourcing mentioned in Sect. 1.4.1, we are looking for an approach that better meets the cross-divisional characteristics and nature of information security. In addition, this approach could easily provide the possibility to compare if an organization, for example, spends too much money on management tasks or too little on a certain type of security measure.

An overview of this approach is shown in Fig. 1.3. Next to each of the five layers of this approach, gray-scale circles are located indicating the ratio of the costs that are directly accountable to information security (information security cost ratio). In the same way as described in Sect. 1.4.5, these are vague, approximate values that were set provisorily. These values also derive from the practical experience of the authors and should be scientifically and comprehensibly determined in future research work (cf. Sect. 1.5.2).

Details for each of the layers and their information security cost ratio, in ascending order, are presented in the following paragraphs.

The bottom layer builds the basis of this approach and describes the prerequisites of an efficiently and effectively working information security management. Since these prerequisites (e.g., inventory of assets, or the introduction of information ownership) are necessary management tasks to provide security for an enterprise's assets, its costs should not be accounted to information security since this would falsify the results. However, information security management is often the trigger for providing these basics or at least it requires the documentation of some additional attributes. As a compromise, we suggest an information security cost ratio of approximately 25 % for this layer.

The second layer is called "Operational Measures"; it consists of what one would typically call information security measures: measures like virus-protection

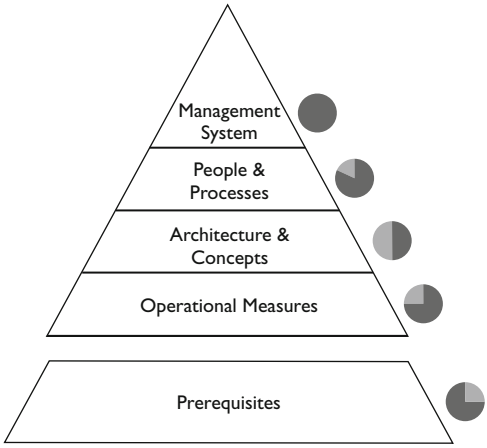

Fig. 1.3 The ISMS-layers approach

or encryption software. Mostly, measures of this layer are fully and directly accountable to information security. The ratio needed to determine the costs of information security highly depends on the specific information security control. In the case of a firewall that also acts as an e-mail gateway, this ratio can be set to a middle to high value. Thus, we suggest a security cost ratio of 75 %.

The third layer of this approach is called "Architecture & Concepts". It consists of concepts like data protection, data leakage prevention, reporting, IDS or providing security in an enterprise's products, enhanced by architecture tasks like encryption, Public Key Infrastructure (PKI) or Identity Management (IdM). In this layer, the line between information security tasks and other tasks with – maybe only partly – security relevant aspects is much harder to draw. On the one hand, an identity management infrastructure can clearly reduce information security risks through approval processes and reporting features. On the other hand, by enabling automated provisioning of access rights it also helps to reduce IT-operations costs. The determination of the information security cost ratio in this layer is quite tricky. However, we suggest a cost ratio of approximately 50 %.

The next layer is called "People & Processes". It deals with security aspects related to people and processes, such as awareness and security related training, but also the development and implementation of security controls, policies and guidelines. In general, the optimization or adaption of existing processes and services may in some cases have an influence on a process that is easily measurable (e.g., cycle time), in other cases the influence may not be directly measurable (e.g., long-term awareness campaigns, training activities, changes in processes). Adding up, a very high percentage of the mentioned costs can be accounted to information security, respectively we suggest a value of 80–85 %.

The top layer of this approach is called "Management System". This layer includes the ISMS including risk management, audits, but also costs for information security related audits, and costs for ISMS software. In our opinion, the costs for ISM, security audits and risk management actions are 100 % costs of

Table 1.4 Adequacy of the approaches for categorizing costs of information security

	App. 1	App. 2	App. 3	App. 4	App. 5
Focus single measure	0	+	0	−	−
Focus whole organization	0	−	0	+	+
IT-security centric	+	+	+	0	−
Information security centric	0	−	0	+	+
Benchmarking	0	−	−	+	+
Cost-benefit-analysis	0	+	0	−	−
Comparing measures	0	+	0	0	−
Compatibility with existing data sources	+	+	0	−	−
Differentiation by determinability	0	−	0	+	+
Differentiation by cost ratio	0	−	0	+	+

$+ \mathrel{\hat{=}}$ appropriate; $0 \mathrel{\hat{=}}$ partially appropriate; $- \mathrel{\hat{=}}$ inappropriate

information security. ISMS software are usually highly specialized solutions that do not have any other purpose than supporting the ISMS. There may be some parts where measures of this layer can be of use or beneficial for other departments or business functions. Mostly, those should be negligible, thus a security cost ratio of close to 100 % is suggested.

In general, this approach is best-suited for benchmarking and for research regarding the performance of different strategies in an organization's security management. The advantage of this approach is the fact that areas with a high information security cost ratio are separated from areas with a low one. In practice, this supports comparisons well.

1.5 Discussion

A cost model is an important step towards a common understanding and comparability of costs of information security. Table 1.4 summarizes the approaches presented in this chapter with regard to their adequacy for different purposes. However, the analysis above has shown that one single model will not satisfy all information needs. Therefore, combined models are needed and future research on the practicability of cost models, as well as on the nature of costs of information security is required. Those aspects are briefly discussed in this concluding section.

1.5.1 Conclusion

In our opinion, it could be useful to combine two – or possibly even more – of the approaches that have been presented in this chapter with other ways of classification. It is conceivable that, for example, a combination with the four aspects

personnel, investment (hardware, software), maintenance and outsourcing/MSS, or also possibly with the four aspects mentioned by Gartner in Sect. 1.4.1, would lead to an improved understanding, comparability and ease of use.

The results of cost-benefit analyses especially can vary widely depending on the related measure. The measure itself can still be classified, for example, with the help of the approach, mentioned in Sect. 1.4.6, but this is simply not enough for a cost-benefit analysis. After the classification, the costs of the measure need to be further broken down. In the case of an operational measure, the application of a classical TCO approach could already be sufficient. In the case of architectural topics like IdM or PKI this is much harder since only parts of the overall costs can be accounted to the acquisition or operation of the security measure. Other aspects like the adaption of internal processes play a much bigger role here and the measure itself might not even be seen as an information security measure. Even in this case the complexity of ISM increases; tests, audits, but also concepts for the introduction of these measures are not seen as measures themselves, but still additional costs may occur. Similar problems may occur also on other layers.

After application of the combination, one would achieve an overview in the form of a matrix, where one dimension would describe the information security costs according to the design of an ISMS (cf. Sect. 1.4.6), or according to the classification of costs according to the ISO27001 standard (cf. Sect. 1.4.5). The other dimension would describe the aforementioned four aspects personnel, investment, maintenance and outsourcing. This would lead to a further breakdown of the costs of the several layers or measures. Detailed research of these combinations should be the subject of future work towards a cost model for the costs of information security.

This section shows several approaches for tackling the aim of being able to determine the costs of information security and achieve comparability. The advantages and the aim of the presented approaches have been identified and visualized. Depending on the aim of its use, another perspective, and thus also another cost model may be relevant for an organization:

- For budget planning, the most relevant costs for an organization will most likely be expenses that have to be paid to externals, e.g., managed services, costs for hardware and software, license costs, consulting (cf. Fig. 1.2, Sect. 1.4.1).
- For the evaluation of a project or the implementation of a certain measure, basically all costs are highly relevant, i.e., TCO.
- For the benchmarking of IT costs, the most relevant costs will probably be costs for operation, maintenance, etc., while costs for the introduction or implementation may be less relevant.
- For the benchmarking of costs of information security, a differentiated view depending on the build-up of an ISMS may be the most relevant for an organization. This would provide the possibility to determine whether an organization, for example, spends too much money on management tasks, or too little on baseline architecture projects (cf. Sects. 1.4.5 and 1.4.6).

This list represents only a few examples. Section 1.3 shows that the topic of cost models in the field of information security has mostly been left out of research

so far. The identification of additional purposes for using cost models may be part of future research work.

1.5.2 Directions for Future Research

1.5.2.1 Empirical Evaluation of the Results

The development of a process for determining security costs in a consistent way is a major prerequisite for an empirical evaluation of the costs of information security. Benchmarking can only be successful if the method for measuring the values for the different cost categories is easy to repeat and described clearly.

In this chapter we suggested new approaches to determine and classify costs for information security or for certain information security measures. The next step would be to research the results of the presented approaches in practice. An empirical study among Chief Information Officers (CIOs) or CISOs could either prove or disprove the information security cost ratios provisorily set, and also give an idea of whether the presented approaches will really work in practice or not.

1.5.2.2 Determination of Information Security Cost Ratios and Determinability

Several possible classifications and ways to determine information security costs were presented in Sect. 1.4. Some of these approaches use cost ratios to ease the determination of costs. The ratios that have been used in this chapter derive from the practical experience of the authors. In general, this means that those values are only estimations to present the idea of the real information security cost ratio. The used values for several controls, measures or ISMS layers cannot be seen as proven values. The detailed and scientific determination of these exact values should be the subject of further research.

In general, the same applies to determinability. By using a more detailed and systematic research method, the suggested values for the determinability of the costs that are accountable to information security could be verified. Regarding this issue, conducting a survey among CISOs or CIOs could be an appropriate method to evaluate the provisional values set in this chapter.

1.5.2.3 Determination and Evaluation of Possible Combinations

As previously mentioned in Sect. 1.5.1, we propose an evaluation of different combinations of the approaches presented. A combination of two or more dimensions could improve the comparability of results between different organizations. The resulting matrix would help to further break down and analyze the costs

of information security. For researchers this could be an important step towards analyzing the effectiveness of information security expenditures. Even if the overall budget in two organizations is almost equal, its distribution over the matrix may be entirely different.

1.5.2.4 Reference Parameter

We have identified several problems and discrepancies of surveys (cf. Sect. 1.3). One of the identified problems is the appropriate baseline. In the case of information security, the IT budget of an organization is often used as the baseline in the literature. As shown in this chapter, using this value is not applicable for information security costs as only parts of information security measures and costs can be seen as IT measures, and thus seen as IT costs. The turnover of an enterprise is also often used as a reference parameter. A detailed inquiry regarding the significance, availability and accuracy of reference parameters in the field of information security and its costs may produce interesting results that could be used in future research.

1.5.2.5 Identification of Differences Between Several Industries

Following a practical test of the approaches suggested in this chapter, another topic for future research could be the possible differences between several industries. A software development company will encounter much higher costs for secure programming than, for example, a consulting company. On the other hand, a company that mostly focuses on mechanical engineering faces lower costs for security in products than a company that develops wireless access points or other network components. Extensions or adaptions of the cost model to make it more suitable for different industries can be a topic for further research. Especially companies that do not only apply information security measures to protect their own information assets, but also implement security in their products, require a further differentiation of costs for security. This differentiation between "conventional" products (e.g., cars), costs for security products (e.g., smartcards), and costs for measures to secure an enterprise's information, data or internal systems (this also includes ISM) seems to be essential.

References

1. Amoroso, E.: Hearing before the US Senate Commerce, Science, and Transportation Committee. Senate Hearing, pp. 111–143. U.S. Senate Committee on Commerce, Science, and Transportation (2009). http://www.commerce.senate.gov/public/index.cfm?p=Hearings&ContentRecord_id=d59f00d0-0ad9-41cd-bde8-b96babb08b7e&ContentType_id=14f995b9-dfa5-407a-9d35-56cc7152a7ed&Group_id=b06c39af-e033-4cba-9221-de668ca1978a&YearDisplay=2009

2. Anderson, R.: Why information security is hard – an economic perspective. In: ACSAC'01: Proceedings of the 17th Annual Computer Security Applications Conference, New Orleans, pp. 358–365. IEEE Computer Society (2001)
3. Barthélemy, J.: The hidden costs of IT outsourcing. Sloan Manage. Rev. **42**(3), 60–69 (2001)
4. Berinato, S.: Finally, a Real Return on Security Spending. CIO Magazine (2002). Available Online: http://www.cio.com.au/article/52650/finally_real_return_security_spending/
5. Capgemini: IT-Trends (2008)
6. Cavusoglu, H., Mishra, B., Raghunathan, S.: A model for evaluating IT security investments. Commun. ACM **47**(7), 87–92 (2004)
7. Commission, F.T.: Identity theft survey report. http://www.ftc.gov/os/2003/09/synovatereport.pdf (2003)
8. Commission, F.T.: 2006 identity theft survey report. www.ftc.gov/os/2007/11/SynovateFinalReportIDTheft2006.pdf (2007). Accessed 20 Sep 2012
9. Faisst, U., Prokein, O., Wegmann, N.: Ein Modell zur dynamischen Investitionsrechnung von IT-Sicherheitsmaßnahmen. Zeitschrift für Betriebswirtschaft **77**, 511–538 (2007)
10. Feigenbaum, A.: Total quality control. Harv. Bus. Rev. **34**, 93–101 (1956)
11. Florêncio, D., Herley, C.: Sex, lies and cyber-crime surveys. In: Ed: Bruce Schneier (ed.) Economics of Information Security and Privacy III. Springer, New York (2013). http://link.springer.com/book/10.1007/978-1-4614-1981-5zitieren?
12. Gartner: Distributed computing – chart of accounts. http://www.arsys-europe.net/Propalms/Datasheets/Propalms_WhitePaper_Gartner_TCO_Analyse_for_Distributed_Computer.pdf (2003). Accessed 20 Sep 2012
13. Gartner: IT budget: information security & risk management spend metrics. http://www.gartner.com/technology/metrics/it-security-risk-spending.jsp (2011). Accessed 20 Sep 2012
14. Gordon, L., Loeb, M.: The economics of information security investment. ACM Trans. Inf. Sys. Secur. (TISSEC) **5**(4), 438–457 (2002)
15. Gordon, L., Loeb, M.: Managing Cybersecurity Resources: A Cost-Benefit Analysis, 1st edn. McGraw-Hill, New York (2005)
16. Holthaus, M.: Management der Informationssicherheit in Unternehmen. PhD thesis, Universität Zürich (2000)
17. Hoo, K.J.S.: How much is enough? A risk management approach to computer security. PhD thesis, Stanford University (2000)
18. Humpert-Vrielink, F., Vrielink, N.: Ganzheitliches sicherheitskosten-controlling. http://www.kes.info/archiv/online/kostencontrolling.html (2011). Accessed 20 Sep 2012
19. ISO: ISO/IEC 27001:2005 Information Technology – Security Techniques – Information Security Management Systems – Requirements (2005)
20. Kendrick, S.: The morphing IT security landscape. https://vishnu.fhcrc.org/security-seminar/IT-Security-Landscape-Morphs.pdf (2010). Accessed 20 Sep 2012
21. Kovacich, G., Halibozek, E.: Security Metrics Management: How to Manage the Costs of an Assets Protection Program. Butterworth-Heinemann, Oxford (2006)
22. Kütz, M.: Controlling der Information Security, 19th edn. TÜV Media – Dieter Burgartz and Ralf Röhrig, chap. 03710. No. 32. Aktualisierung September 2011 in Praxiswissen IT-Sicherheit: Praxishandbuch für Aufbau, Zertifizierung und Betrieb (2011)
23. Langfield-Smith, K., Smith, D.: Managing the IS outsourcing relationship. In: Rivard, S., Aubert, B.A. (eds.) Advances in Managing Information Systems. Information System Outsourcing, chap. 10, pp. 163–188. M.E. Sharpe, Armonk (2008)
24. Locher, C.: Ein Steuerungsmodell für das Management von IV-Sicherheitsrisiken bei Kreditinstituten. In: Ferstl, O.K., Sinz, E.J., Eckert, S., Isselhorst, T. (eds.) Wirtschaftsinformatik, pp. 1207–1225. Physica-Verlag, Heidelberg (2005)
25. Longstaff, T., Chittister, C., Pethia, R., Haimes, Y.: Are we forgetting the risk of information technology. IEEE Comput. **33**(12), 43–51 (2000)
26. Lubich, H.P.: IT-Sicherheit: Systematik, aktuelle Probleme und Kosten-Nutzen-Betrachtung. HMD, Praxis der Wirtschaftsinformatik **43**(248), 6–15 (2006)
27. Mercuri, R.T.: Analyzing security costs. Commun. ACM **46**(6), 15–18 (2003)

28. New Scientist: Cybercrime toll threatens new financial crisis. http://www.newscientist.com/article/dn16092-cybercrime-toll-threatens-new-financial-crisis.html (2008). Accessed 04 June 2012
29. NIST – National Institute of Standards and Technology: Risk Management Guide for Information Technology Systems. NIST Special Publication 800–30 (2004)
30. Nowey, T.: Konzeption eines Systems zur überbetrieblichen Sammlung und Nutzung von quantitativen Daten über Informationssicherheitsvorfälle. PhD thesis, Universität Regensburg (2010)
31. Penn, J.: The State of Enterprise IT Security: 2008 to 2009 (2009). http://www.forrester.com/The+State+Of+Enterprise+IT+Security+2008+To+2009/fulltext/-/E-RES47857
32. Pohlmann, N.: Wie wirtschaftlich sind IT-Sicherheitsmaßnahmen. HMD, Praxis der Wirtschaftsinformatik **43**(248), 26–34 (2006)
33. Schaffry, A.: Die IT-Sicherheitsausgaben bis 2015. http://www.cio.de/knowledgecenter/security/2294879/index.html?r=2616952702416512&lid=152021 (2011). Accessed 20 Sep 2012
34. Schiffauerova, A., Thomson, V.: A review of research on cost of quality models and best practices. Int. J. Qual. Reliab. Manage. **23**, 647–669 (2006)
35. Scholtz, T.: Articulating the business value of information security. Tech. rep., Gartner Inc. (2011)
36. SSG Inc: Cyber crime – the facts. http://www.ssg-inc.net/cyber_crime/cyber_crime.html (2012). Accessed 20 Sep 2012
37. Sullivan, T.: The surprisingly small percentage health orgs spend on data security. http://govhealthit.com/news/surprisinlgy-small-percentage-health-orgs-spend-data-security (2011). Accessed 20 Sep 2012
38. Weigelt, M.: Security could consume 10 percent of IT budget. http://fcw.com/articles/2008/02/07/security-could-consume-10-percent-of-it-budget.aspx (2008). Accessed 20 Sep 2012

Chapter 2
To Invest or Not to Invest? Assessing the Economic Viability of a Policy and Security Configuration Management Tool

Lukas Demetz and Daniel Bachlechner

Abstract The threat of information security (IS) breaches is omnipresent. Large organizations such as Sony or Lockheed Martin were recently attacked and lost confidential customer information. Besides targeted attacks, virus and malware infections, lost or stolen laptops and mobile devices, or the abuse of the organizational IT through employees, to name but a few, also put the security of assets in jeopardy. To defend against IS threats, organizations invest in IS countermeasures preventing, or, at least, reducing the probability and the impact of IS breaches. As IS budgets are constrained and the number of assets to be protected is large, IS investments need to be deliberately evaluated. Several approaches for the evaluation of IS investments are presented in the literature. In this chapter, we identify, compare, and evaluate such approaches using the example of a policy and security configuration management tool. Such a tool is expected to reduce the costs of organizational policy and security configuration management and to increase the trustworthiness of organizations. It was found that none of the analyzed approaches can be used without reservation for the assessment of the economic viability of the policy and security configuration management tool used as an example. We see, however, considerable potential for new approaches combining different elements of existing approaches.

2.1 Introduction

The perils of information security (IS) breaches are ubiquitous. In 2011, large companies were subject to attacks and IS breaches were discussed in public (e.g., [11, 12, 21]). Besides attacks, other reasons, for instance, virus and malware infections,

L. Demetz (✉) · D. Bachlechner
Department of Information Systems, School of Management, University of Innsbruck, Innsbruck, Austria
e-mail: lukas.demetz@uibk.ac.at; daniel.bachlechner@uibk.ac.at

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_2,

lost or stolen mobile devices, human errors, and forces of nature [14, 15, 26, 43] urge organizations to invest in IS. The question today is not whether an organization will face an IS breach, but rather when a breach will occur [20]. As a result, organizations invest in countermeasures that prevent IS breaches or reduce their probability and impact.

Facing constrained budgets and an increasing number of assets to protect, organizations have to decide how much to invest in IS, and how to allocate the IS budget [2, 19]. As the probability of being exploited as well as the criticality differ from asset to asset, not all assets should receive the same level of attention [3].

Investments in IS, unlike other investments, do not generate monetary returns but result in cost savings by preventing IS breaches or by reducing the probability of their occurrence and their impact [23,32,34,37]. Analyzing the cost-benefit tradeoffs of alternative IS investments, however, is challenging [9], as not only the costs but also the benefits of IS investments with respect to known and unknown threats have to be assessed [16]. Organizations, however, need to pay attention to the economic viability of IS investments. They have to find the balance between the risks of threats on one side and the possibility to mitigate the risks and the costs thereof on the other side [8]. Mizzi [29, p. 19] defines an IS investment as economically viable if

$$E_S < L_T$$

where E_S represents security expenditures and L_T the total annual losses. That is, an IS investment is economically viable if and only if the security expenditures are smaller than the total annual losses. Mizzi [29], however, assumes that the security expenditures E_S aim to fix all vulnerabilities to assets at stake (i.e., to completely remove these vulnerabilities). Thus, security expenditures E_S, which include the costs to build IS countermeasures and the costs to fix vulnerabilities, clearly need to be lower than the sum of the losses expected from an IS breach of the vulnerabilities and the costs to repair breached assets. In the case of investments that only aim to fix a certain vulnerability (i.e., not all vulnerabilities) to an asset, the equation by Mizzi [29] does not hold. In such cases, the security expenditures need to be lower than the *reduction* of expected losses conditional the investment to fix the respective vulnerability. Similar to Mizzi, Huang et al. [23] argue that for risk-averse decision makers expenditures for IS investments increase with, however, never exceed the expected losses associated with IS threats. Gordon and Loeb [17] even argue that the optimal amount to invest in IS never exceeds 37 % of the expected losses associated with an IS breach. Willemson [44], however, shows that in some cases expenditures of nearly 100 % of the expected losses can be reasonable.

Fortunately, the literature provides a myriad of approaches (e.g., [6, 10, 20, 22]) that help decision makers in deciding whether or not to invest in a certain IS countermeasure. Among the most frequently cited approaches is the one presented by Gordon and Loeb [17] for which also several extensions have been proposed (e.g., [28, 45]).

In this chapter, we identify approaches which are suitable to assess the economic viability of a specific countermeasure, namely, a policy and security configuration

management tool. Such a tool helps, first, to reduce the costs associated with policy and security configuration management and, second, to increase the trustworthiness of an organization by automating or providing decision support for critical activities related to policy and security configuration management. We describe and compare selected approaches, evaluate them with respect to their suitability for assessing the economic viability of such a tool based on a set of criteria, and discuss the approaches' advantages and disadvantages.

The remainder of this chapter is structured as follows: Sect. 2.2 introduces the policy and security configuration management tool which is the subject of the investment decision. Section 2.3 is devoted to the research methods used to collect and analyze data about approaches. We present the results of the analysis in Sect. 2.4, where we also outline the determining characteristics of the different approaches. In Sect. 2.5, we discuss the approaches with respect to their suitability for assessing the economic viability of the policy and security and configuration management tool, and highlight commonalities and differences of the approaches. Finally, Sect. 2.6 concludes this chapter and gives a short outlook on possible future work.

2.2 Policy and Security Configuration Management

Today, organizations are confronted with an increasing number of regulatory (e.g., SOX or PCI-DSS) and contractual requirements they need to comply with. As a result, they have to increase their expenditures on compliance activities [31]. This situation is particularly exacerbated for service providers offering services to clients as they are faced with a myriad of additional contractual requirements requested by their clients. Management costs, including costs for policy and security configuration management, steadily increased over recent years [25]. Currently, policy and security configuration management is mainly done manually, which often turns out to render related activities inefficient and error prone [30]. In this respect, a policy is a declarative description of an outcome. A security policy comprises rules that specify how security is established and maintained [33]. Each security policy is associated with at least one security configuration that describes imperatively how the respective goal is to be reached. While inefficiencies often lead to unnecessary high costs, a lack of trust is often the consequence of error-proneness. Disrespecting or ignoring security policies may be the causes for many IS breaches [39].

To deal with this myriad of requirements, organizations in general and service providers in particular could benefit from a tool supporting them in policy and security configuration management. Such a tool establishes and maintains a consistent and transparent link between high-level security and compliance requirements at one end and low-level technical configurations of IT landscape components on the other. This end-to-end link is maintained automatically where possible, and, in case human interaction is necessary, decision support is offered. The aim of such a tool is two-fold: reducing costs (e.g., management costs and losses due to IS breaches) and increasing an organization's trustworthiness by increasing its level of security

and compliance. Both goals may be achieved by partially or fully automating activities related to policy and security configuration management such as detecting misconfigurations or checking whether different security countermeasures are equivalent with respect to security level, performance and costs. Additionally, the tool would ease audits as the information necessary for audits can be provided directly by the tool.

The policy and security configuration management tool would support two different modes of operation. The first mode is a static mode, in which the end-to-end link between security and compliance requirements and configurations is planned and initially established. Additionally, the tool can be operated in a dynamic mode, in which configurations are constantly monitored for deviations from the ideal configuration. Such an automated monitoring allows organizations to detect misconfigurations quicker and thus to reduce the risk of IS breaches or of problems caused by non-compliance. As the tool's functions would be tightly coupled, we assume that the tool is available only as a whole, that is, there are no modules which could be added at a later point in time.

Ideally, the tool is run not only at one organization, but also at its suppliers and clients. This way, each involved party could easily share information about requirements and configurations. As a result, each party is able to assess the fulfillment of requirements at its suppliers and to also assess whether certain requirements can be fulfilled by a supplier. Operating such a tool across several parties in a cross-organizational setting would increase the benefits of the tool for all parties involved.

Such a policy and security configuration management tool would certainly have its advantages. Nevertheless, the decision to invest in such a tool must be well justified, for instance, by applying an approach for assessing investment decisions found in the literature. Based on the policy and security configuration management tool's characteristics and its application in cross-organizational settings, we derive a set of mandatory and optional criteria a suitable approach must or should meet, respectively. The derived criteria are:

1. *The approach must be able to deal with investments made as a whole.* As we assume that the tool is only available as a whole and that there are no modules that could be added at a later point in time, a suitable approach must support decisions regarding investments made as a whole.
2. *The approach must be able to consider financial measures.* As the tool aims, among other things, at reducing the costs of policy and security configuration management, a suitable approach must support financial measures.
3. *The approach should be able to consider non-financial measures.* As the tool also aims at increasing the trustworthiness of an organization and since increased trustworthiness cannot be easily expressed in financial measures, a suitable approach should support non-financial measures.
4. *The approach must be able to support one-time costs and benefits.* As costs crucial for decision making incur immediately whenever the tool is used for planning and initially establishing the end-to-end-link between security and

compliance requirements and configurations, a suitable approach must support such one-time costs and benefits.
5. *The approach should be able to support running costs and benefits.* As the tool is also operated in a dynamic mode and since costs and benefits thus also incur over time, a suitable approach should support running costs and benefits.
6. *The approach must be applicable without explicitly considering attacks.* Some approaches rely on the provision of information on a particular attack. As the tool's primary focus is on policy and security configuration management and information on attacks is generally neither relevant nor available, a suitable approach must be applicable without considering attacks.
7. *The approach should be able to consider network effects of investments.* The more organizations are involved in a cross-organizational setting, the higher are the benefits of the tool for all parties involved. Thus, network effects should be supported by a suitable approach.

We chose a policy and security configuration management tool as the subject of the investment decision because of the tool's broad relevance for organizations in general and service providers in particular, and our insight into the unique characteristics of such a tool resulting from prior research. Assessing the economic viability of another IS investment would certainly lead to other criteria to be met by suitable approaches.

2.3 Data Collection and Analysis

In this section, we describe the methods used to collect relevant articles and to select approaches for assessing the economic viability of IS investments. Subsequently, the detailed analysis of the selected approaches is outlined.

2.3.1 Collection of Approaches

We started collecting approaches described in the literature with an unsystematic search using Google Scholar. In this step, we identified 30 relevant articles discussing IS investments. We extracted their keywords, combined them under more general terms, and ranked the terms with respect to their frequency of appearance.

Subsequently, we used the two most frequent terms – *economics of security* and *security investment* – for a systematic search, again using Google Scholar. For both terms, we looked for peer-reviewed articles with matching titles and abstracts within the first 200 search results and created a collection of articles. Since the term *security* has different connotations in other domains, and for the sake of completeness, we additionally queried Google Scholar with variations of the terms. More concretely, we replaced *security* with *information security*, *computer security* and *IT security*

in both terms. Search queries using these variations, however, did not result in additional articles. Apart from that, as suggested by Webster and Watson [42], we examined the articles referenced by the already collected articles. The entire collection process resulted in 83 articles focusing on IS investments.

In the next step, we discarded articles that did not focus on approaches for supporting IS investment decision making. For instance, articles dealing with empirical analyses of IS investments (e.g., [18, 27]) were discarded. Furthermore, we excluded articles discussing approaches that help to optimally allocate a fixed budget. Additionally, articles that present an overview of several approaches (e.g., [34, 38]) were discarded. Substantial extensions to existing approaches (e.g., [28] extends [17]) were treated as individual approaches. Approaches tailored to specific countermeasures incomparable with the policy and security configuration management tool were removed. Cavusoglu et al. [10], for instance, present an approach to determine the value of intrusion detection systems and was thus not considered for detailed analysis. In case an approach was described in several articles by the same author or group of authors, newer publications were favored over older, and journal articles over articles in conference proceedings. We made sure that the newer articles did not only extend the older ones. In the case of extensions, both articles were treated separately. Eleven approaches for assessing IS investments, each described in an individual article, were finally considered for detailed analysis.

2.3.2 Analysis of Approaches

First, for each approach, the corresponding article was read carefully. While reading, information regarding the criteria introduced in Sect. 2.2 was marked and extracted. For the identification of relevant information, the descriptions of the approaches' procedures proved to be particularly valuable. For instance, for information regarding financial and non-financial measures, we looked primarily at the approaches' input and output parameters. There, we analyzed whether they solely represent financial measures or also non-financial ones. We proceeded similarly to determine whether one-time and running costs and benefits are considered in the approaches.

2.4 Results

In the following, we present the analyzed approaches in alphabetical order. For each approach, we first give a short description of the approach and then show to what extent it meets the criteria presented in Sect. 2.2. Table 2.1 lists the analyzed approaches and the degree to which they meet the criteria. Each criterion is represented by a dedicated column. A checkmark (✓) indicates that a criterion

Table 2.1 Overview of the approaches for IS investment decisions analyzed in detail

Approach presented by	Made as a whole	Financial	Non-financial[a]	One-time costs	Running costs[a]	Attacks	Network effects[a]
Al-Humaigani and Dunn [1]	✓	✓		✓		✓	
Bodin et al. [4]	✓	✓	✓	✓	∼	✓	∼
Butler [7]	✓	✓	✓	✓		✓	
Cremonini and Martini [13]	✓	✓		✓			
Gordon and Loeb [17]	✓	✓		✓		✓	
Gordon et al. [20]	✓	✓		✓		✓	
Huang et al. [23]	✓	✓		✓		✓	
Mizzi [29]	✓	✓		✓	∼	∼	
Sonnenreich et al. [35]	✓	✓		✓		✓	
Tallau et al. [37]	✓	✓	✓	✓	∼	✓	∼
Wang et al. [40]	✓	✓		✓		✓	

✓ Criterion is met completely; ∼ Criterion is met partially

[a] Criterion is optional

is met completely, whereas a tilde (∼) indicates that a criterion is met partially. An empty cell denotes that a criterion is not met or nothing is mentioned in the respective article. In the column labeled "Attacks", a checkmark indicates that the corresponding approach does not rely on information on attacks. Columns marked with a letter "a" denote optional criteria.

In contrast to other surveys on approaches for the assessment of IS investments (e.g., [5, 36]), this chapter objective is not to present an overview of all approaches found in the literature. Rather, the objective of this chapter is to collect and analyze approaches that are suitable for determining the economic viability of a specific IS investment, namely, of a policy and security configuration management tool as presented in Sect. 2.2. As the total number of approaches presented in the literature would be too exhaustive for a detailed analysis, we reduced the number of approaches as described in Sect. 2.3. Accordingly, the approaches analyzed in detail (i.e., the results of this chapter) represent only a number of approaches found in literature. The analyzed approaches, however, are those considered most suitable for assessing the economic viability of the policy and security configuration management tool.

2.4.1 Approach of Al-Humaigani and Dunn

A rather simple approach for assessing IS investments is presented by Al-Humaigani and Dunn [1]. They argue that the maximum return of an IS investment is reached

when the total costs of security, including losses due to IS breaches and costs of countermeasures, is minimal. Al-Humaigani and Dunn use measures representing expenditures for the investment and costs incurring if no investment is made.

In their approach, Al-Humaigani and Dunn calculate the return on investments based solely on financial measures. In their calculations, however, they use financial measures for non-financial aspects, for instance, losses in reputation and goodwill. Al-Humaigani and Dunn determine the return on security investment *(ROSI)* using the following equation:

$$ROSI = \sum [K_T \times (C_{T6} + C_{T7} + C_{T8} + C_{T9} + C_{T10}) + C_{T11} - (C_{T1} + C_{T2} + C_{T3} + C_{T4} + C_{T5})]$$

where T is the threat or risk the IS investment is intended to address; C_{T1} denotes costs of procuring the countermeasures, C_{T2} costs of additional hardware and facilities, C_{T3} costs of training, C_{T4} losses due to limitations placed on business, C_{T5} costs of adopting a secured-by-design strategy, C_{T6} costs to recover from an IS breach, C_{T7} losses due to business interruption, C_{T8} losses in human casualties, C_{T9} losses in data from business and legal aspects, C_{T10} losses in reputation and goodwill, C_{T11} the amount paid by the insurance and K_T the probability of the realization of the threat without the investment.

Just like all the other approaches investigated in detail, the approach proposed by Al-Humaigani and Dunn is not only able to deal with investments made as a whole but also to consider financial measures. With respect to financial measures, the approach incorporates 11 predefined costs to determine the $ROSI$. The approach is, however, not able to consider non-financial measures. While the approach supports one-time costs and benefits, it is not per se able to support running costs and benefits. However, running costs and benefits may be discounted to their present value and added to the one-time measures. The approach proposed by Al-Humaigani and Dunn is applicable without explicitly considering attacks. Like most of the other approaches, the approach is not per se able to consider network effects of investments. However, network effects may be taken into account by users of the approach when specifying the measures used. To sum up, the approach meets the four mandatory criteria but does not meet any of the optional ones.

2.4.2 Approach by Bodin et al.

Bodin et al. [4] present an approach based on the analytic hierarchy process *(AHP)*. The *AHP* uses besides financial measures also non-financial measures for analyzing multi-criteria decision problems. The approach by Bodin et al. is predominantly used in comparative analyses, where several investment alternatives are compared with each other.

This approach starts with the determination of criteria and sub-criteria along with intensity levels denoting the level of fulfillment (e.g., high or very high). On the basis of these criteria and sub-criteria, IS investments are compared. Therefore, weights $C(i, j)$ for a pairwise comparison are assigned to the criteria, sub-criteria and intensity levels. The larger a weight $C(i, j)$, the more important is (sub-)criteria i over j. Then, each alternative is evaluated with respect to the criteria and sub-criteria. Simultaneously, the corresponding intensity levels are recorded. Finally, for each alternative, the weights of all criteria and sub-criteria are summed up resulting in the alternatives' total scores. The alternative yielding the highest total score is recommended.

Just as all other approaches analyzed, the approach presented by Bodin et al. is able to handle investments made as a whole. The approach allows the decision maker to choose the measures to be used for decision support. As such, the approach is able to support financial and non-financial measures as well as measures for one-time and running costs and benefits. The approach is applicable without explicitly considering attacks. Even though network effects are not proposed as measures, the approach is able to support measures for network effects of investments. To sum up, the approach meets all four mandatory criteria. Of the three optional criteria, two are met and one is partially met.

2.4.3 Approach by Butler

The comparative approach described by Butler [7] is called the Security Attribute Evaluation Method *(SAEM).* This approach is a quantitative cost-benefit analysis for IS investment decisions comprising four steps. For the initial data collection, structured interviews with information technology and IS managers are conducted.

The first step of the analysis is an IS technology benefit assessment. In this step, several investment alternatives are collected and their benefits are assessed. Subsequently, each alternative is evaluated with respect to its capability to mitigate IS risks. That is, an alternative's effectiveness in reducing the probability and impact of an IS breach is assessed. These estimations are done by IS managers who rate the effectiveness based on their working experience. In the following step, an IS architecture coverage assessment is conducted. Here, each alternative is assessed with respect to the breadth of IS risks the alternative covers. In the final step, the costs of each alternative are compared with each other.

Similar to the other approaches we analyzed, the approach proposed by Butler is able to deal with investments that are made as a whole. The approach is able to support financial as well as non-financial measures. With respect to these measures, one-time costs and benefits are supported, while running costs and benefits are not per se supported. The approach is applicable without explicitly considering attacks. Just as most other approaches, the approach proposed by Butler does not per se support network effects of investments. These, however, may be considered while

specifying the measures used. To sum up, the approach meets the four mandatory criteria and all but one of the optional ones.

2.4.4 Approach by Cremonini and Martini

Cremonini and Martini [13] discuss an approach to IS investment decision making which is similar to that of Sonnenreich et al. [35]. They also use a return on investment *(ROI)* based approach using the annual loss expectancy *ALE*. Additionally, they couple *ROI* with a measure referred to as return on attack *(ROA)* representing the convenience of attacks. *ROA* comprises costs faced by an attacker willing to breach a system. This allows us to compare alternatives from an attacker's point of view and to choose the alternative with the highest disadvantage for an attacker.

Cremonini and Martini define *ROI* as

$$ROI = \frac{ALE_{\text{before}S} - ALE_{\text{after}S}}{\text{costs of security measure } S},$$

where $ALE_{\text{before}S}$ and $ALE_{\text{after}S}$ denote the annual costs related to all IS incidents that security countermeasure S is destined to mitigate, before and after S was implemented, respectively. *ROA*, on the other hand, is equal to

$$ROA = \frac{\text{gain from successful attack}}{\text{costs before } S + \text{losses caused by } S}.$$

The approach proposed by Cremonini and Martini is able to deal with investments made as a whole. The approach considers three financial measures to determine the *ROI*. Non-financial measures, however, are not taken into account by the approach proposed by Cremonini and Martini. While the approach is able to support one-time costs and benefits, running costs and benefits are per se not supported. Nevertheless, running costs and benefits may be discounted to their present value and considered in the determination of the *ROI*. Contrary to the other analyzed approaches, the approach proposed by Cremonini and Martini relies on information about attacks and is not easily applicable without such information. Similar to most other approaches investigated, also this approach does not account for network effects. To sum up, the approach meets three of four mandatory criteria and does not meet any of the optional ones.

2.4.5 Approach by Gordon and Loeb

Gordon and Loeb [17] present an approach for determining the optimal amount to invest to protect single assets. The authors assume a risk-neutral decision maker

and a one-period model (i.e., all decisions and outcomes occur instantaneously). Each asset is associated with monetary losses λ in case an IS breach occurs, a threat probability t and an inherent vulnerability v denoting the probability that without additional security an attack is successful. The expected losses L associated with an asset represent the product of the threat probability t and the monetary losses λ and are calculated as $L = t \times \lambda$. To reduce the vulnerability v of an asset, an organization invests $z > 0$ monetary units. In this respect, $S(z, v)$ represents an IS breach probability function denoting the probability that the asset with vulnerability v is compromised given the investment z to secure the asset.

The expected benefit from an IS investment z $EBIS(z)$ is calculated as

$$EBIS(z) = [v - S(z, v)]L;$$

the expected net benefit $ENBIS(z)$ reads

$$ENBIS(z) = EBIS(z) - z = [v - S(z, v)]L - z.$$

Just as all other approaches analyzed, also the approach proposed by Gordon and Loeb is able to deal with investments made as a whole as well as to consider financial measures. For determining the expected benefit of an IS investment, the approach takes two predefined costs coupled with probabilities into account. Non-financial measures, in contrast, are not considered by the approach. The approach is able to support one-time costs and benefits; running costs and benefits, however, are not per se supported. Running costs and benefits may, nevertheless, be discounted to their present value and considered within the determination of the expected benefits of the IS investment. The approach proposed by Gordon and Loeb is applicable without explicitly having information on attacks. Similar to most of the approaches analyzed, the approach does not per se consider network effects of investments. To sum up, while the approach proposed by Gordon and Loeb meets all four mandatory criteria, none of the optional ones is met.

2.4.6 Approach by Gordon et al.

Gordon et al. [20] present a wait-and-see approach based on real options. The basic idea of their approach is that in case of uncertainty regarding expected benefits, it may be better to wait for key events to occur. Often higher expected benefits can be yielded this way. Thus, before investing in IS, it may be advisable to wait for an IS breach to happen. As soon as an IS breach occurs, more information to assess the expected benefits of an IS investment is available, which makes the assessment more accurate.

Gordon et al. state that to make an investment, the net present value *(NPV)* of the investment made today must be greater than the *NPV* of the deferred investment. Determining the costs and benefits of an IS investment before an IS breach occurs

is, however, uncertain. For instance, Gordon et al. [20, pp. 3–4] provide an example of an organization about to make an investment of $1,000,000 in IS for 1 year. The benefits of this investment, are, however, uncertain. Either the benefits are $40,000 or $200,000 per month, both equally probable. Then, the expected value of the investment is equal to $(12 * \$40{,}000 * 0.5) + (12 * \$200{,}000 * 0.5) - \$1{,}000{,}000 = \$440{,}000$. They assume that 1 month later an IS breach occurs and the benefits of the investment become known. Now, the expected value for both savings can be determined: In the case of the lower benefits, the expected value of the investment is $EV_{low} = 11 * \$40{,}000 - \$1{,}000{,}000 = -\$560{,}000$, which is negative and the investment should not be made. When looking at the higher benefits, the expected value yields $EV_{high} = 11 * \$200{,}000 - \$1{,}000{,}000 = \$1{,}200{,}000$ making the investment economically viable. This example illustrates how the expected value of an IS investment increases from $440,000 to $\$1{,}200{,}000 * 0.5 = \$600{,}000$ by deferring the decision to invest by 1 month.

Just as all other approaches analyzed, the approach proposed by Gordon et al. is able to deal with investments made as a whole as well as with financial measures. Regarding financial measures, the approach uses two predefined costs coupled with probabilities for determining the economic viability of an IS investment. The approach, is, however, not able to support non-financial measures. While one-time costs and benefits are supported by the approach, running costs and benefits are not per se considered by the approach. These, however, may be discounted and added to one-time measures. The approach proposed by Gordon et al. is applicable without explicitly considering attacks. Like most other approaches analyzed, the approach is not able to support network effects per se. They, however, may be taken into account by considering them as financial measures. To sum up, the approach meets all four mandatory criteria but none of the optional ones.

2.4.7 Approach by Huang et al.

Huang et al. [23] present an approach for determining the optimal amount to invest in IS based on the investment's expected utility. As in the approach proposed by Gordon and Loeb [17], in this approach the level of investment also depends on the asset to be protected, its vulnerability, and the associated potential losses. In their approach, Huang et al. assume a single-event, single-period IS breach of an asset. An IS breach is associated with a probability function ρ and potential losses L including direct financial and indirect non-financial losses from, for instance, bad reputation. ρ is a function of the threat probability t external to the organization and determined by the attractiveness of the asset; the vulnerability v of the asset is determined by the configuration of the information system providing the asset; and the investment S in IS countermeasures to protect the asset. That is,

$$\rho = \rho(S, v, t).$$

The expected losses due to an IS breach is denoted by X with

$$X = \begin{cases} L, \rho, \\ 0, (1-\rho) \end{cases}$$

With respect to the calculation of the optimal amount to invest, Huang et al. assume that with increasing investment S the breach probability ρ decreases, and that the marginal improvement on security decreases with a higher investment S. They further assume a risk-averse decision maker, whose aim is to maximize the expected utility u, determined by the organization's wealth w. That is, $u = u(w)$. To determine the optimal amount to invest, the expected utility of the investment, written as

$$E[u(w - S - X)] = \rho u(w - S - L) + (1 - \rho)u(w - S)$$

needs to be maximized. To do so, the equation needs to be differentiated with respect to S and set equal to zero. Besides determining the optimal amount to invest, the approach by Huang et al. can also be used to calculate the upper bound of investments (i.e., the maximum amount to invest). Even for a risk-averse decision maker, the maximum amount to invest should never exceed the expected losses of a potential IS breach.

The approach by Huang et al. assumes that the investment is made as a whole. Cases in which the investment is partitioned into smaller parts are not considered. As inputs for supporting investment decisions, the approach uses one-time, financial measures. Non-financial measures, and running costs and benefits, however, are neglected. For decision support, the approach by Huang et al. does not rely on information on attacks. This allows us to apply the approach without considering attacks. Information on network effects of the investment are, however, not reflected by any of the approach's input parameters. To sum up, the approach meets all four mandatory criteria, but does not meet any of the optional ones.

2.4.8 Approach by Mizzi

Mizzi [29] presents an approach for IS investment decisions based on accounting figures. In his approach, Mizzi focuses solely on financial measures comprising the annual costs F to fix a vulnerability, the one-time costs B to implement a security countermeasure, and the annual maintenance costs M. To decide about an IS investment, the costs of the total annual IS expenditures E_S and the expected total annual losses L_T of a given security vulnerability are compared. More concretely, an investment should be made, if the expenditures are lower than the expected total annual losses, that is,

$$E_S < L_T \text{ with } E_S = F + B + M.$$

In subsequent years, B does not incur. Thus, the term is dropped from the equation. L_T can be calculated in several ways: One way is to account for the instantaneous losses L_I and the losses of asset I over t days of unavailability, that is

$$L_T = L_I + I * t/365;$$

the losses resulting from unavailability over t days may also be modeled as a function $A(t)$ making the total annual losses equal to

$$L_T = L_I + A(t);$$

additionally, the costs R to rebuild a compromised asset can also be taken into consideration as

$$L_T = L_I + A(t) + R$$

in case the rebuild costs do not include man-hour costs. Alternatively, if man-hour costs are the dominant rebuild costs, R can be substituted by $R(t)$.

If the approach is used as described by Mizzi, costs and benefits that incur over the course of time discounted to the present point in time are not considered. Mizzi, however, notes that one could additionally use *NPV* or internal rate of return *(IRR)* to better account for running costs and benefits. In contrast to other approaches presented, Mizzi presents an extension to his approach in which the costs to break a security countermeasures *CTB* for an attacker can be taken into account. In all calculations, however, this approach neither takes probabilities of IS breaches nor the success rate of IS countermeasures into consideration.

Just as all other approaches analyzed, the approach proposed by Mizzi is not only able to deal with investments made as a whole but also financial measures are supported. With respect to financial measures, the approach considers eight predefined costs to determine the economic viability of an IS investment. The approach, however, does not take non-financial measures into account. While the approach is able to deal with one-time costs and benefits, running costs and benefits are not taken into account. Mizzi, however, notes that these can be discounted to their present value and added to their one-time counterparts. Contrary to most other approaches analyzed, the approach presents an optional extension in which costs seen by the attacker are considered. The approach, nevertheless, is still applicable without explicitly considering attacks. Network effects are per se not considered by the approach, but may be taken into account when specifying the measures used. To sum up, the approach meets three of the four mandatory criteria. The remaining mandatory criterion and one of the optional criteria are partially met.

2.4.9 Approach by Sonnenreich et al.

Sonnenreich et al. [35] propose an approach similar to the traditional accounting figure return on investment *(ROI)* termed return on security investment *(ROSI)*.

In contrast to other approaches, Sonnenreich et al. do not split the costs used for the calculation further into different types of costs. For supporting investment decisions, they calculate *ROSI* as

$$ROSI = \frac{(\text{risk exposure} \times \text{risk mitigated}) - \text{solution costs}}{\text{solution costs}},$$

where

$$\text{risk exposure} = ALE = SLE \times ARO;$$

ALE denotes the annual loss exposure, that is, the single loss expose, *SLE*, times the annual rate of occurrence, *ARO*, of an IS breach the security investment should mitigate.

Just as all approaches analyzed, the approach proposed by Sonnenreich et al. is able to deal with investments made as a whole as well as with financial measures. In total, the approach considers five predefined costs to determine the *ROSI*. The approach, however, is not able to incorporate non-financial measures. For the determination of the *ROSI*, the approach incorporates one-time costs and benefits, while the approach is not per se able to take running costs and benefits into consideration. These, however, may be discounted to their present value and combined with one-time costs and benefits. Just as most other approaches, the approach presented by Sonnenreich et al. is applicable without explicitly considering attacks. Furthermore, the approach is not per se able to consider network effects of IS investments. To sum up, the approach meets all four mandatory criteria, but does not meet any of the optional ones.

2.4.10 Approach by Tallau et al.

Another approach to IS investment decision making is presented by Tallau et al. [37]. In contrast to the other approaches analyzed, Tallau et al. base their approach on the Balanced Scorecard proposed by Kaplan and Norton [24]. In general, the Balanced Scorecard is a performance measurement system that does not only consider financial measures, but also non-financial ones related to internal processes, customers, and innovation and learning. The Balanced Scorecard allows us to view business from four different angles, thus providing a balanced view of an organization's performance.

Tallau et al. use the perspectives as were used for the original Balanced Scorecard, *financial, customer, internal processes,* and *innovation and learning*, to support IT investment decisions. For each perspective, goals and measures for the investment are established. For instance, the authors use "Reduce hacks/intrusions in past year by 90 %" as a goal and "Server downtime (in hours)" as a measure in their exemplary application [37, p. 47]. Additionally, each goal is weighted indicating the

importance relative to the other goals. Next, the degree to which each goal is fulfilled is determined, the goals are weighted and the average of all weighted degrees of fulfillment is calculated. If this approach is applied in a non-comparative way (i.e., only one investment is evaluated), a minimum average degree of fullfilment of the goals can be set. If the investment's average degree is above the threshold, an investment is considered to be economically viable. If the approach by Tallau et al. is used in a comparative analysis (i.e., several investments are compared with each other), the investment yielding the highest average degree is recommended.

Just as all other approaches analyzed, the approach proposed by Tallau et al. is able to deal with investments made as a whole. As the approach is based on the Balanced Scorecard, the approach is able to consider financial as well as non-financial measures. The approach allows the decision maker to freely choose the measures used for decision support. Therefore, measures for one-time and running benefits and costs, network effects and attacks can be freely chosen even though they are not predefined by the approach. To sum up, the approach meets all mandatory criteria, and partially meets the optional ones.

2.4.11 Approach by Wang et al.

Wang et al. [40] present an approach supporting IS investment decisions based on value-at-risk (*VaR*), a tool originally developed for the assessment of the risk associated with financial assets. With their approach, Wang et al. are able to measure the risk of daily losses and, by using extreme value analysis, to assess the value that is at risk.

VaR denotes the upper limit for daily losses L caused by an IS breach. The loss of the IS breach exceeds *VaR* with probability p. In other words, with a proper IS investment the probability that the daily losses L exceed *VaR* is p. That is,

$$p = Pr[L \geq VaR] = 1 - Pr[L \leq VaR].$$

The daily losses L at a given investment level I is

$$L = \sum_{j=1}^{T} n_j C_j(I),$$

where n_j is the number of occurrences of incident type j, and C_j denotes the costs caused by an incident of type j. Both n_j and C_j assume that the IS investment is in place. The approach by Wang et al. can be applied in two ways. First, in a non-comparative way (i.e., only one investment alternative is evaluated), where *VaR* and the expected daily costs of the investment, consisting of the average daily losses and daily solution costs, are compared with the current situation. Second, in a comparative analysis, in which *VaR* and expected daily costs are calculated and

compared for each alternative IS investment. In both ways, the decision maker then chooses either of the alternatives (or the current status) based on whether he or she strives to decrease the expected daily costs or *VaR*.

Just as the other approaches analyzed, the approach proposed by Wang et al. is able to deal with investments made as a whole as well as with financial measures. With respect to financial measures, the approach considers four predefined costs for the assessment of an investment's economic viability. Non-financial measures, however, are not taken into account by the approach. The approach is able to support one-time costs and benefits, while running costs and benefits are not per se supported. Just as most of the other approaches, the approach presented by Wang et al. is applicable without explicitly considering attacks. Network effects are not per se supported by the approach. To sum up, the approach meets the four mandatory criteria but none of the optional ones.

2.5 Discussion

In this section, we discuss the analysis of approaches supporting investment decisions with respect to the policy and security configuration management tool. More concretely, we highlight the degree to which the analyzed approaches meet the criteria derived from the tool's characteristics and its application in cross-organizational settings. Furthermore, we show commonalities and differences of the approaches.

We start with a general discussion of the approaches. Then, for each criterion the degree to which it is met by the analyzed approaches is discussed. Emphasis is put on the consequences that result from meeting or not meeting the criterion. At the end of this section, we summarize the suitability of each approach to support investment decisions with respect to the policy and security configuration management tool. Finally, we address in more detail the two approaches that at least partially meet all mandatory and optional criteria.

The analyzed approaches can be divided into comparative and non-comparative approaches. The approaches by Bodin et al. [4], Butler [7], Tallau et al. [37], and Wang et al. [40] are intended for comparative analyses. In comparative analyses, several investments are compared to each other. Comparative approaches may be unsuitable in case only one investment needs to be evaluated. In such cases, the investment can be compared to the current situation without the investment being made. Alternatively, as for instance proposed by Tallau et al. [37], a single investment is evaluated and compared to a certain threshold of an overall score. The investment can be made if its score exceeds the threshold. The problem, however, is to determine this threshold. As comparative approaches compare alternative investments with each other, they do not necessarily say whether an investment is economically viable. The other approaches are non-comparative. Such approaches can be used to evaluate a single investment. These approaches yield one result based on which the investment decision can be made. When comparing several

investments using a non-comparative approach, the results of the approaches, for instance, the *ROSI*, are compared.

Comparing the assistance provided by the approaches, we see that the approaches by Gordon and Loeb [17], Huang et al. [23], and Wang et al. [41] help to calculate the optimal as well as the maximal amount that should be invested. The approaches, however, do not say whether one should make a certain investment. Nevertheless, if the costs of an investment are between the optimal and maximal amount to invest, the investment seems reasonable, the nearer to the optimal amount the better. Similarly, the approach by Wang et al. [40] does not say whether an investment should be made. The approach compares alternatives with respect to the investment's costs and the *VaR* of expected losses. It is up to the decision maker to choose an investment based on his or her risk appetite. The three approaches by Bodin et al. [4], Butler [7], and Tallau et al. [37] give an overview of alternative investments and indicate which investment should be favored. Again, the decision to invest remains with the decision maker. The accounting figure based approaches by Al-Humaigani and Dunn [1], Cremonini and Martini [13], and Mizzi [29] provide the expected return of the investment as the result. In case the return is positive, an investment can be made as its benefits are higher than its costs and it thus can be considered economically viable; in case the return is negative, the investment should be neglected; in case the investment equals zero, it remains with the decision maker to invest or not. The same reasoning is applied in the approach by Gordon et al. [20], except that additionally a deferment of the investment decision is taken into account.

All analyzed approaches for supporting IS investment decisions assume that the investment is made as a whole. That is, the investment is not split into smaller parts, where the decision to invest in some parts may be deferred to a later point in time. This criterion is important as the policy and security configuration management tool is provided as a whole only and cannot be split into modules.

All approaches use financial measures for costs and benefits. This is important as the decision to invest is mostly based on financial figures and as an organization's upper management is particularly interested in financial measures.

Investments, however, do not only have financial benefits. The policy and security configuration management tool aims at increasing an organization's trustworthiness, which is hardly expressible in financial measures. Three of the investigated approaches consider non-financial measures: The approach by Tallau et al. [37] considers besides the financial perspective also the customer, the internal process and the innovation and learning perspective to provide decision support. The approach by Butler [7] allows the decision maker to freely choose the measures that will be used to evaluate the investment. The approach by Bodin et al. [4] does not allow such a freedom in selecting measures but assesses, for instance, an investment's security architecture coverage. Finding appropriate measures for assessing investments, however, is difficult, time consuming, and depends on the person responsible for selecting the measures [37]. Allowing the decision maker to freely choose the measures used for evaluation, however, may bear some disadvantages. For instance, relationships between measures may not be obvious.

All approaches take one-time costs and benefits into account. This is important as one-time costs, for instance, for acquiring and deploying the policy and security configuration management tool, incur in any case.

Running costs and benefits, in contrast, are not per se supported by all approaches. The approach presented by Mizzi [29] gives formulas for costs incurring after the first year, however, does not discount them. As the measures can be chosen freely when applying the approaches described by Bodin et al. [4] and Tallau et al. [37], respective measures may be selected. Considering running costs is important, as the policy and configuration management tool offers a dynamic mode in which costs and benefits incur over time. Running costs and benefits may however be considered by discounting them to their present value and by adding them to the one-time costs and benefits.

Only two of the analyzed approaches directly consider attacks. First, the approach described by Mizzi [29] provides an extension that takes the attacker's cost to break a countermeasure into consideration. The approach, however, can be applied without the extension. Second, the approach described by Cremonini and Martini [13] which uses the attacker's return on an attack in the decision support.

As expected, none of the analyzed approaches considers network effects of investments per se. The approaches described by Bodin et al. [4] and Tallau et al. [37] allow the decision maker to freely choose measures to be used in the approach. Therefore, measures focusing on the investment's network effects may be selected and taken into consideration. This way, network effects can be taken into account. The more of an organization's suppliers and clients use the policy and security configuration management tool, the higher will be the overall benefit for all involved parties. This is because information about requirements and configurations can be easily exchanged via the tool. Taking the tool's network effects into account is important as the network effects substantially influence the benefits of the tool.

All things considered, the approach for supporting IS investment decisions presented by Cremonini and Martini [13] is the least suitable of the analyzed approaches. The approach meets three mandatory criteria (i.e., the tool is acquired as a whole, financial measures as well as one-time costs and benefits are supported); it neglects, however, important criteria such as non-financial measures, and running costs and benefits. The approaches presented by Al-Humaigani and Dunn [1], Gordon et al. [20], Gordon and Loeb [17], Huang et al. [23], Sonnenreich et al. [35], and Wang et al. [40] meet all four mandatory criteria. They are, thus partially suitable to assess the economic viability of the policy and security configuration management tool used as the subject of the investment decision. The approach described by Mizzi [29] meets three mandatory criteria (i.e., the tool is acquired as a whole, financial measures as well as one-time costs and benefits are supported) and partially fulfills two optional criteria (i.e., running costs and benefits and attackers). The approach presented by Butler [7] is a comparative approach that meets all four mandatory criteria and the optional criterion regarding non-financial measures. Only the approaches by Bodin et al. [4] and Tallau et al. [37] at least partially meet all criteria. They are, nevertheless, both not perfectly suitable to support investment decisions such as the one regarding the policy and security configuration

management tool. Both approaches are intended for comparative analyses. Thus, the two approaches do not determine the investment's expected return. They also do not calculate the optimal amount to invest given the value and vulnerability of the asset to be protected. Applying one of those approaches, therefore, does not determine the economic viability, but determines which investment should be favoured over other investments. To have an evaluation with respect to financial and non-financial measures, and to determine the return of the investment or the optimal amount to invest, the approaches presented by Bodin et al. [4] and Tallau et al. [37] could be combined with one of the other approaches. For instance, the approach by Gordon and Loeb [17] or Cremonini and Martini [13] seem to be suitable for such a combination.

2.6 Conclusion

In this chapter, we presented and analyzed a set of approaches for supporting IS investment decisions. More concretely, we evaluated and compared approaches with respect to their suitability for assessing the economic viability of a policy and security configuration management tool. Such a tool helps organizations in general and service providers in particular to ensure compliance with the myriad of regulatory and contractual requirements and to reduce the risk of IS breaches. The tool aims at reducing the costs for policy and security configurations management and at increasing the trustworthiness of organizations. Derived from the tool's characteristics and its application in cross-organizational settings, we evaluated and compared the approaches with respect to whether they support investments made as a whole, consider financial and non-financial measures, are able to take one-time and running costs and benefits into account, are applicable without considering attacks, and take network effects into account.

The findings show that there is no approach which meets all criteria. There are, however, approaches, such as those presented by Bodin et al. [4] and Tallau et al. [37] that meet, at least partially, all criteria. They are, however, intended for comparative analyses and thus need to be adapted before they can be used to assess the economic viability of a single investment. It is very likely that two or more of the investigated approaches could be used in combination to assess the economic viability of the policy and security configuration management tool well. Evaluating different combinations of approaches and determining their suitability for the tool is, however, left to future work. As we focused on a specific policy and security configuration management tool in this chapter, the results are specific to the characteristics of this tool. Using another tool as the subject of the investment decision would most certainly lead to other results.

One issue to be kept in mind is that we focused on approaches that help to assess the economic viability of a certain investment. We simply presupposed that the budget to make economically viable IS investments is available. In practice,

however, IS budgets are not inexhaustible. The objective then is to determine how to best spend this fixed budget.

Acknowledgements The research leading to these results was partially funded by the European Commission under the 7th Framework Programme (FP7) through the PoSecCo project (project no. 257129).

References

1. Al-Humaigani, M., Dunn, D.B.: A model of return on investment for information systems security. In: Proceedings of the 46th IEEE International Midwest Symposium on Circuits & Systems, Cairo, vols. 1–3, pp. 483–485 (2003)
2. Anderson, R., Schneier, B.: Guest editors' introduction: economics of information security. IEEE Secur. Priv. **3**(1), 12–13 (2005)
3. Bagchi, K., Udo, G.: An analysis of the growth of computer and Internet security breaches. Commun. Assoc. Inf. Syst. **12**, 684–700 (2003)
4. Bodin, L.D., Gordon, L.A., Loeb, M.P.: Evaluating information security investments using the analytic hierarchy process. Commun. ACM **48**(2), 78–83 (2005)
5. Böhme, R.: Security metrics and security investment models. In: Echizen, I., Kunihiro, N., Sasaki, R. (eds.) Security Metrics and Security Investment Models. Lecture Notes in Computer Science, vol. 6434, pp. 10–24. Springer, Berlin/Heidelberg (2010)
6. Böhme, R., Moore, T.: The iterated weakest link – a model of adaptive security investment. In: Proceedings of the 8th Workshop on the Economics of Information Security (WEIS), London (2009)
7. Butler, S.A.: Security attribute evaluation method: a cost-benefit approach. In: Proceedings of the 24th International Conference on Software Engineering, Orlando, pp. 232–240. ACM (2002)
8. Cavusoglu, H., Cavusoglu, H., Raghunathan, S.: Economics of IT security management: four improvements to current security practices. Commun. AIS **14**, 65–75 (2004)
9. Cavusoglu, H., Mishra, B., Raghunathan, S.: A model for evaluating IT security investments. Commun. ACM **47**(7), 87–92 (2004)
10. Cavusoglu, H., Mishra, B., Raghunathan, S.: The value of intrusion detection systems in information technology security architecture. Inf. Syst. Res. **16**(1), 28–46 (2005)
11. Computerworld: Honda Canada breach exposed data on 280,000 individuals. Website: http://www.computerworld.com/s/article/9217094/Update_Honda_Canada_breach_exposed_data_on_280_000_individuals (2011). Last access 1 Feb 2012
12. Computerworld: RSA warns SecurID customers after company is hacked. Website: http://www.computerworld.com/s/article/9214757/RSA_warns_SecurID_customers_after_company_is_hacked (2011). Last access 1 Feb 2012
13. Cremonini, M., Martini, P.: Evaluating information security investments from attackers perspective: the Return-On-Attack (ROA). In: Proceedings of the 4th Workshop on the Economics of Information Security (WEIS), Cambridge (2005)
14. CSI Computer Survey: 14th Annual CSI Computer Crime and Security Survey, San Francisco (2009)
15. Deloitte: Raising the bar: 2011 TMT Global security study – key findings. http://www.deloitte.com/assets/Dcom-Global/Local%20Assets/Documents/TMT/dttl_TMT%202011%20Global%20Security%20Survey_High%20res_191111.pdf (2011)
16. Franqueira, V., Houmb, S., Daneva, M.: Using real option thinking to improve decision making in security investment. In: Meersman, R., Dillon, T., Herrero, P. (eds.) On the Move to Meaningful Internet Systems. Lecture Notes in Computer Science, vol. 6426, pp. 619–638. Springer, Berlin/Heidelberg (2010)

17. Gordon, L.A., Loeb, M.P.: The economics of information security investment. ACM Trans. Inf. Syst. Secur. **5**(4), 438–457 (2002)
18. Gordon, L.A., Loeb, M.P.: Budgeting process for information security expenditures. Commun. ACM **49**(1), 121–125 (2006)
19. Gordon, L.A., Loeb, M.P.: Economic aspects of information security: an emerging field of research. Inf. Syst. Front. **8**(5), 335–337 (2006)
20. Gordon, L.A., Loeb, M.P., Lucyshyn, W.: Information security expenditures and real options: a wait-and-see approach. Comput. Secur. J. **19**(2), 1–7 (2003)
21. Guardian, T.: Sony suffers second data breach with theft of 25 m more user details. Website: http://www.guardian.co.uk/technology/blog/2011/may/03/sony-data-breach-online-entertainment (2011). Last access 1 Feb 2012
22. Herath, H.S.B., Herath, T.C.: Investments in information security: a real options perspective with Bayesian postaudit. J. Manage. Inf. Syst. **25**(3), 337–375 (2008)
23. Huang, C.D., Hu, Q., Behara, R.S.: An economic analysis of the optimal information security investment in the case of a risk-averse firm. Int. J. Prod. Econ. **114**(2), 793–804 (2008)
24. Kaplan, R.S., Norton, D.P.: The balanced scorecard–measures that drive performance. Harv. Bus. Rev. **70**(1), 71–79 (1992)
25. Kark, K., Orlowv, L.M., Bright, S.: Forrester Research: The change and configuration management software market (2007)
26. Liginlal, D., Sim, I., Khansa, L.: How significant is human error as a cause of privacy breaches? An empirical study and a framework for error management. Comput. Secur. **28**(3–4), 215–228 (2009)
27. Liu, W., Tanaka, H., Matsuura, K.: Empirical-analysis methodology for information-security investment and its application to reliable survey of Japanese firms. Inf. Media Technol. **3**(2), 464–478 (2008)
28. Matsuura, K.: Productivity space of information security in an extension of the Gordon-Loeb's investment model. In: Proceedings of the 7th Workshop on the Economics of Information Security (WEIS), Hanover (2008)
29. Mizzi, A.: Return on information security investment: the viability of an anti-spam solution in a wireless environment. Int. J. Netw. Secur. **10**(1), 18–24 (2010)
30. Oehrlich, E., Lambert, N.: Forrester Research: How to manage your information security policy framework (2006). http://www.forrester.com/The+Change+And+Configuration+Management+Software+Market/fulltext/-/E-RES42580
31. Sadiq, S., Governatori, G., Namiri, K.: Modeling control objectives for business process compliance: business process management. In: Alonso, G., Dadam, P., Rosemann, M. (eds.) Business Process Management. Lecture Notes in Computer Science, vol. 4714, pp. 149–164. Springer, Berlin/Heidelberg (2007)
32. Schneier, B.: Security ROI. Website: http://www.schneier.com/blog/archives/2008/09/security_roi_1.html (2008). Last access 1 Feb 2012
33. Shirey, R.: Internet security glossary – RFC 2828. Tech. rep., The Internet Engineering Task Force – Network Working Group. http://www.ietf.org/rfc/rfc2828.txt (2000)
34. Sklavos, N., Souras, P.: Economic models and approaches in information security for computer networks. Int. J. Netw. Secur. **2**(1), 14–20 (2006)
35. Sonnenreich, W., Albanese, J., Stout, B.: Return on security investment (ROSI) – a practical quantitative modell. J. Res. Pract. Inf. Technol. **38**(1), 55–66 (2006)
36. Su, X.: An overview of economic approaches to information security management. Tech. rep., Centre for Telematics and Information Technology, University of Twente (2006)
37. Tallau, L.J., Gupta, M., Sharman, R.: Information security investment decisions: evaluating the balanced scorecard method. Int. J. Bus. Inf. Syst. **5**(1), 34–57 (2010)
38. Tsiakis, T.K., Pekos, T.: Analysing and determining return on investment for information security. In: Proceedings of the International Conference on Applied Economics (ICOAE), Chania, Crete, pp. 879–883 (2008)
39. Vroom, C., von Solms, R.: Towards information security behavioural compliance. Comput. Secur. **23**(3), 191–198 (2004)

40. Wang, J., Chaudhury, A., Rao, H.R.: A value-at-risk approach to information security investment. Inf. Syst. Res. **19**(1), 106–120 (2008)
41. Wang, S.L., Chen, J.D., Stirpe, P., Hong, T.P.: Risk-neutral evaluation of information security investment on data centers. J. Intell. Inf. Syst. **36**(3), 329–345 (2011)
42. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q **26**(2), xiii–xxiii (2002)
43. Whitman, M.E.: Enemy at the gate: threats to information security. Commun. ACM **46**(8), 91–95 (2003)
44. Willemson, J.: On the Gordon and Loeb model for information security investment. In: Proceedings of the 5th Workshop on the Economics of Information Security (WEIS), Cambridge (2006)
45. Willemson, J.: Extending the Gordon and Loeb model for information security investment. In: Proceedings of the 5th International Conference on the Availability, Reliability, and Security (ARES'10), Krakow, pp. 258–261 (2010)

Chapter 3
Ad-Blocking Games: Monetizing Online Content Under the Threat of Ad Avoidance

Nevena Vratonjic, Mohammad Hossein Manshaei, Jens Grossklags, and Jean-Pierre Hubaux

Abstract Much of the Internet economy relies on online advertising for monetizing digital content: Users are expected to accept the presence of online advertisements in exchange for content being free. However, online advertisements have become a serious problem for many Internet users: while some are merely annoyed by the incessant display of distracting ads cluttering Web pages, others are highly concerned about the privacy implications – as ad providers typically track users' behavior for ad targeting purposes. Similarly, security problems related to technologies and practices employed for online advertisement have frustrated many users. Consequently, a number of software solutions have emerged that block online ads from being downloaded and displayed on users' screens as they browse the Web.

We focus on these advertisement avoidance technologies for online content and their economic ramifications for the monetization of websites. More specifically, our work addresses the interplay between users' attempts to avoid commercial messages and content providers' design of countermeasures. Our investigation is substantiated by the development of a game-theoretic model that serves as a framework usable by content providers to ponder their options to mitigate the consequences of ad avoidance techniques. We complement our analytical approach with simulation results, addressing different assumptions about user heterogeneity.

N. Vratonjic (✉) · J.-P. Hubaux
School of Computer and Communication Sciences, EPFL, Lausanne, Switzerland
e-mail: nevena.vratonjic@epfl.ch; jean-pierre.hubaux@epfl.ch

M.H. Manshaei
Department of Electrical and Computer Engineering, Isfahan University of Technology, Isfahan, Iran
e-mail: manshaei@gmail.com

J. Grossklags
College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA
e-mail: jensg@ist.psu.edu

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_3,

Our findings show that publishers who treat each user individually, and strategically deploy fee-financed or ad-financed monetization strategy, obtain higher revenues, compared to deploying one monetization strategy across all users. In addition, our analysis shows that understanding the distribution of users' aversion to ads and valuation of the content is essential for publishers to make a well-informed decision.

3.1 Introduction

> It is difficult to produce a television documentary that is both incisive and probing when every twelve minutes one is interrupted by twelve dancing rabbits singing about toilet paper. (Rod Serling 1997)

Consumers and content providers have a love-hate relationship with advertisements. In the area of online news sites, 81 % of a surveyed consumer sample report the acceptance of the presence of online advertising in exchange for content being free. At the same time, 77 % state that they would hardly ever click on these ads [36]. More significantly, across all media channels, 69 % say they are "interested in products and services that would help them *skip* or *block* marketing messages [40]."

Each media genre is affected with its own specific advertisement circumvention challenges. During TV commercial breaks, viewers can leave the room to do small chores. Ads in video recordings can be manually skipped with fast-forwarding or are automatically marginalized with advanced functions of digital video recorders (e.g., TiVo) and VCRs [23]. This trend has accelerated with the availability of Home Theater PC systems such as Windows Media Center, SageTV Media Center and MythTV where available third-party add-ons allow consumers to conveniently skip ads (e.g., Comskip and ShowAnalyzer). In telemarketing, consumers are able to screen calls with CallerID or utilize software tools that act on their behalf (e.g., Telemarketing Blocker). Further, regulatory intervention can have a significant impact, for example, with the US Do-Not-Call list that upon registration allows consumers to opt out from unsolicited telephone marketing calls [51].

We focus on advertisement avoidance technologies (AATs) for Web content and their economic ramifications. In the past few years, a number of effective software solutions have emerged of which the most prominent is perhaps the Adblock Plus third-party extension for the Firefox browser family [5,39]. According to up-to-date statistics provided by Mozilla, Adblock Plus has been downloaded over 172 million times since July 2006, and has an active daily user base of about 14 million consumers. Further, it is also among the most popular add-ons for the Google Chrome browser with more than 100,000 weekly installs. Observers from the advertising business have predicted that the "importance of Adblock is its potential for extreme menace to the online-advertising business model [10]". However, many other technology options exist to block ads.

The emergence of behavioral ad-targeting and the associated increase in advertisers' incentives for user tracking, has led to what some observers call

a "data collection arms race" (see, for example, [12]). Most recently, Google's proposed changes to its privacy policy that would allow for more pervasive user data aggregation have refreshed privacy concerns in consumers' minds (see, for example, [21]). And consumers object to such practices [28, 48]. However, in the absence of truly effective *and* widespread technologies to opt-in/opt-out from tracking and the later usage of such information for ads, consumers only have the option to decide on their own personal mix of avoidance technologies. For example, while consensus for a powerful and broadly applicable Do-Not-Track mechanism is still absent, some users might seek to disable scripting languages, Flash or cache cookies.[1] Others might use advanced privacy-enhancing technologies such as Tor just for the purpose of evading such commercially-motivated tracking. Finally, to be effective, avoidance of tracking does frequently necessitate also the blocking of the display of ads since ad campaigns almost always involve some form of campaign management tools. While this is trivially necessary to allow for ad-related payment flows, consumers cannot easily distinguish between different degrees of tracking severity.

So far, the impact of the circumvention of online tracking and advertisements has been moderated by the overall growth of the market for Internet commercials. The Interactive Advertising Bureau estimates that online advertising in the United States in 2011 totaled $31.7 billion and has grown by 22 % compared to the previous year [24]. Nevertheless, many content sites suffer from the burden of ad-blocking tools, in particular, if they cater to a technology-savvy audience (see, for example, [17]). The search for an adequate response to this threat has so far proven inconclusive. In particular, monetization approaches do not only have to be economically sensible, but need to be accompanied by technically sound implementations. So far, ad-block deterrence solutions have been notably absent from the marketplace, even though the cost of development and deployment of simple approaches would be very manageable. In fact, as the majority of ad-blocking tools are based on filtering out elements whose URLs contain keywords like *ad* or *click*, omitting these keywords would make existing ad-blocking tools ineffective. In addition, existing tools cannot automatically detect URLs likely to be ads. Therefore, if publishers start using different keywords, ad-blocking systems would not work [39].

The stakes described in this chapter are very high and are relevant beyond the discussions about the effectiveness of marketing or commercial mechanisms. In fact, the popularity of Adblock-style add-ons represents only the tip of the iceberg, as many related challenges are consuming the attention of content producers. For example, applications such as Flipboard allow users to conveniently grab pictures and articles from many different content resources to display them in a variety of user-defined formats, and ads could be left behind (or replaced).

[1] It is unlikely that a meaningful compromise on Do-Not-Track will be reached quickly. See, for example, the counterarguments on such technology brought forward by leading content providers [4].

Our work studies in detail and in a quantitative manner the implications of a (likely to happen) growing usage of ad-blocking technologies and addresses the economic justification for effective countermeasures concerning ad avoidance. To achieve that goal, we develop a game-theoretic model that takes into account the most relevant parameters, identifies different canonical options (strategies) that the content providers and the users can choose from and forecasts the most likely outcome of such situations. The models we provide rely on Subgame Perfect Nash Equilibria (SPNE) and on Perfect Bayesian Nash Equilibria (PBNE). We complement our analytical approach with simulation results by addressing different assumptions about user heterogeneity. We make "common sense" assumptions in terms of cost and show that in general, content providers are better off when they make use of a "mixed approach", namely when they simultaneously rely on fee-funded and ad-funded monetization strategies.

The chapter is structured as follows. We survey the related work in Sect. 3.2. In Sect. 3.3, we introduce the reader to background information relevant to the problem area of ad avoidance. After briefly laying out the roadmap for our analysis in Sect. 3.4, we delve into the details of our game-theoretic models in Sect. 3.5. We present simulation results in Sect. 3.6 and concluding remarks in Sect. 3.7.

3.2 Related Work

Closely related to our work is an economic model by Tåg [45]. Content providers decide whether to offer to users a subscription option that eliminates advertisements as an alternative to the content with advertisements. The content provider would introduce such an option only if the revenue gained from those customers who are willing to pay the subscription fee is greater than the revenue that the content provider would earn by only offering the basic advertisement model. According to the model, if the subscription option is introduced, it causes an increase in advertising quantity in the free version, thus increasing the annoyance due to ads and reducing the perceived quality of the free version. Moreover, consumers' aggregate utility decreases, while content providers' and advertisers' profits increase. By increasing the amount of advertisements to non-subscribers, the content provider can further increase the differentiation between the two options. Prasad et al. [35] analyze the incentives to price discriminate when consumers are of two given types and a content provider offers two versions differing in advertising quantity and price. They show that offering two versions (price discrimination) tends to be optimal in most cases.

In another model, Shah accounts for ad avoidance technologies [37]. Users can invest in ad avoidance options but will still see a certain fraction of the commercials. A content provider can make use of this fact by optimally differentiating the amount of advertisements catered to the two groups (i.e., users with and without ad avoidance products). In a two-sided market model for television advertising, Anderson and Gans similarly show that content providers could increase the number

of ads to those users who do not invest in avoidance technologies, as they are less averse to advertising [3]. They note that this effect is not solely due to the incentive of content providers to regain the revenue, but rather due to revealed preferences of those who do not invest in ad avoidance technologies. In practice, this may be one of the contributing reasons that a larger number of ads per hour are observed in US television recently (the US does not impose a cap on the number of commercials, in contrast to the EU). As a result, overall welfare and program quality could decrease and programming would be tailored to appeal to a broader range of viewers.

In [52], Wilbur presents a two-sided, empirical model of television advertising and models the effects of an ad-avoidance technology on an advertisement-supported media industry. The model considers the following two possibilities. First, to overcome the loss caused by ad avoidance technologies, networks could increase the quantity of ads, which makes AAT even more valuable to ad-adverse viewers. Therefore, this scenario leads to mutually reinforcing increases in AAT penetration and advertising time. Second, if advertisers value users with AAT less, as they fast-forward through ads, then non-AAT users become scarce and more valuable. Due to this self-selection, the remaining market is composed of viewers who accept ads which might lead to increased ad prices for advertising space. The competition for non-ad-avoiding viewers can lead to lower advertising levels, rendering ad-avoidance technologies less valuable and slowing down its rate of growth. The author uses a counterfactual experiment to gain insight into how AAT affects the industry. It is shown that when AAT penetration increases, then ad levels rise as well. Nevertheless, increased AAT levels lead to revenue loss, which implies that AAT might decrease a content provider's incentives to invest in program quality. Another model analyzes the impact of ad-avoidance behavior considering two alternative schemes by which media channels are financed: free-to-air and pay-TV [44]. The model also considers market competition in the two scenarios. The analysis shows that increased AAT levels lower profits and decrease entry in the free-to-air model. In contrast, in the pay-TV regime, lower income from ads is compensated by higher subscription fees, therefore the profits and the number of channels are unaffected.

In our model, we explicitly consider the limited information aspects related to ad avoidance technology and its detection. As a result, content providers must invest in detection technologies to be able to distinguish between consumers that utilize AATs and those who do not engage in such activities. Such user differentiation enables content providers to deploy a personalized approach, treating each user individually and applying an appropriate monetization strategy per user. It also enables deployment of countermeasures that affect only the AAT users (e.g., preventing access to the content unless they turn off AATs or subscribe). A personalized approach is not possible in the traditional TV market, as providers do not have technological means to detect who is using AAT (e.g., fast-forwarding through ads). Therefore, the previous work has only considered an aggregate strategy for a content provider, which is applied across all the users, regardless of whether they use AATs or not. In such a scenario, instead of impacting only AAT users, the countermeasures taken to offset losses due to AATs either affect all, or even

worse, only the non-AAT users. For example, an increased advertisement level only impacts non-AAT users (while AAT users can fast-forward through ads). Thus, there are no incentives for AAT users to change their behavior. On the contrary, such an approach increases incentives to adopt AATs. In our model, the countermeasures directly affect the AAT users and therefore discourage their use of AATs. Moreover, our model leads to stronger differentiation since AAT users are not of any value to advertisers as online AATs block all available ads, whereas in the TV market, users who fast-forward through ads are still exposed to traces of marketing content.

Further academic works on advertisement circumvention have been undertaken in the context of "old media" from a legal or ethical perspective [23, 41, 49]. Additional recent work has been focused on improvements of the mechanisms for ad allocations and techniques to lower the impact of manipulation by malicious actors. See, for example, research papers on ad auctions (e.g., [14, 50]) and click fraud [27, 29].

3.3 Background

In this section, we discuss the drivers of consumer resistance to advertisements and their propensity for ad blocking. We also review existing technologies for ad avoidance and approaches by website owners to detect ad-blocking software.

3.3.1 Why Do Consumers Block Ads?

Previous research has studied a variety of ad avoidance behaviors such as eliminating, ignoring or quickly flipping past commercial messages [43]. Graphical and auditory stimuli are frequently considered annoying or unconvincing, irrespective of the actual information content [43]. Online ads are more likely to be avoided if consumers hold expectations of a negative experience, are generally skeptical towards the advertisements or contest their relevance [26]. Further, if a user perceives an interruption in his primary interaction objective or considers ads to clutter his workspace, marketing messages are more likely to be blocked or ignored [9].

Further, sophisticated online advertising approaches such as personalized, behavioral or targeted delivery mechanisms rely on the collection and use of data about users' Web interactions. Different studies have documented users' misgivings and privacy concerns about these practices. For example, in an interview study of 1,000 adult consumers, 66 % objected to tailored ads [48]. Due to the pervasiveness of these concerns, (self-)regulatory and technical proposals are under consideration, e.g., that would allow users to opt-out from such data collection practices by signing up for a Do-Not-Track list [11]. At the same time, users can attempt to block advertisements altogether when suspecting that they are triggered by the

Table 3.1 Survey results: why do consumers use Adblock Plus?

Reasons	No opinion (%)	Not important (%)	Somewhat important (%)	Important
Distracting animations and sound	4.3	5.6	15.6	74.5
Offensive/inappropriate ad content	8.0	20.1	23.3	48.6
Reduce page load time and bandwidth use	5.7	10.1	22.6	61.6
Missing separation between ads and content	13.2	11.5	27.5	47.8
Privacy concerns	8.3	9.9	27.5	54.3
Security concerns	8.0	9.7	26.1	56.3
Ideological reasons	20.2	32.0	24.2	23.7

tracking of their online trails. In addition to privacy issues, online advertisements also present security threats. Infected online ads are often used to compromise ad viewers' machines and spread malware [42] or direct the machines to participate in ad-fraud scams. Users do not even have to click on ads to trigger malware and the consequences can be devastating. In a sophisticated ad-fraud scheme discovered in 2012, shutting down malicious servers that orchestrate the fraud and control victims' machines would lead to all the victims losing their Internet service [32]. Most of these users were even unaware that their machines had infected and mitigation of the effects of the scam represented a big challenge.

A survey of 1543 Adblock Plus users further evidenced that privacy and security concerns are major factors to select this application [33]. Avoiding distractions and improving website load time performance, however, are the dominating reasons. Interestingly, the lowest score of importance was given to ideological reasons. See Table 3.1 for the full results [33].

3.3.2 What Technologies Are Involved?

Ad-blocking tools prevent online ads from being downloaded and displayed on users' screens as they browse the Web. They can also be considered privacy-preserving tools as some forms of online tracking (e.g., via cookies) can be evaded. Typically, ad-blocking tools are available as free downloadable plug-ins and exist for several Web browsers. For example, Adblock Plus is open-source and maintained by an international community of voluntary helpers. Internet Explorer 9 includes a directly embedded functionality primarily used for *tracking protection*, but also allows one to block some unwanted content.

Ad-blocking tools rely on two mechanisms to block ads: (i) prevent loading of elements whose URLs match *filter rules* used to classify elements as ads, and (ii) hide page elements that match a Cascading Style Sheets (CSS) selector.

Users can subscribe to different community-generated filter lists or manually specify filtering rules themselves. They can also decide to allow loading of some elements of a page or to turn-off ad-blocking on specific pages or websites. However, this feature is not widely used among Adblock users [33].

Ad-blocking causes revenue loss for advertisers and ad networks but it has the most significant impact on websites whose business model is based on online advertising. The majority of websites today rely on ad revenue, whereas only a few websites have successfully implemented- subscription and membership-based systems for revenue. Therefore, it is understandable that site operators might want to discourage or thwart ad-blocking. In particular, a website can detect the use of ad-blocking tools with a JavaScript that executes after the page is loaded and verifies that the ads are displayed. Then, the website could take one of the following countermeasures: (i) inform users about adverse effects of ad-blocking on the website and ask them to turn it off; (ii) prevent users from accessing the content unless they disable ad-blocking; (iii) embed ads in a way that ad-blocking filters cannot easily differentiate ads from content; (iv) tie the functionality of websites to the download of ad elements; and (v) offer users to pay subscription fees for ad-free content.

Both the ad-blocking and detection tools currently come at a very low cost. The former requires the user to install a browser plug-in and subscribe to filter lists. As for detecting ad-blocking, the required JavaScript code is easily available online.

3.4 Analysis Overview and Assumptions

We propose a game-theoretic model of the informational consequences of consumers' ad circumvention and website owners' detection of these practices. In our analysis, we model the strategic interactions between a generic website W and a user U and we iteratively consider the following three cases: (i) without the presence of ad blocking and ad circumvention detection technologies; (ii) with ad blocking but no detection, and (iii) where both technologies are available to consumers and website owners, respectively. Throughout the rest of the chapter, we use the terms "website" and "website owner" interchangeably.

A key assumption we make is that the website attempts to analyze users *individually*. A number of technologies exist to implement various forms of conditional content and ad delivery (see, for example, [22]) ranging from tailoring a website's appearance to the type of browser and operation system in use by the consumer. Note that the individualized analysis does not necessarily translate into unique monetization strategies for each user.

Website owners can utilize two canonical types of monetization strategies in response to a particular user: either employ ad-financed content delivery or propose a micropayment for access to content (as a representative subcase of a wider range of payment-based strategies, such as subscriptions). The consideration of micropayments for newspaper content is extremely timely. Not only has the debate

about micropayment schemes for news and other digital content been fought very passionately over the last few years [25, 38]; but from an actual deployment point of view, easy-to-manage systems are now available, for example, One Pass from Google [20] or PayPal for Digital Goods [34]. And consumers seem more willing than ever to accept small charges in response for immediate content or entertainment needs [38].

We further assume that the website is aware of the user's valuation of content, for example, because of the cooperation with ad networks, inference about the resources the user is trying to access or previous interactions. In practice, websites work on obtaining such information and use it to, for example, compute appropriate prices for their services or content (e.g., The New York Times' subscription price is based on the estimates of readers' valuations of the content, which is set such that the current paywall system should be accepted by a certain fraction of their readership [30]). Our analysis can also be easily extended to introduce uncertainty about user's content preferences from the content providers' perspective.

Not all aspects about user behavior are immediately observable without sophisticated detection technologies. In particular, the website cannot easily deduce whether the consumer is taking advantage of ad-blocking tools. This is especially the problem in the impression-based ad revenue model, in which the website obtains ad revenue for each ad displayed to its visitors. For example, if the feedback cycle between the ad network and the website is not real-time then payoff consequences of ad avoidance are only realized at a later time. In the click-based ad revenue model, a website gets paid for users' clicks that get reported to the ad network, thus perhaps enabling more direct and immediate control. The absence of signals could indicate to ad networks (and websites) a change in the user's behavior (e.g., use of AB software). The website can mitigate this information disadvantage by investing in technologies to detect ad avoidance. In this work, we focus on impression-based model and we note that a similar analysis can be provided for the click-based model.

Based on these assumptions, we model each website visit as a sequential game between the two players, a website W and a user U, to highlight the informational and strategic aspects of the interactions. We represent the different cases as game trees (see Figs. 3.1 and 3.2) with the notation provided in Table 3.2. In each game, the players can choose from the corresponding strategy sets and the payoffs achieved at the end of the game are represented in a format (P_W, P_U), where P_W and P_U are total payoffs of W and U, respectively.

3.5 Game-Theoretic Models

In this section, we introduce game-theoretic models that capture strategic interactions of a website W and a user U. For each model, we present analysis methodology and the obtained results.

Table 3.2 Symbols for the game-theoretic models

Symbol	Definition
b	User's "benefit" of viewing content
c	User's "cost" of viewing ads
s	Subscription fee
r_i	Ad Network's per-impression ad revenue
C_B	Cost of using AB software
C_D	Cost of detecting AB software
α	Belief about the reached information set
P_W	Website's total payoff
P_U	User's total payoff
AF	Ad-financed content
FF	Fee-financed content (micropayments)
DI	Invest in detection of AB software
NI	No investment in detection of AB software
B	Block ads
A	Abstain from blocking ads
P	Pay subscription
N	Not pay subscription
$(x\|y)$	First (x) and second (y) action of a player
(x, y)	Strategy profile: (first mover, second mover)

3.5.1 Model 1: No Blocking and No Detection

We introduce the reader to our approach by first proposing a basic model of the interaction between websites and users in which no ad-blocking and no detection technologies are used by users and websites, respectively. Afterwards, we slowly increase the complexity of the model to account for ad avoidance and countermeasures.

3.5.1.1 Model Setup: An Extensive Form Game with Complete Information

The content provider selects between fee-financed (e.g., micropayments) and ad-financed monetizing scheme for his content (denoted by FF and AF, respectively). If presented with a website that solicits a fee to access its content, users can elect to transmit a payment, P, or to deny payment and forfeit access, N. The website will either earn positive revenues from the ad impression, r_i, or from the micropayment, s. The consumer receives a benefit, b, from accessing the content and pays either the fee, s, or has a cost c due to accepting ads. The subscription fee s is determined by the content provider and it is the same for all the users, because it has been shown that price discrimination is not economically optimal for providers [1] and because of users' protest (e.g., the case of Amazon [7]). Determining the optimal price is not the goal of this work, but is certainly noteworthy to explore.

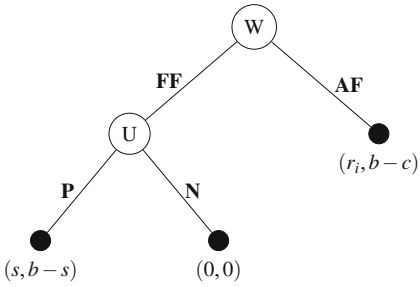

Fig. 3.1 Model 1: Game tree for the basic model with no blocking and no detection

The cost c captures all the negative aspects of receiving ads from the users' point of view (summarized in Table 3.1). Figure 3.1 summarizes the characteristics of the basic model.

3.5.1.2 Analysis Methodology: Subgame Perfect Nash Equilibrium

The basic model belongs to the class of *perfect and complete information extensive form* games. In these games, each player always knows the previous moves of all players when he has to make his move. In [18], it is proven that every finite extensive-form game of perfect information has a pure-strategy Nash equilibrium. We use a *Subgame Perfect Nash Equilibrium* (SPNE) solution concept that is a refinement of a Nash equilibrium in dynamic games. In game theory, a strategy profile is a SPNE if it represents a Nash equilibrium of every subgame of the original game.

A common method for determining SPNE is *backward induction* and we apply it in our analysis. Backward induction can be applied to any finite game of perfect information. This technique eliminates incredible equilibria and assumes that: (i) the players can reliably forecast the behavior of other players, and (ii) the players believe the other players can do the same. In the game defined by Fig. 3.1, the user knows that he is the player that has the last move. Hence, for each possible move of the website the user selects his best response. For example, if the website plays FF, the user concludes that with move P he obtains the best payoff if and only if $b > s$.

Now we consider how the website chooses its best strategy using backward induction. Let us assume that $b > s$. The website then knows that if it plays FF, the user's best response is P, which results in the payoff of s for the website. However, if the website plays AF, its payoff would be r_i. Hence, the website's best response is FF, if $s > r_i$. In summary, if $b > s$ and $s > r_i$ strategy profile (FF, P) is the SPNE of the game in Fig. 3.1. Table 3.3 summarizes all possible SPNE of the defined game.

3.5.1.3 Results

Following this methodology, Table 3.3 summarizes all possible SPNE of the defined game, considering different values of game parameters.

Table 3.3 SPNE of Model 1

$b > s$	$s > r_i$	(FF, P)
	$s < r_i$	(AF, P)
$b < s$		(AF, N)

It follows that a website owner would only implement a fee-financed revenue scheme when users' value of the provided content is sufficiently high, $b > s$, and the expected ad-revenue does not exceed fee payments, $s > r_i$. The first condition is relatively difficult to assess for a large number of diverse users if the revenue policy cannot be set adaptively for each consumer. In contrast, the second condition allows for a more straightforward calculation – at least for an impression-based ad model. We address the impact of the heterogeneity of the users in simulations (Sect. 3.6).

3.5.2 Model 2: Blocking, Detection Versus No Detection

In the following, we extend the analysis to include consumers having the opportunity to utilize ad-blocking software and website owners to potentially respond by investing in detection technologies. The expanded game is represented in Fig. 3.2.

3.5.2.1 Model Setup: An Extensive Form Game with Imperfect Information

Consumers now have the option to block ads, B, at cost C_B, or to abstain from ad-blocking, A, which does not incur any direct cost. We assume that websites are aware of the possibility of ad-blocking, but without an investment in detection technologies, NI, are not able to differentiate between users with and without AB tools and thus hold only *imperfect* information about the user's action and its payoff consequences. In contrast, when the website is equipped with detection technologies, DI, at cost C_D, the information barrier is resolved. The informational consequences are easily discernible in Fig. 3.2 by observing the dotted lines between information sets that indicate the website's uncertainty about the reached state in the game and the eventual outcomes. Websites have to formulate a probabilistic assessment α of the reached state of the game, following the user's decision to block ads or to abstain.

We further break down the game into two subgames concerning the website's decision to invest or not in detection of AB software, as highlighted by the left and right boxes in Fig. 3.2, respectively. The analysis of the left-hand-side subgame in Fig. 3.2 (i.e., when website plays DI) is similar to the calculation of SPNE, presented in Sect. 3.5.1.2. Using the same methodology we obtain SPNE of this subgame and present the obtained results later in Table 3.5.

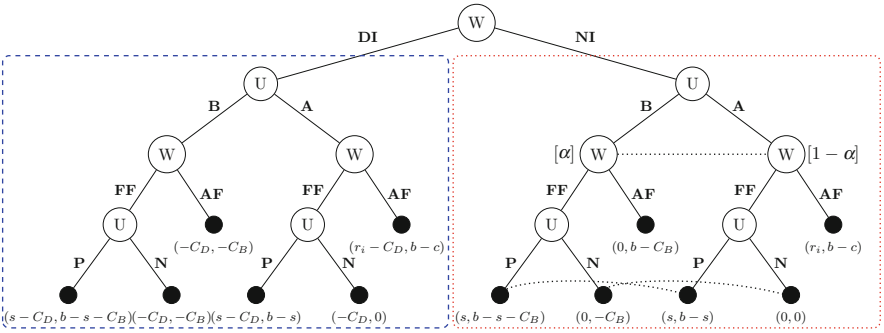

Fig. 3.2 Model 2: Game tree for ad-blocking with and without detection technologies

3.5.2.2 Analysis Methodology: Perfect Bayesian Nash Equilibrium

The subgame in the right-hand side of the game in Fig. 3.2 belongs to the class of *complete imperfect sequential* games, because one player does not have information about the opponent's action played in the previous stage of the game. In other words, the website owner does not know whether the user has already installed AB software or not, when he wants to choose the monetization strategy for the website's content (i.e., use ad-financed or fee-financed strategy).

Next, we discuss the game-theoretic concept of the *Perfect Bayesian Nash Equilibrium* (PBNE) that helps us get an insight into the strategic behavior of players in such games. PBNE was developed in order to refine the Bayesian Nash equilibrium concept and remove implausible equilibria in sequential games [19]. More specifically, the concept of PBNE is defined by four Bayes requirements that eliminate unwanted subgame-perfect equilibria [31]. We discuss these requirements considering the defined subgame represented in the right-hand side box in Fig. 3.2.

Requirement 1: The player with the move must have a belief about which node in the information set has been reached by the play of the game. For example, in Fig. 3.2 the website believes that the user installed AB with a probability of α.

Requirement 2: At the PBNE strategy profile, players must be sequentially rational given the players' beliefs. A strategy profile is sequentially rational if and only if the expected payoff of the player who has the move at that information set is maximal given the strategies played by all the other players. For example, in Fig. 3.2 the website should calculate its expected payoff for playing AF and FF, given its belief α and choose the strategy that maximizes its expected payoff. Given website belief, the expected payoff from playing FF is $\alpha \times s + (1-\alpha) \times s = s$. The expected payoff from playing AF is $\alpha \times 0 + (1-\alpha) \times r_i = (1-\alpha) r_i$. Hence if $\alpha > \frac{r_i - s}{r_i}$, the website plays FF to be sequentially rational.

Requirement 3: The player must update his belief at the PBNE to remove implausible equilibria of BNE on the equilibrium path. These beliefs are determined by Bayes' rule and the players' equilibrium strategies. In other

Table 3.4 PBNE of submodel without detection

		$C_B < c$	$C_B > c$
$b > s$	$s > r_i$	(A|P, FF; $\alpha = 0$)	(A|P, FF; $\alpha = 0$)
	$s < r_i$	(A|P, FF; α) $\alpha > \frac{r_i - s}{r_i}$	(A|P, FF; α) $\alpha > \frac{r_i - s}{r_i}$
		(B|P, AF; α) $\alpha < \frac{r_i - s}{r_i}$	(A|P, AF; $\alpha = 0$)
$b < s$		(B|N, AF; $\alpha = 1$)	(A|N, AF; $\alpha = 0$)

Table 3.5 SPNE of submodel with detection

		$C_B < c - s$	$C_B > c - s$
$b > s$	$s > r_i$	(A|P, FF)	
	$s < r_i$	(B|P, FF)	(A|P, AF)
		$C_B < c - b$	$C_B > c - b$
$b < s$		(B|N, FF)	(A|N, AF)
		(B|N, AF)	

words, players should first calculate the equilibrium paths of the complete perfect information game. If the calculated strategy that satisfies sequential rationality is on the equilibrium path, there is no uncertainty for the player at the PBNE (i.e., α equals 0 or 1).

Requirement 4: Finally, the belief should be updated considering the sequential rationality and players' equilibrium strategies where possible.

In the right-hand subgame presented in Fig. 3.2, if $b > s$, $s < r_i$, and $C_B > c$ there exists an equilibrium path of $(A|P, AF)$. Although, the user cannot play P when the website deploys AF strategy, we use $A|P$ notation to represent the full strategy profile of the user at the equilibrium path. This means that if $\alpha < \frac{r_i - s}{r_i}$, the PBNE is $(A|P, AF; \alpha = 0)$ (i.e., Requirement 3). Requiring that each player has a belief and acts optimally given this belief, it suffices to eliminate the implausible equilibria for the belief of $0 < \alpha < \frac{r_i - s}{r_i}$. But, if $\alpha > \frac{r_i - s}{r_i}$, the PBNE is $(A|P, FF; \alpha)$, because we cannot eliminate any implausible equilibria for this strategy profile (i.e., Requirement 4). Similar calculations can be made for other cases.

3.5.2.3 Results

Applying this methodology, we can derive results presented in tabular fashion for the right-hand side (Table 3.4) and the left-hand side (Table 3.5) of Fig. 3.2.

If website owners do not invest in detection, we observe that ad blocking happens in two instances (see Table 3.4). First, when consumers do not value the content highly enough to pay a fee ($b < s$), and ad-blocking is cheap relative to the cost of viewing ads ($C_B < c$). Second, if website owners believe it to be unlikely that consumers block ads ($\alpha < \frac{r_i - s}{r_i}$) and ad-blocking is cheap, then ad avoidance can persist even when users value the content sufficiently ($b > s$). In both cases, the

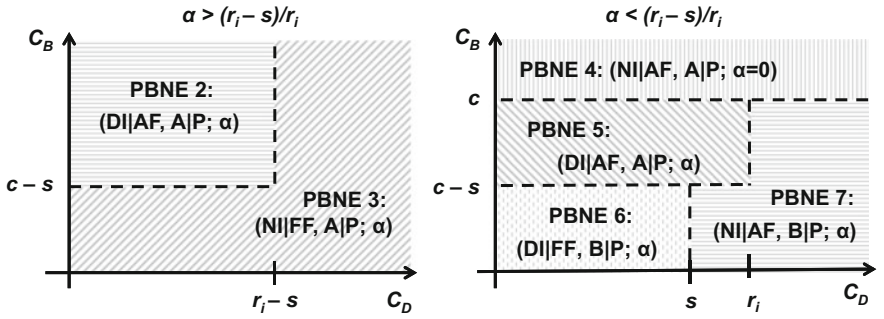

Fig. 3.3 Case 2: Users value the content and are willing to pay subscription fees ($b > s$). The website prefers ad-financed to fee-financed monetization strategy ($r_i > s$)

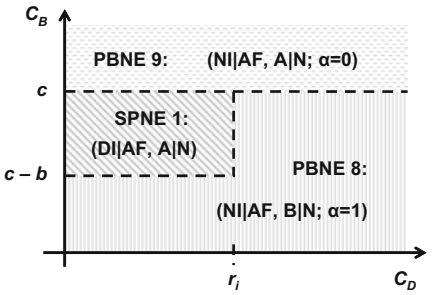

Fig. 3.4 Case 3: Users do not value content sufficiently to pay subscription fees ($b < s$)

user will exploit his information advantage to avoid ad clutter while the website will gain nothing through the interaction (because it mistakenly relies on ad-financed strategy, AF).

In contrast, with an investment in detection technology the website owner can partially crowd out the ill-effects of ad avoidance. He can successfully solicit a micropayment even when ad-blocking technology is cheap as long as the user values the content sufficiently (see Table 3.5). However, the website will still not extract any benefits from a user who does not value the content highly and has access to cheap ad-blocking technology. Interestingly, the website is indifferent in the latter case about allowing the user to access the content freely (with blocked ads) or not. Importantly, the introduction of detection technology also lowers the threshold of what a user considers to be cheap ad-blocking, i.e., the consumer now internalizes the cost of the expected micropayment when making the assessment ($C_B < c - s$).

We now proceed to visualize the space of equilibria from a different perspective in Figs. 3.3 and 3.4 by integrating the results of the subgames from the left-hand and right-hand side of Fig. 3.2. The figures show how the equilibrium strategies of the players depend on the cost of detection, C_D, and ad-blocking, C_B, technologies, respectively. We break down the results based on the equilibrium beliefs of the

website, i.e., Fig. 3.3 is split according to the threshold belief, $\alpha^* = \frac{r_i - s}{r_i}$. Figure 3.4 shows the cases where the website is certain about the consumer's strategies. In addition (and not visualized), for the case of high content value, $b > s$, and low ad-revenue, $s > r_i$, we also find that the website and the user select PBNE $1 = (NI|FF, A|P; \alpha = 0)$, independently of C_B and C_D.

3.6 Simulation Approach and Results

Our analysis in Sect. 3.5 provides a framework that websites can use to determine which countermeasures concerning ad avoidance they should use to maximize the revenue. Our results show that the best response depends on the type of users that a given website serves. In this section, we illustrate how our framework can be used to determine the best response while taking into account different assumptions about user heterogeneity with respect to user perception of content and ads.

3.6.1 Simulation Setup

We model the application of our framework to a popular website with specific and unique content that is of a high value to its visitors (e.g., the Financial Times). The Financial Times is a good example as it is a content provider that deploys both monetization strategies: fee-financed and ad-financed. Our game-theoretic analysis shows that the outcome of the game depends mostly on the parameters that characterize visitors of a given website: users' benefit of viewing the content, users' cost of viewing ads with the content and ad revenue that the website earns for each pageview. As discussed in Sect. 3.4, the values of per-impression ad revenue and users' benefit of viewing the content are available to the stakeholders, namely websites and ad networks. It is more difficult to obtain exact values for users' cost of viewing ads and, to do so, websites could perhaps position themselves with respect to the reasons users have named in the survey on why they block ads (Table 3.1). Depending on how much they match users' criteria, they can estimate their visitors' costs. In addition, as we will show, knowing the distribution of such a variable for which the relevant parameter is the fraction of users who use ad-blocking software (e.g., available from Firefox statistics) is sufficient for the model.

We rely on Web analytics providers, Alexa and Google's DoubleClick Ad Planner, to obtain the data based on which we can estimate the parameter values. We use the following values in our evaluations:

1. The website receives one million pageviews per day, as reported by Google's DoubleClick Ad Planner [13].
2. In the case of fee-financed content, we consider a micropayment of $s = \$0.321$ per pageview. We compute this value based on the Financial Times' subscription

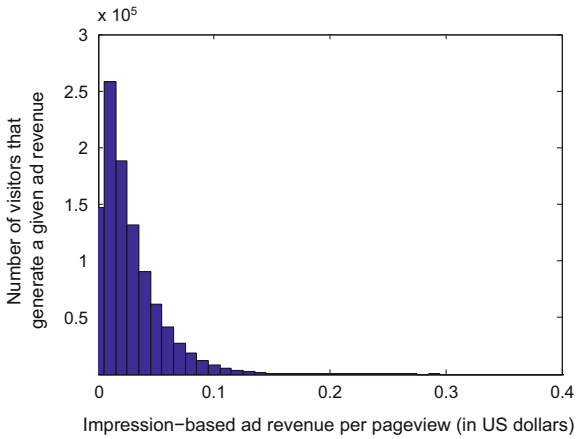

Fig. 3.5 Distribution of user-generated impression-based ad revenue per pageview

fee of $4.99 per week [16] and the 2.22 average number of pageviews per visitor per day, as reported by Alexa [2]. As explained in Sect. 3.5, the subscription fee is the same for all users.

3. We model the impression-based ad revenue per pageview with a beta distribution represented in Fig. 3.5 based on the estimated cost-per-mille (CPM) between $1 and several tens of dollars [46]. CPM is a cost that advertisers pay for a 1,000 impressions and thus we compute the per-pageview ad revenue as $CPM/1{,}000$ for the considered values of CPM. We select skewed distribution as most of advertisers pay CPM in the range of a couple of dollars and only a very few major advertisers pay a high CPM in the order of tens of dollars. The total ad revenue that the website can earn in our model is in the range of the reported ad revenues by the top blog websites [47] with a similar number of daily pageviews [13].
4. Benefit b (expressed in US dollars) of users viewing the content (Fig. 3.6) is drawn from a beta distribution (in the range of values comparable to the impression-based ad revenue per pageview), such that 25 % of the visitors would opt for fee-financed content (i.e., has $b > s$). This number is in compliance with 25 % of Financial Times' visitors paying for digital subscriptions [15]. In addition, for most of the websites users' benefits are high due to users' self-selection bias. The exact values are not necessary, the important parameter is the fraction of users accepting to pay the subscription fees.
5. We consider a population of visitors that consists of: (i) a fraction $(1 - \gamma)$ of users who are indifferent about ads and therefore do not use AB software, and (ii) a fraction γ of users who are heterogeneous in how much they like or dislike ads and therefore might use AB software. Users who are indifferent about ads associate a small cost (expressed in US dollars) to viewing online ads. Other users, who are not indifferent about ads, have a higher cost of viewing ads, that can even surpass the benefit they associate to viewing the content. However, it does not necessarily mean that all of them use AB software. Their decision on whether to use AB software (Block) or not (Abstain) then depends on the cost of

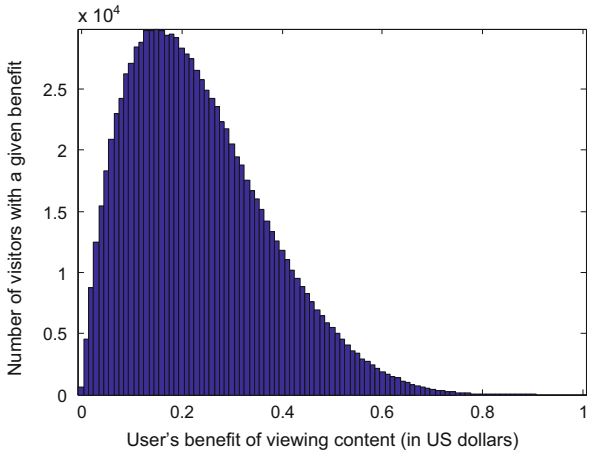

Fig. 3.6 Distribution of users' benefits of viewing content per pageview

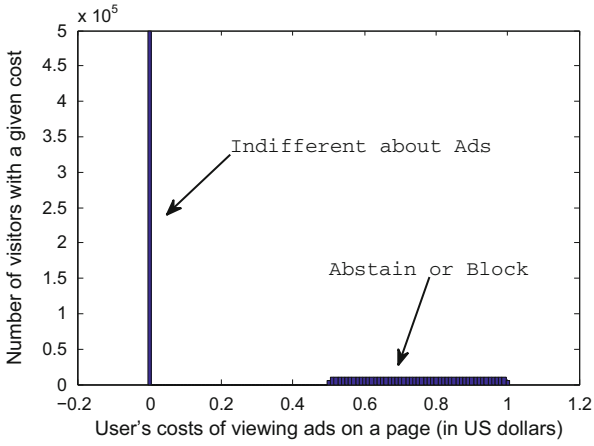

Fig. 3.7 Distribution of users' costs of viewing ads per pageview: fraction $(1-\gamma)$ of users indifferent to ads; fraction γ of users who choose between Abstain (no AB software) and Block (using AB software)

viewing ads with respect to the values of other parameters (e.g., their valuation of the content or the cost of using AB software). Therefore, the parameter c that represents users' costs of viewing ads is drawn from a bimodal distribution (Fig. 3.7), that assigns a small cost to the users indifferent about ads (the first mode of the distribution) and higher costs to other users (the second mode). The values of c are in the range comparable to the impression-based ad revenue per pageview and users' valuation of the content. Figure 3.7 depicts the distribution for $\gamma = 0.5$. We vary the value of γ in the simulations.

6. In practice, the cost of blocking ads (C_B) corresponds to the cost of installing and maintaining a browser add-on and subscribing to filter lists that define blocking rules. At the moment, the cost (C_D) of detecting AB software on users' machines corresponds to the cost of including a specific Javascript into Web pages. Nowadays, both of these costs (expressed in US dollars) are very small and we use values of $C_B = \$0.01$ $C_D = \$0.001$ for our simulations.

Note that these values represent costs per interaction and have such a low value as they are factored out on millions of users (for C_D) and a number of pageviews per day (for C_B). These costs could increase if an arms race develops between AB softwares and detection tools, as was the case with pop-up ads and pop-up blockers [8]. We evaluate the effect of higher costs of blocking and detection later in the analysis.

3.6.2 Results

We simulate the interaction between the website and the population of users, based on our game-theoretic model and parameter values described above. The website treats each user individually and applies the framework to each of the visitors. We then aggregate the results of the interactions to represent the outcomes for the entire population of visitors. The fraction (γ) of users that might potentially install AB software is a variable in our simulations. For each value of $\gamma \in \{0.05, 0.1, 0.2, 0.3, 0.4, 0.5\}$, we generate a corresponding bimodal distribution (as in Fig. 3.7) that assigns the values to users' costs of viewing ads (c). The values of all other parameters remain fixed.

First, we compare the revenues that the website obtains by deploying three different monetizing strategies: (i) serving ad-financed content (AF model) to all visitors, regardless of whether they use AB software or not; (ii) serving fee-financed (FF) content, where users have to pay a subscription fee in order to access the content; (iii) game-theoretic approach (GT model) where a website chooses an appropriate strategy according to our analysis, and can either serve ad-financed or fee-financed content to different users. Figure 3.8 depicts the daily revenue of the website, for the three models, depending on the fraction of users that might potentially block ads. We observe that the revenue that the website obtains with GT monetizing model is superior to using pure fee-financed (FF) or ad-financed models (AF). The reasoning behind such a result is as follows. In the AF model, users with AB software do not generate ad revenue for the website, as ad impressions are blocked on their machines. The higher the potential number of users with AB software (γ), the higher the revenue loss for the website. In the FF model, only users who value the content more than the subscription fee are willing to pay, thus the revenue is not influenced by the users who use AB, only by the number of subscriptions. FF revenue depends on the subscription fee that the website can charge, which mostly depends on the content it serves and how valuable it is to its visitors. The GT model represents a compromise between AF and FF models. For users who dislike ads, but value content enough to pay subscription fees, the website will apply the FF strategy. With AF, the website cannot make profit out of these users as they block ads. For users who do not dislike ads as much, the website might either use the FF or the AF strategy, whichever is more profitable. Thus, the GT model enables the website to take into account users' heterogeneity and maximize its profit. In Fig. 3.9 we show the fraction of users that generate profit

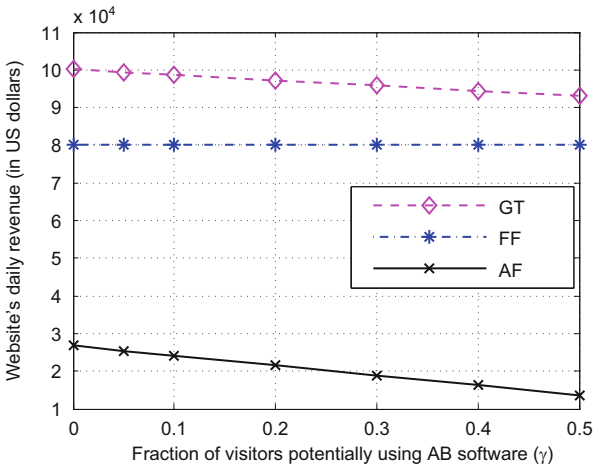

Fig. 3.8 Website's daily revenue (in US dollars) with different monetizing models

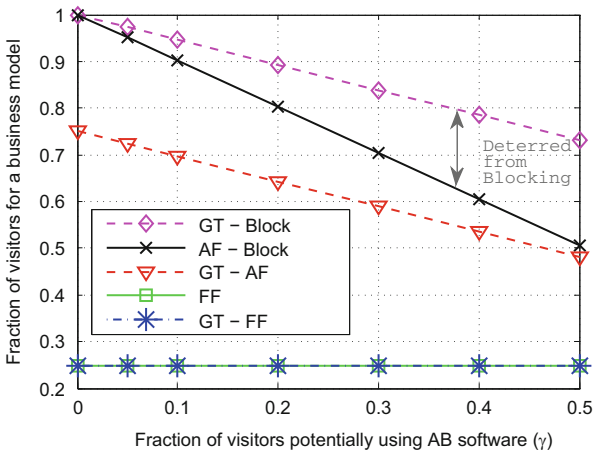

Fig. 3.9 Fraction of visitors that generate revenue for each monetizing model

for the website with the three monetizing models. The curve labeled *AF-Block* represents the fraction of users from which the website profits in the AF model. In this model, the ad revenue is generated only by the users without AB software. Note that nevertheless all users obtain the content. The difference between the *AF-Block* curve and 1 corresponds to the fraction of users who use AB software in the AF model. In the fee-financed (FF) model, only the users who opt to pay the subscription generate the revenue for the website and obtain the content (*FF* curve[2]). In the GT model, the website profits from serving ad-financed content to a fraction of users (*GT-AF* curve) and fee-financed content to another fraction of users (*GT-FF* curve). The sum of these two corresponds to the total fraction of users

[2]Note that the *FF* curve overlaps with the *GT-FF* curve.

that the website can generate revenue from, represented with the *GT-Block* curve. The remaining fraction of users (i.e., the difference between the *GT-Block* curve and 1) corresponds to the users with AB software in the GT model. Users served with ad-financed content are those who: (i) accept to view ads in exchange for free content (for which the outcome of the game is PBNE 9: $(NI|AF, A|N; \alpha = 0)$), or (ii) value the content more than they dislike ads, but not enough to pay the subscription fee for ads-free content (for which the outcome of the game is SPNE 1: $(DI|AF, A|N)$). Users who are served fee-financed content are those who: (i) dislike ads but value the content, or (ii) users who accept ads but also value the content, thus leaving the choice to the website that could decide to offer the subscription model to such users as it might be more profitable. These are the users for which the outcome of the game is PBNE 1: $(NI|FF, A|P; \alpha = 0)$. We observe that the total fraction of users that generate the revenue for the website in the GT model (*GT-Block*) is higher than in either the AF or the FF model.

Users who do not generate revenue and do not obtain the content in the GT model are those who dislike ads and do not value the content enough to pay subscription fees. This case corresponds to PBNE 8: $(NI|AF, B|N; \alpha = 1)$. Note that the impact of the users with AB software is smaller in the GT model, and we see that in the worst case about 27 % block ads (and generate revenue loss for the website) compared to the 50 % in the AF model. These results are in line with the results in Fig. 3.8 and explain why the website earns more with the GT monetizing model. In the worst case, the GT revenue is around 16 % higher than the FF revenue and it may not seem justified to deploy the GT model for that increment in the revenue. However, one major advantage of the GT model is that it maximizes the number of users who obtain the content (73 % in the GT model compared to 25 % in the FF model, in the worst case). We conclude that the GT model allows the website to adapt its monetizing strategy such that it maximizes the number of visitors from whom it profits, as well as its visibility or impact factor.

As discussed previously, the website can deploy a strategy of making it more difficult for AB software to filter out and block ads. In our GT model, this action can be represented with an increase in the users' cost of blocking ads and a higher investment in the detection. We simulate the effect of higher ad-blocking and investment costs ($C_B \in 0.01, 0.1, 0.5, 0.7, 1$; and $C_D = \$0.1$) and represent the results in Fig. 3.10. Different curves correspond to the fraction of users that the website can profit from in the GT model, considering different costs of ad-blocking. We observe that the fraction of users that will block ad-financed content decreases with the increase in the cost of blocking ads. As both the website and users are behaving strategically in the GT model, with the higher cost rational users deter from blocking ads and it shows that the website has a good return-on-investment with the strategy of making ad-blocking more difficult.

In summary, we have illustrated how a website can use our framework in practice as a decision help in addition to the content provider's overall business strategy and factors that are outside the scope of our model. We have demonstrated how a website maximizes its revenue with a strategic choice of its best response when facing users with different preferences with respect to ads and content. Such a strategic behavior

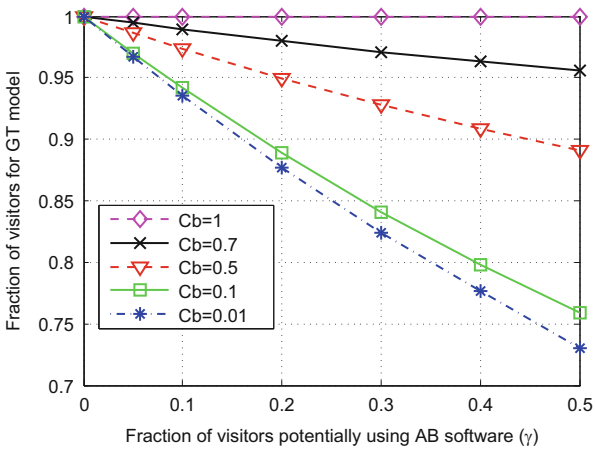

Fig. 3.10 Fraction of visitors that generate revenue in the GT model, considering higher blocking and detection costs

allows for the website to maximize the number of users from which it can profit, as well as to apply the strategy that maximizes the profit. Users' strategic behavior allows them to maximize their utility as well, by having a choice of viewing ad-financed or fee-financed content.

3.7 Conclusion

In this chapter, we conduct a systematic study of the consequences of ad avoidance on the business model of content providers. We develop a framework usable by content providers to ponder their options to mitigate the consequences of ad-avoidance technologies. We carefully devise and analyze a game-theoretic model of the impression-based ad revenue mechanism and illustrate with simulations the impact of different strategies under parameter assumptions motivated by real-world data. Our analysis shows that deploying a game-theoretic approach, i.e., strategically applying a fee-financed or an ad-financed monetization strategy, and treating each user individually yields higher revenues for publishers, compared to deploying one strategy across all users. Also, understanding the distribution of users' aversion to ads and valuation of the content is essential for publishers to make well-informed decisions. We expect that our modeling and simulation assumptions are a reasonable, but likely not a perfect fit for every situation involving content providers and ad avoiders. In future work, we intend to further explore deviations from our modeling assumptions and expand our framework to additional problem areas.

Our contribution is only a first step to account for the complicated interactions between ad avoidance and content monetization. For example, a promising area for additional work is to more carefully address the impact of the negative feedback spiral caused by the adoption of ad-blocking under the presence of

limited information. A loss of revenue through an increase of website visitors who use ad-blocking software will frequently trigger a more aggressive pursuit of advertisement opportunities. Those might even include consumer-unfriendly affiliate marketing schemes. While this may create short-term benefits, additional consumers will depart or try to avoid these practices.

In addition, we aim to consider measures of concentration and interdependency in the ad industry. For example, a recent study shows that Google-controlled cookies were present on 97 of the top 100 websites [6]. The same study also documents the growing intricacy of tracking attempts that will make it very difficult for users to find adequate countermeasures in the absence of market (self-)regulation.

In conclusion, we expect content providers that serve a technology-minded audience to suffer most from ad avoidance technologies. And, in the absence of a broad consensus between the ad and content industry, on the one side, and consumers, on the other side, the trend towards blocking of advertisements is likely to grow. Resistance to user tracking and the desire for ad avoidance are tightly interwoven, even though we do not model the related long-term trends at the moment, i.e., users rarely become technology-savvy ad avoiders overnight. However, the potential for a significant shift in consumer behavior is large and should not be under appreciated.[3]

Acknowledgements We thank the anonymous reviewers and participants at the Workshop on the Economics of Information Security (WEIS) 2012 for their valuable comments and feedback. The presentation at WEIS was partially supported by travel funding from the Volkswagen Foundation. Jens Grossklags gratefully acknowledges the support from the Swiss National Science Foundation's International Short Visit Program and from Google's Faculty Research Award Program.

References

1. Acquisti, A., Varian, H.R.: Conditioning prices on purchase history. Mark. Sci. **24**, 367–381 (2005)
2. Alexa, The Web Information Company: Available online at http://www.alexa.com/siteinfo/ft.com
3. Anderson, S., Gans, J.: Platform siphoning: Ad-avoidance and media content. Am. Econ. J. Microecon. **3**(4), 1–34 (2011)
4. Arstechnica Opposition Letter: Available online at http://static.arstechnica.com/oppositionletter.pdf (2011)
5. Aycock, J.: Spyware and Adware. Advances in Information Security. Springer, New York (2010)
6. Ayenson, M., Wambach, D.J., Soltani, A., Good, N., Hoofnagle, C.J.: Flash cookies and privacy II: now with HTML5 and ETag respawning. In: World Wide Web Internet and Web Information Systems (2011). http://ssrn.com/abstract=1898390 or http://dx.doi.org/10.2139/ssrn.1898390

[3] A 2010 study revealed that up to 40 % consumers are willing to change their online behavior if advertisers were collecting data [28].

7. Bezos Calls Amazon Experiment “a Mistake”: Available online at http://www.bizjournals.com/seattle/stories/2000/09/25/daily21.html (2000)
8. Can’t Stop the Pop-Ups: Available online at http://news.cnet.com/2100-1024_3-5226273.html (2004)
9. Cho, C., Cheon, H.: Why do people avoid advertising on the Internet? J. Advert. **33**(4), 89–97 (2004)
10. Cohen, N.: Whiting Out the Ads, but at what cost? The New York Times (2007). http://www.nytimes.com/2007/09/03/technology/03link.html
11. Commission, F.T.: Protecting consumer privacy in an era of rapid change: a proposed framework for businesses and policymakers. Preliminary FTC Staff Report (2010)
12. Data Collection Arms Race Feeds Privacy Fears: Available online at http://www.reuters.com/article/2012/02/19/us-data-collection-idUSTRE81I0AP20120219 (2012)
13. DoubleClick Ad Planner by Google: Available online at https://www.google.com/adplanner/ (2012)
14. Edelman, B., Ostrovosky, M., Schwarz, M.: Internet advertising and the generalized second-price auction: selling billions of dollars worth of keywords. Am. Econ. Rev. **97**(1), 242–259 (2007)
15. Financial Times, Digital Subscribers: Available online at http://aboutus.ft.com/corporate-information/ft-company/ (2012)
16. Financial Times, Subscription Fees: Available online at https://registration.ft.com/signup/standard?execution=e1s1 (2012)
17. Fisher, K.: Why Ad blocking is devastating to the sites you love. Ars Technica (2010). http://arstechnica.com/business/2010/03/why-ad-blocking-isdevastating-to-the-sites-you-love/
18. Fudenberg, D., Levine, D.: Subgame-Perfect equilibria of finite- and infinite-horizon games. J. Econ. Theory **31**(2), 251–268 (1983)
19. Fudenberg, D., Tirole, J.: Perfect Bayesian equilibrium and sequential equilibrium. J. Econ. Theory **53**(2), 236–260 (1991)
20. Google One Pass: Available online at http://www.google.com/landing/onepass/ (2011)
21. Google Privacy Changes Must Be Stopped, Group’s Lawsuit Says: Available online at http://www.businessweek.com/news/2012-02-13/google-privacy-changes-must-be-stopped-group-s-lawsuit-says.html (2012)
22. Ha, S.: An intelligent system for personalized advertising on the Internet. In: Proceedings of the 5th International Conference on E-commerce and Web Technologies (EC-Web), Zaragoza, pp. 21–30 (2004)
23. Haskins, J.: Commercial skipping technology and the new market dynamic: the relevance of antitrust law to an emerging technology. Duke Law Technol. Rev. **8**(1) (2009)
24. Internet Advertising Bureau: IAB Internet Advertising Revenue Report, 2011 Full Year Results. Available online at http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_Report_FY_2011.pdf (2011)
25. Isaacson, W.: How to save your newspaper. TIME Magazine (2009). http://content.time.com/time/magazine/article/0,9171,1877402,00.html
26. Kelly, L., Kerr, G., Drennan, J.: Avoidance of advertising in social networking sites: the teenage perspective. J. Interact. Advert. **10**(2), 16–27 (2010)
27. Kshetri, N.: The economics of click fraud. IEEE Secur. Priv. **8**(3), 45–53 (2010)
28. McDonald, A.M., Cranor, L.F.: Americans’ attitudes about Internet behavioral advertising practices. In: Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society (WPES), Chicago, pp. 63–72 (2010)
29. Mungamuru, B., Weis, S.: Competition and fraud in online advertising markets. In: Proceedings of the 12th International Conference on Financial Cryptography and Data Security (FC), Cozumel, pp. 187–191 (2008)
30. NYTimes’ “Fair” Prices: Available online at http://www.mondaynote.com/2011/03/21/nytimes-%E2%80%9Cfair%E2%80%9D-prices/ (2011)
31. Okada, A.: Perfect bayesian equilibrium and sequential equilibrium. In: Wiley Encyclopedia of Operations Research and Management Science. Wiley, Hoboken (2010)

32. Operation Ghost Click: Available online at http://www.fbi.gov/news/stories/2011/november/malware_110911 (2012)
33. Palant, W.: Adblock Plus User Survey. Available online at http://adblockplus.org/blog/adblock-plus-user-survey-results-part-2 (2011)
34. PayPal Merchant Services: Available online at https://merchant.paypal.com/cgi-bin/marketingweb?cmd=_render-content&content_ID=merchant/digital_goods (2012)
35. Prasad, A., Mahajan, V., Bronnenberg, B.: Advertising versus pay-per-view in electronic media. Int. J. Res. Mark. **20**(1), 13–30 (2003)
36. Rainie, L., Purcell, K.: State of the news media 2010: online economics and consumer attitudes. Report produced by the Pew Internet Project and the Pew Research Center's Project for Excellence in Journalism (2010)
37. Shah, S.: Ad-skipping and time-shifting: a theoretical examination of the digital video recorder (2011). Working paper, University of Virginia
38. Sindik, A., Graybeal, G.: Newspaper micropayments and millennial generation acceptance: a brand loyalty perspective. J. Media Bus. Stud. **8**(1), 10–20 (2011)
39. Singh, A., Potdar, V.: Blocking online advertising – a state of the art. In: Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Monash University, Gippsland (2009)
40. Smith, W.: Consumer Resistance to Marketing Reaches All-Time High, Marketing Productivity Plummets, Yankelovich Partners, AAAA Conference (2004)
41. Snow, N.: The TiVo question: does skipping commercials violate copyright law? Syracuse Law Rev. **56**(1), 27–84 (2005)
42. Sood, A.K., Enbody, R.J.: Malvertising – exploiting web advertising. Comput. Fraud Secur. **2011**(4), 11–16 (2011)
43. Speck, P., Elliott, M.: Predictors of advertising avoidance in print and broadcast media. J. Advert. **26**(3), 61–76 (1997)
44. Stühmeier, T., Wenzel, T.: Getting beer during commercials: adverse effects of Ad-avoidance. Inf. Econ. Policy **23**(1), 98–106 (2011)
45. Tåg, J.: Paying to remove advertisements. Inf. Econ. Policy **22**(4), 245–252 (2009)
46. The Average CPM Rates Across Different Verticals. Available online at http://www.labnol.org/internet/average-cpm-rates/11315/ (2010)
47. Top Earning Blogs: Available online at http://onlineincometeacher.com/money/top-earning-blogs/ (2011)
48. Turow, J., King, J., Hoofnagle, C., Bleakley, A., Hennessy, M.: Americans reject tailored advertising and three activities that enable it. Available at University of Pennsylvania Scholarly Commons (2009)
49. Vallade, J.: Adblock plus and the legal implications of online commercial-skipping. Rutgers Law Rev. **61**(3), 823–853 (2009)
50. Varian, H.: Position auctions. Int. J. Ind. Organ. **25**(6), 1163–1178 (2007)
51. Varian, H., Wallenberg, F., Woroch, G.: The demographics of the do-not-call list. IEEE Secur. Priv. **3**(1), 34–39 (2005)
52. Wilbur, K.: A two-sided, empirical model of television advertising and viewing markets. Mark. Sci. **27**(3), 356–378 (2008)

Chapter 4
Software Security Economics: Theory, in Practice

Stephan Neuhaus and Bernhard Plattner

Abstract In economic models of cybersecurity, security investment yields positive, but diminishing, returns. If that were true for software vulnerabilities, fix rates should decrease, whereas the time between successive fixes should go up as vulnerabilities become fewer and harder to fix.

In this work, we examine the empirical evidence for this hypothesis for Mozilla, Apache httpd and Apache Tomcat over the last several years. By looking at 292 vulnerability reports for Mozilla, 66 for Apache, and 21 for Tomcat, we find that the number of people committing vulnerability fixes changes proportionally to the number of vulnerability fixes for Mozilla and Tomcat, but not for Apache httpd.

Our findings do not support the hypothesis that vulnerability fix rates decline. It seems as if the supply of easily fixable vulnerabilities is not running out and returns are not diminishing (yet).

Additionally, software security has traditionally been viewed as an arms race between attackers and defenders. Recent work in an unrelated field has produced precise mathematical models for such arms races, but again the evidence we find is scant and does not support the hypothesis of an arms race (of this kind).

4.1 Introduction

In standard economic models of cybersecurity, security can be bought incrementally, but there are diminishing returns: some investment later buys less security than the same investment earlier [20]. From comparing several cybersecurity models, Rue et al. find that the security function (the function that says how much return

S. Neuhaus (✉) · B. Plattner
Eidgenössische Technische Hochschule Zürich, Zürich, Switzerland
e-mail: stephan.neuhaus@tik.ee.ethz.ch; bernhard.plattner@tik.ee.ethz.ch

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_4,

one gets for how much investment) that underlies every economic model of cybersecurity is increasing and concave.[1]

In order to view software security in economic terms, we assume this workflow:

1. Programmer A commits a *vulnerability*, by which we mean an error that has security consequences.
2. That vulnerability is discovered and *assigned* to programmer B for fixing. We might have $A = B$.
3. Programmer B *fixes* the vulnerability, either by making all the required changes himself (even in code not written by him), or by making other people contact the people who own the code that is to be corrected.
4. A fix might not solve the vulnerability completely, causing steps 2 and 3 to be *repeated*.

If fixing vulnerabilities could be modeled as a security function, one expends a certain amount of work and reduces the total number of vulnerabilities by one. As time goes by, vulnerabilities become fewer and harder to fix, so returns will be diminishing. Certainly, adding more people to fix vulnerabilities will at some point only bring diminishing returns as administrative overhead eats up the potential for increased productivity. That this occurs is clearly the expectation of at least some economists [20].

It is also of economic interest to know how often vulnerability fixes need to be repeated for a single vulnerabilty. Clearly, the fewer the better.

Another concern is the size of the *security team*, by which we mean the number of developers who fix vulnerabilities. The size of the security team ought to depend on the code ownership model. In one code ownership model that we call *individual code ownership*, the person having written the code in question retains ownership of that code and is expected to fix any bugs pertaining to it, including vulnerabilities. In this case, we will have $A = B$ in step 2, and we will have A contact other code owners in the workflow above. In the *communal code ownership* model, in principle anyone can fix anybody else's code, so we will in general have $A \neq B$, and B will fix the vulnerability all by himself.

We can see from this that individual code ownership should lead to a large security team, since many code owners will commit vulnerability fixes, whether they are interested in security or not, whereas in communal code ownership, checkins will be made by people who like fixing vulnerabilities, and the security team will therefore be smaller. This has staffing consequences, since in individual code ownership, *all* developers will have potentially to be schooled in how to fix

[1] The authors use 'convex' instead of 'concave', and by 'convex' they mean "any twice continuously differentiable function". But unless that function has a negative second derivative, the diminishing returns don't happen. However, a negative second derivative is a criterion of *concavity*, not convexity, and two times continuous differentiability is not needed for concavity.

security issues. This is different from writing secure code, which all developers ought to know anyhow.

From a different perspective, software security is often seen as an arms race between hackers and programmers, sometimes termed "Red Queen" (for the attackers) and "Blue King" (for the programmers). The term comes from Lewis Carroll's *Through the Looking-Glass*, where the Red Queen complains that "it takes all the running you can do, to keep in the same place" [4]. In our case, the Red Queen tries to find and exploit vulnerabilities, which the Blue King then eventually patches. If the Blue King is successful, the Red Queen would then have to expend time and energy, just to keep her supply of possible exploits constant ("running [. . .] to keep in the same place"). In this case, the time between successive vulnerability fixes should increase and obey a power law [9].

Our objects of study are three large software systems: first, the Mozilla suite (including Firefox and Thunderbird, but also lesser-known products like Seamonkey); second, the Apache HTTP server httpd; and third, the Apache Tomcat application server. We chose these projects because they are widely used, because they represent a wide variety of vulnerability-prone, internet-facing software, and because development data such as source code repositories and bug and vulnerability databases are freely available.

Our contribution is twofold. First, from the theoretical considerations above, we have the following predictions, which we check on our three software systems:

- Vulnerability fix rates should decrease as diminishing returns set in (we find no evidence that this prediction is true);
- The size of the security team should be correlated with the vulnerability fix rate or not, depending on the code ownership model (again we find no evidence that this is true); and
- The time between vulnerability fixes should increase and obey a power law (the time tends very roughly to increase, but the data do not support a power law).

Second, we also find that:

- The quality of vulnerability fixes is good: most vulnerabilities are fixed once, with no need to fix the fix; and
- The distribution of vulnerability fixes per day obeys some, and perhaps the same, heavy-tailed distribution for all three applications. This is a result that was not derived by a research question, but came from the data analysis. We do not yet have a theoretical explanation for this behaviour, but speculate on possible causes; see Sect. 4.4.2.

The rest of this chapter is organised as follows. First, we describe our data collection (Sect. 4.2) and evaluation methodology (Sect. 4.3). Then we describe our findings and speculate about possible causes for the results (Sect. 4.4). Finally, we describe related work (Sect. 4.5), analyse threats to validity (Sect. 4.6) and conclude with conclusions and future work (Sect. 4.7).

The dataset and the scripts that were used to generate the figures in this section are available from ftp://ftp.tik.ee.ethz.ch/pub/publications/WEIS2012/fixrates.tar.gz.

4.2 Data Collection

For our purposes, a *checkin* or *commit* is the uploading of a set of related changes to source files to a source code repository. Checkins are usually perceived to be atomic, but this is not always the case; see below.

When programmers fix a vulnerability, or part of a vulnerability, they commit those changes to the source code for which they have the responsibility or authority to change. The systems where checkins and vulnerability information are stored are usually separate, so in order to find out which checkins fixed vulnerabilities, they have first to be united. This section spells out in more detail how this process works for the three software systems under observation.

4.2.1 Mozilla: Reconstituting Checkins

The Mozilla Foundation uses Mozilla Foundation Security Advisories (MFSAs) to report security issues in Mozilla products.[2] Our dataset contains 417 MFSAs, starting with MFSA 2005-01, issued January 21, 2005, and ending with MFSA 2011-18, issued April 28, 2011.

MFSAs can contain references to bug reports, and there can be zero or more such references; see Fig. 4.1. These references in turn contain the bug identifier, chosen by the bug tracking software. (The identifier does not carry any semantics.) While not all bugs are vulnerabilities, all vulnerabilities are bugs, so we will identify a vulnerability in Mozilla by its bug identifier. If the programmer committing the fix for a vulnerability mentions its bug identifier in the commit message—as is usual in Mozilla—we can identify it as fixing that vulnerability. Typical commit messages contain "bug *n*", "bug=*n*", or "b=*n*".

Not all bug identifiers that are mentioned in MFSAs appear in checkin messages. We call a bug identifier *unassigned* when there is no checkin carrying that identifier; otherwise we call it *assigned*.

Also, not all MFSAs contain such bug identifier references. For example, MFSA 2010-79 references a Java LiveConnect bug and an entry in the Common Vulnerabilities and Exposures database, but no bug identifier. In all, only 382 out of 417 MFSAs have such references. In analogy to bug identifiers, we call an MFSA *assigned* if it contains a bug identifier, and *unassigned* otherwise.

[2] http://www.mozilla.org/security/announce/

Fig. 4.1 A typical MFSA, showing the bug identifier

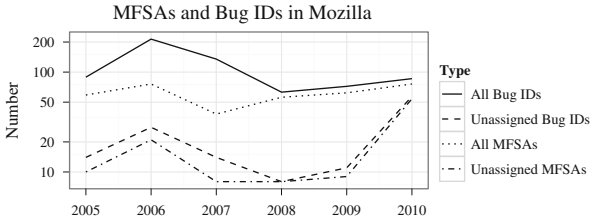

Fig. 4.2 Number of MFSAs and their associated bug identifiers, by year. The 'Year' column contains the year in the MFSA name, which is not necessarily the same as its publication year. Note the logarithmic *y* axis

Even worse, only 292 MFSAs were fixed using assigned bug identifiers. The earliest example of an MFSA that we could not associate with a checkin is MFSA 2005-12 ("javascript: Livefeed bookmarks can steal cookies"). This MFSA is associated with the bug identifier 265668, which appears nowhere in the CVS log messages.

One particular aspect of the Mozilla checkins is that they have been historically stored in a CVS version archive. After March 2007, there has been a Mercurial repository too, but the main work went on in the CVS repository, which was considered to be authoritative. Only since October 2010 has the "relic master copy" officially switched to the Mercurial repository.[3]

In the top two lines in Fig. 4.2, we see that since 2008, there is a trend towards "one MFSA, one bug identifier". This is unlikely to happen by chance, so we suspect either a conscious effort by the Mozilla team to assign only one bug identifier per vulnerability, or a bug identifier now tends to be assigned only after the corresponding vulnerability is discovered.

Generally, the gap between the total number of bug identifiers or MFSAs on the one hand and the number of unassigned bug identifiers or MFSAs on the other is

[3] See comment on changeset 56642:882525a98119.

closing since 2009. This seems to indicate that the main work is now going on in the Mercurial repository.

Mercurial commits all changed files as a single changeset. CVS, however, does not have this notion of a checkin transaction; files are individually version-controlled by RCS, so there is no easy way to ask "which files were affected by the last checkin?". This is perfectly fine in itself, but since from a developer perspective, fixes occur in checkins, we therefore need to reconstruct those checkins from the individual version-controlled files. Also, since there are 639,783 individual file checkins, or *CVS log entries*, in our dataset, the reconstruction algorithm needs to be efficient. For the purposes of our analysis, individual file changes in the CVS log are considered part of the same commit if they have the same commit message and if the time between them is less than 5 min. When applied to our dataset, we found 186,170 checkins in the Mozilla CVS, ranging from March 28, 1998 to February 22, 2011.

4.2.2 Apache httpd and Tomcat: Assigning Checkins

The source code and other development artefacts for Apache httpd (from now on simply called httpd) does not offer a way to link vulnerabilities and their fixing checkins. Apache publishes security information about httpd[4] using the Common Vulnerabilities and Exposures (CVE) database. Each report also contains timestamps when the security team was made aware of the vulnerability, then the issue was published, when the update was released, and which versions are affected. What is missing is the link to a bug report (called a 'Problem Report' by the developers, and abbreviated PR). Therefore, we went manually through all reported vulnerabilities and tried to find the fixing commit ourselves.

Such an approach naturally introduces uncertainties. We have therefore labeled our vulnerability-to-commit mapping with the degree of certainty that the commit in question is actually the one that fixes the named vulnerability. There are four certainty levels:

- **Certain.** The commit message contains the CVE name and asserts that it fixes the issue described in it.
- **High.** The commit message mentions issues thematically related to the CVE message and the timeline fits (commits must come after the security team was made aware of the issue, but before a fix is released), but the CVE identifier is not explicitly mentioned.
- **Low.** One or more of the above indicators (commit message, timeline) fits, but not all of them. The CVE identifier is not mentioned.
- **Unassigned.** No commit had any of the above indicators.

[4] http://httpd.apache.org/security/vulnerabilities_x.html, where x is either 13, 20, 22, or 23.

Out of the 100 CVE entries for Apache, 34 were categorised as *unassigned*, leaving 66 reports.

The situation for Tomcat is better. From 2008 onwards, the vulnerability reports[5] always contain the revision number of the fixing checkin. Conversely and unfortunately, none of the reports from 2007 or earlier have any such attribution, so that fully 68 out of the 89 CVEs for Tomcat are unassigned, leaving 21 CVEs, whose attribution to a fixing checkin, however, is absolutely certain.

Httpd and Tomcat both reside in SVN repositories. SVN also lacks the ability to retrieve changesets, but the situation is much improved over CVS, since SVN labels commits with their own unique revision number. Therefore, one can simply group all log file entries with the same revision number into one checkin.

We found 18,803 checkins in the Apache SVN, ranging from November 18, 1996 to March 10, 2011, and 17,688 checkins in the Tomcat SVN, ranging from May 12, 2001 to April 23, 2011.

4.3 Methodology

4.3.1 Vulnerability Fixes

Trends in vulnerability fix rates cannot be read directly off the checkins, since it is not a priori clear that every checkin is a fix. However, this is a reasonable assumption because of three observations:

- Each such checkin is usually *believed* to be a fix by the developer doing the checkin, even if the fix needs to be modified later;
- Checkins that are backports of vulnerability fixes to earlier releases constitute vulnerability fixes in themselves, since the fix will have to be adapted to the peculiarities of the earlier release; and
- The number of checkins for any given vulnerability is small (but see below for a discussion of actual results).

Plotting the number of vulnerability-fixing checkins on any given day against time directly gives only a confused picture. For example, plotting this number for Mozilla gives Fig. 4.3. While it is clear that fix rates have peaked in 2007—during the lifetime of Mozilla 2.0—and have declined since then, some smoothing would give a clearer picture.

Instead of plotting fixes directly against time, we plot the moving average and the moving average fraction. Both move a window of constant size over the data: the moving average computes the sample mean in each window, and the moving average fraction computes the ratio of two moving averages.

[5] http://tomcat.apache.org/security-x.html, where x is either 5, 6, or 7.

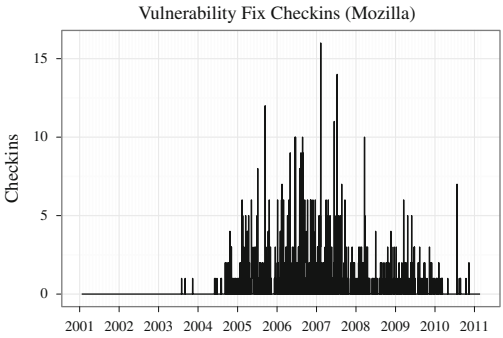

Fig. 4.3 Fixes against time for Mozilla. No trend is directly apparent

We plot moving averages for checkins with a window size of 365 days, therefore this moving average will start 1 year after the start of the original series and will average values of the previous year. We also plot moving average fractions in order to express vulnerability-fixing checkins as a fraction of all checkins, also with a window size of 365 days.

In order to analyse the distribution of checkins per day, we plot the number of vulnerability-fixing checkins on the x axis and the number of days with this number of checkins on the y axis on a log-log scale. We chose to bin checkins by day because it is a small enough bin size so that important effects would still be present in the data and not lost in larger bins.

4.3.2 Size of the Security Team

In order to investigate the size of the security team, we look at the number of unique committers in a given time window, in our case with a window size of 365 days. In addition, in order to see how many vulnerabilities a person fixes in a given year, we divide the number of vulnerabilities in the last year by the number of committers.

4.3.3 Red Queen/Blue King

Johnson et al. have used a "Red Queen" model to explain the power laws they found when they looked at the time between successive insurgent attacks with coalition fatalities in Iraq and Afghanistan: not only did the time between days with Alliance fatalities obey a power law, the exponent could also be predicted from the interval between the first two fatal days! [9]. If software security were such a Red Queen race, we should see the same power laws, and we should be able to tell who is winning the race by looking at the trend: if the time between successive vulnerability fixes is growing, the defending Blue King wins; otherwise, the attacking Red Queen wins.

They explained this regularity with a dynamic Red Queen model. In Red Queen models in evolutionary biology, an entity needs to adapt continuously in order to survive. In the context of insurgent attacks, Johnson et al. model the race between the Red Queen insurgents and the Blue King alliance with a random walk model, which is then responsible for the observed power laws.

More formally, if vulnerabilities are fixed at times $t_1, \ldots, t_N$, we then look at the sequence $\langle \delta_1, \ldots, \delta_{N-1} \rangle$, where $\delta_k = t_{k+1} - t_k$ for $1 \leq k < N$, and fit this sequence to the model $\log \delta_k = \log a + b \log k$.

4.4 Findings and Discussion

4.4.1 Vulnerability Fixes Over Time

Figure 4.4 (left) shows the moving average of vulnerability-fixing checkins for Mozilla, httpd, and Tomcat. We can see that httpd and Tomcat have had the same fix rates (within half an order of magnitude, or a factor of about 3.2), but Mozilla's fix rates are about two orders of magnitude higher.

As we suspected already from the raw data in Fig. 4.3, the fix rate for Mozilla peaked in 2007 and then declined. What was not apparent in that figure, however, is how dramatic the decline is. Fix rates decline from about 1.4 per day to under 0.1 per day, which is a stunning 93 % reduction. However, it is possible that most vulnerability-fixing checkins now occur in Mercurial repository, and not in the CVS. This is supported by looking at a moving average of all checkins (not shown), where we can see that not just the vulnerability fixes per day have gone down, but also the number of all checkins per day. There is very little activity left in the Mozilla CVS.

Having established that the strong decline in Mozilla is probably an artifact of repository usage, the next interesting feature of that graph is the incline and peak in 2007. The period with the highest vulnerability fix rates in Mozilla corresponds to the lifetime of Firefox 2 (released on 24 October 2006, end of life in December 2008) [14]. We know from other studies [13] that Firefox 2 is unlike other Firefox versions because almost none of its source code was inherited by later versions. Together with the high number of vulnerability fixes during Firefox 2's lifetime, we conclude that something in its architecture must have made it inherently unsafe to use so that it was completely phased out.

The general shape for httpd is similar to that of Mozilla, but the peak is in 2005, not in 2007, and the decline in fix rates is not as strong. In fact, one could argue that the fix rate has been about constant since 2007.

Tomcat is a comparative latecomer, but the fix rate seems to be stable, with an upward trend in 2010. It is apparent that the data is much grainier than the data for either Mozilla or httpd, simply because there are much fewer checkins, so each checkin represents a rather larger jump. As we can see, the fix rates rose until 2009, stayed essentially constant during 2009 and 2010, and took off in 2011.

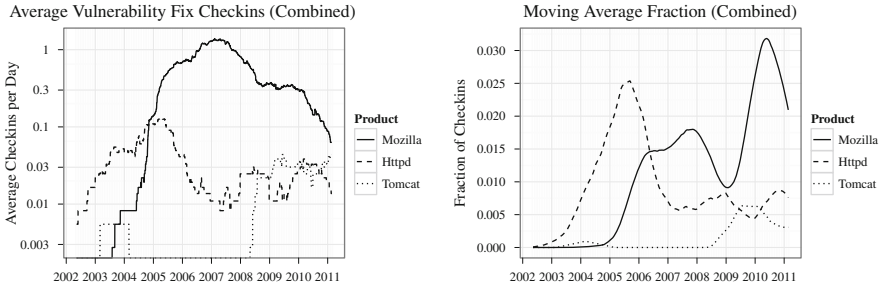

Fig. 4.4 Combined moving average of vulnerability-fixing checkins for all three products, log-linear scale (*left*), and combined moving average of fraction of vulnerability-fixing checkins in all checkins, for all three products, linear (*not* log-linear!) scale

In Fig. 4.4 (right), we see the moving average fraction of vulnerability-fixing checkins in all checkins. For Mozilla, we can see that vulnerability-fixing came into prominence by 2007, and, after a brief dip in 2009, rose to a new height in 2010, after which it fell again. We attribute the comparatively large proportion of vulnerability fixes in 2010 to a gradual move to the Mercurial repository, because vulnerability patches would have a proportionately greater need of being backported to the CVS than other checkins. For httpd, the long-term trend seems to go down, whereas for Tomcat, it seems to be rising.

There is no evidence in the data that vulnerability fix rates are declining overall, neither over time nor as a fraction of all checkins. The decline in Mozilla is explained by development occurring elsewhere, and the rates for httpd and Tomcat simply show no decline.

4.4.2 Distribution of Number of Checkins per Day

Figure 4.5 shows the distribution of the number of fixes per day on a log-log scale. It is evident that the data points lie very nearly on straight lines. We take this as evidence that all three distributions are heavy-tailed.[6]

If we view the set of vulnerabilities as a reservoir, there are *on periods* of mean length μ_{on}, in which new code is written and new vulnerabilities added to the

[6]Even though a linear regression on the model $\log \text{days} = \log a + b \log(\text{checkins} + 1)$ gives excellent p- and R^2-values, we cannot infer from this that the distribution obeys a power law. This is because (1) parameter estimation for power law distributions from linear regression is prone to large systematic biases, (2) the data do not span sufficiently many orders of magnitude for a reliable check, and (3) even with much data, power laws are very hard to distinguish from other heavy-tailed distributions such as the log-normal distribution [5]. Fortunately, the precise nature of the distribution is not important for this work, since we are here concerned with an empirical description and not with forecasting. The problems with estimating power laws with linear regression were brought to our attention by one of the anonymous reviewers.

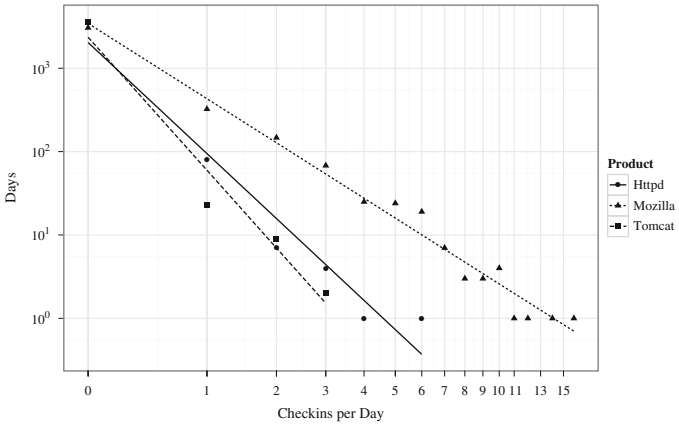

Fig. 4.5 Number of days versus number of checkins, on a log-log scale, together with their respective least-squares regression lines

reservoir, and *off periods* of mean length μ_{off}, in which vulnerabilities are fixed and hence removed from the reservoir. If vulnerabilities are added at unit rate during on periods and removed at a rate greater than $\mu_{\text{on}}/(\mu_{\text{on}} + \mu_{\text{off}})$ during off periods, then the number of vulnerabilities V will be heavy tailed with $P(V > x) \sim cx^{-(\alpha-1)}L(x)$, where c and α are constants and L a slowly varying function.[7] If we now also assume that at any day, a constant fraction of these available vulnerabilities will be fixed, it follows that the number of days will also be heavy tailed. (The idea for this model and the notation used is from a survey paper on heavy tail modeling [19, p. 1807]).

We can also explain the different slopes in Fig. 4.5: they are associated with project size. Mozilla is by far the biggest project, containing at the time of writing 95,617 files in a freshly checked out working directory, whereas httpd only has 21,136 (including tests; the pure source is just 9,300 files), and Tomcat 21,726. Larger project size is thus associated with a shallower slope and this is intuitively pleasing, since we would expect that larger projects are also more active, and hence there are bigger chances that there are two or more vulnerability fixes on any given day. But of course, three projects are too few to enable a meaningful statistical analysis of this relationship.

4.4.3 Number of Checkins per Vulnerability

Figure 4.6 shows the number of checkins per vulnerability for Mozilla, whereas we show the much smaller datasets for Httpd and Tomcat in Table 4.1.

[7] A real function L is slowly varying if for all real $c > 0$ we have $\lim_{x\to\infty} L(cx)/L(x) = 1$.

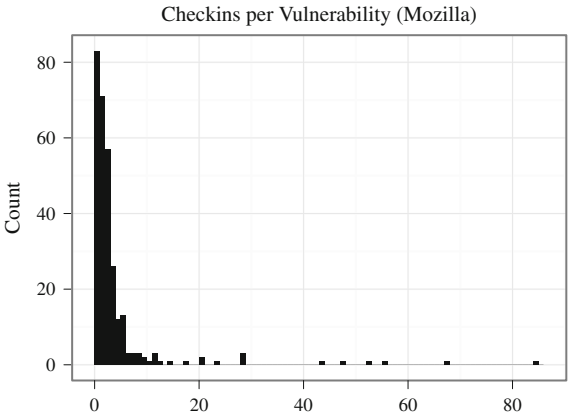

Fig. 4.6 Checkins per vulnerability for Mozilla

Table 4.1 Number of checkins per vulnerability (httpd and Tomcat, with the first four Mozilla values for comparison)

Checkins				
Product	1	2	3	4
Httpd	33	33		
Tomcat	1	14	4	2
Mozilla	83	71	57	26

Some Mozilla vulnerabilities took a large number of commits to fix. To take the most extreme example, MFSA 2006-64 ("Crashes with evidence of memory corruption (rv:1.8.0.7)") took a record 85 commits to fix. Looking at that MFSA, we find that this vulnerability is host to 29 bug identifiers, and that this MFSA is in fact a blanket vulnerability under which to file any bug report that is concerned with crashing due to memory corruption: there are six bug identifiers associated with "crashes involving tables", one on "heap corruption using XSLTProcessor", four involving the "JavaScript engine", fully 17 because of "crashes involving DHTML", and one more which seems to have been a regression. In other words, *any* memory corruption issue at this time was seen as an instance of this vulnerability.

Other vulnerabilities with large numbers of checkins are a consequence of Mozilla's bug-fixing process. A developer first commits a fix to the trunk of the respective product, and pushes a patch out to a *try server*. On the try servers, a large test suite is run on the patched product. If a test fails and a patch is held responsible for the failure, that patch is undone, and a new fix attempted. Due to the repository organisation, this backout counts as a new commit, and will again carry the bug ID in the commit message.

After having made it through the try server, the patched product is upgraded to mozilla-central (recently renamed to mozilla-incoming), a large repository, where all products are again subjected to tests. Again, if a product fails a test, and if a patch is held responsible, the patch is again undone, and a new solution needs again to be attempted.

It is curious that almost all vulnerabilities in Tomcat need two commits to fix. For example, CVE 2010-1157 is fixed in r936540, which was committed on April 22, 2010 at 00:12:05 +0200, and in r936541, which was committed about 90 s later. Analysis reveals however that the both checkins are backports of a fix committed on the trunk (for the then unreleased Tomcat 7). So the large number of vulnerabilities needing two commits is simply because the fixes are committed separately for the different versions.

4.4.4 Size of Security Team

Figure 4.7 shows the size of the security team in the last 365-day period (left) and the number of vulnerability fixes per committer in that period (right). It is difficult to tell visually how well the size of the security team (left graph in Fig. 4.7) tracks the number of vulnerability fixes (Fig. 4.4), but the linear correlation coefficients are clear: there is a strong association between the two quantities for Mozilla and Tomcat ($\rho = 0.77$ and 0.92, respectively), and a very weak one for httpd ($\rho = 0.22$). See also Fig. 4.4.

The findings for Mozilla and Tomcat mean either that there is a staffing process in place that adapts the security team to apparent needs, that there are always enough people there to fix vulnerabilities, or that the original authors of code are responsible for any fixes in that code. The number of vulnerability fixes per committer (Fig. 4.7, right), decreasing for Mozilla, almost constant since 2009 for Tomcat, support the second hypothesis. This is also the reason why the declining rate of commits per committer for Mozilla cannot be used to support increasing marginal cost of fixing vulnerabilities.

The findings for httpd are striking. The lack of correlation between vulnerability fixes and number of people committing such fixes means that there is *no effort to staff the security team proportionally to the needs*, which in turn means that *the amount of work expected of each committer will be proportional to the amount of vulnerabilities discovered*. That is a potentially dangerous situation, since it can lead to overload. On the other hand, the number of checkins per vulnerability is very low, indicating a good quality of vulnerability fix checkins, since they generally do not need to be revised.

The absence of a correlation between vulnerability fixes and fixers could indicate that there are always enough people available to fix vulnerabilities. In that case, it would not make sense to worry about staffing. It could also be that, from a software development perspective, fixing vulnerabilities is no big deal. If fixing a vulnerability is no more difficult than fixing any other bug, then the same people who fix ordinary bugs can also fix vulnerabilities, and it would again make no sense to worry about staffing. Or it could indicate that there is no planning in place to adapt the security-conscious staff to needs. To be fair, httpd is an open-source project, where staffing is likely not to be the result of a cental planning process.

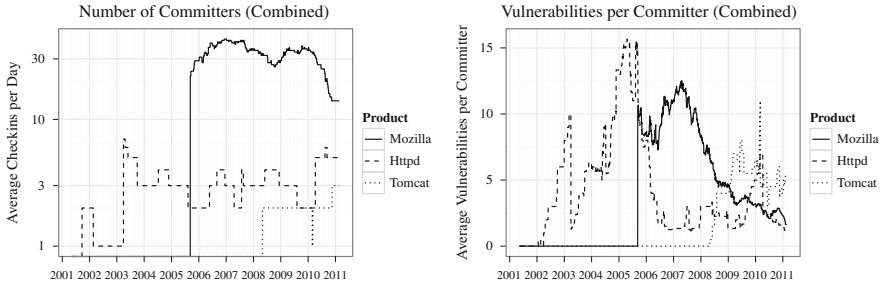

Fig. 4.7 Moving sum of unique vulnerability fix committers on log-linear scale (*left*), and moving rate of commits per committer on linear scale (*right*)

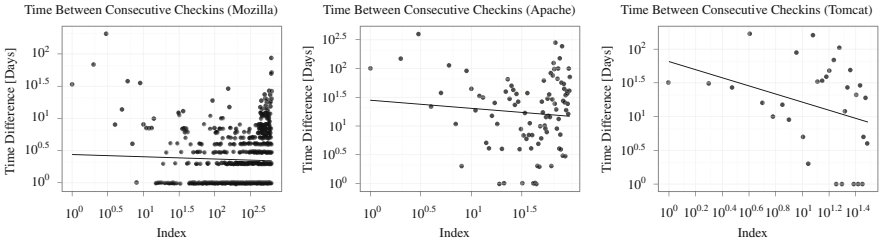

Fig. 4.8 Time between consecutive checkins for Mozilla (*left*), httpd (*middle*), and Tomcat (*right*), on log-log scale with least-squares regression line

4.4.5 Red Queen/Blue King

As explained in Sect. 4.3.3, in a race between a constantly adapting Red Queen and a defending Blue King, the time between consecutive events (vulnerability-fixing checkins in our case) should obey a power law. When we actually plot the time between consecutive checkins, we see in all cases a downward trend that is, however, far removed from the uncannily precise fits in Johnson et al.'s paper [9]. The p-values (0.37 for Mozilla, 0.38 for httpd, and 0.07 for Tomcat) are as disappointing as the R^2 values (2.8×10^{-4} for Mozilla, 2.4×10^{-3} for httpd, and 0.077 for Tomcat) (Fig. 4.8).

With this data and this analysis, we cannot confirm a Red Queen race.

4.5 Related Work

As far as we know, we are the first to investigate vulnerability fix rates over time; none of the papers below address this problem. The closest work is Ozment and Schechter's study of vulnerabilities in OpenBSD [16]. Their aim, like ours, is to find out whether software security gets better as the software ages (wine) or

worse (milk). In order to analyse this, Ozment and Schechter look at the number of *foundational vulnerabilities*, that is, the vulnerabilities that have been in the software from the first release. They conclude that software security in OpenBSD is indeed getting better with age in the sense that fewer and fewer reported vulnerabilities are foundational. Similarly, Massacci et al. presented a study on Firefox evolution, and looked especially at the question of whether vulnerabilities were inherited or foundational [13], but did not compare Firefox with other software.

In our study, we have a different viewpoint: we look at whether a constant effort in fixing vulnerabilities is bringing diminishing returns. In doing this, we not only consider foundational vulnerabilities (which by definition can only become fewer), but also vulnerabilities that are introduced as the software evolves.

In the area of software engineering in general, one of the earliest attempts to show bug fixes—not necessarily vulnerabilities—is in software visualisation work; see, for example, work by Baker [1] or by Ball and Eick [2]. Such visualisations can display the fix-on-fix rate (the rate at which fixes need to be fixed), but not the fix rates themselves.

There is much research on bug introduction and fix rates, but that research in general does not look at how these rates change over time. For example, Phipps analysed bug and productivity rates for Java and C++ [17], and Kim et al. [10] worked on bug introduction rates per author, based on earlier work by Śliwerski et al. identifying bug-introducing changes [22].

Rescorla modeled the vulnerability lifecycle in order to determine if looking for vulnerabilities was a good idea [18], whereas Massacci et al. looked at Firefox vulnerabilities, but with the intent of determining which source of vulnerability information would enable one to answer which type of research question [12].

Frei looked at the dynamics of insecurity by painstakingly analysing over 30,000 vulnerabilities and the paths they take through the security ecosystem [7], and Frei e al. analysed vulnerabilities specifically to find whether "responsible disclosure" works [8].

Neuhaus et al. studied Mozilla vulnerabilities in order to predict so far unknown vulnerabilities [15] and Schryen studies many OSS projects in order to find out whether open-source security is a myth [21].

Heavy tailed distributions have been extensively studied for network structures, the earliest example being Price's 1965 discovery that networks of scientific citations obey a power law [6]. A more mathematical treatment of heavy tail modeling is given in the survey article by Resnick [19].

A particularly interesting application of heavy-tailed distributions specifically to the problem of human irrationality, with obvious applicability to problems of security, is given by Maillart et al. [11]. Here the authors model the appearance of a security vulnerability as an item entering a priority queue, and show that in various plausible scenarios, the time to complete a given work item will be heavy-tailed, and sometimes even obey a power law. The authors have confirmed this model using anonymised data from Google about users with outdated web browsers, but it is plausible that the same model will also hold for developers fixing vulnerabilities.

4.6 Threats to Validity

4.6.1 Study Size

We are looking at only three software projects. It is possible that looking at more (and more diverse) projects would give different results. However, all such studies would be restricted to open source software, and in that area, the chosen projects are representative in terms of market share and attack surface.

4.6.2 Biases in Data Selection

Bird et al. have described the bias that is present in bug-fix datasets, and the consequent bias that results when these datasets are used for prediction [3]. It is certainly true that we have not been able to map all vulnerabilities to their fixing checkins. However, since we are here only concerned with an empirical description, not with prediction, we argue that the data we have is for the most part representative.

4.6.3 Unknown Noise in Data

From previous studies we know that vulnerability information can be noisy, and that the noise is often not even quantifiable. This is because we lack information about the exact process by which vulnerability information is acquired and published. In our case, this information comes directly from the vendor, and since we are only concerned with the time when a vulnerability has been *fixed* (as opposed to when it has been *discovered*), and since we get this information directly from the source code repository, we are confident that our results are robust.

4.6.4 Confusing Data

The picture emerging from analysing these three software systems is far from unified; instead, each system has its own idiosyncrasies. This may create a confusing data set where inter-system comparisons may be difficult. Still, we believe that the phenomena we are investigating should be robust with respect to slight differences in the precise meanings of words like ‘vulnerability’ or ‘fix’.

4.7 Conclusion and Future Work

In this work, we gave evidence that vulnerability fixing is not an activity that can be modeled by an increasing concave function, as we would have in a standard cybersecurity economic model. Instead, the situation is different for each of the three projects: for Mozilla, the vulnerability fix rate decreases as predicted (but there has been a switchover to another repository, so this decrease is suspect, since the decrease is abrupt and is concurrent with the switchover), for httpd it stays constant, and for Tomcat it increases. Also, vulnerability fixing is not a Red Queen race since we did not find any evidence of the resulting power laws. Whatever the reasons for these findings, they are evidence that standard cybersecurity economic models are not easily applied to software security.

We unexpectedly found evidence that the distribution of the number of fixes per day is heavy tailed. While we can at this point only speculate about the causes for this relationship, as far as we know we are the first to notice this.

Looking at the number of people fixing vulnerabilities, we found that the pool of people available for fixing vulnerabilities changes proportionally with the demand for Mozilla and Tomcat, but not for httpd. Generally, vulnerability fixes have good quality, because they do not need to be revised.

In future work, we will:

- Investigate probable models for the heavy tailed distributions exhibited in Fig. 4.5;
- Find models that predict fix rates for the three projects; and
- Explore the Red Queen races in order to see if perhaps the predicted relationship holds in subsets of the data.

Acknowledgements We thank Sandy Clark, Jonathan M. Smith and Matt Blaze for constructive discussions and for finding reference [9]; Brian Trammell for suggesting the title of this chapter; the Tomcat security team for answering our questions; Christian Holler for information about the Mozilla development process; Dominik Schatzmann for excellent suggestions on early drafts of the chapter; Thomas Maillart for excellent and fruitful discussions and a gentle pointer towards reference [11]; and the anonymous reviewers for raising many excellent points and making helpful suggestions.

References

1. Baker, M.J., Eick, S.G.: Visualizing software systems. In: Proceedings of the 16th International Conference on Software Engineering, ICSE'94, Sorrento, pp. 59–67 (1994)
2. Ball, T., Eick, S.: Software visualization in the large. Computer **29**(4), 33–43 (1996)
3. Bird, C., Bachmann, A., Aune, E., Duffy, J., Bernstein, A., Filkov, V., Devanbu, P.: Fair and balanced? Bias in bug-fix datasets. In: Proceedings of the ESEC/FSE'09, Amsterdam, pp. 121–130 (2009)
4. Carroll, L.: Through the Looking-Glass. Macmillan and Co, London (1871)

5. Clauset, A., Shalizi, C.R., Newman, M.E.J.: Power-law distributions in empirical data. SIAM Rev. **51**, 661–703 (2009)
6. de Solla Price, D.J.: Networks of scientific papers. Science **149**(3683), 510–515 (1965)
7. Frei, S.: Security econometrics – the dynamics of (in)security. ETH Zürich, Dissertation 18197, ETH Zurich (2009)
8. Frei, S., Schatzmann, D., Plattner, B., Trammel, B.: Modelling the security ecosystem – the dynamics of (in)security. In: Anderson, R. (ed.) Workshop on the Economics of Information Security (WEIS), Cambridge (2009)
9. Johnson, N., Carran, S., Botner, J., Fontaine, K., Laxague, N., Nuetzel, P., Turnley, J., Tivnan, B.: Pattern in escalations in insurgent and terrorist activity. Science **333**(6038), 81–84 (2011)
10. Kim, S., Zimmermann, T., Pan, K., Jr., E.J.W.: Automatic identification of bug introducing changes. In: Proceedings of the 21st IEEE/ACM International Conference on Automated Software Engineering, Tokyo, pp. 81–90 (2006)
11. Maillart, T., Sornette, D., Frei, S., Duebendorfer, T., Saichev, A.: Quantification of deviations from rationality with heavy-tails in human dynamics. ArXiv e-prints (2010)
12. Massacci, F., Nguyen, V.H.: Which is the right source for vulnerability studies? An empirical analysis on Mozilla Firefox. In: Proceedings of the 6th International Workshop on Security Measurements and Metrics, MetriSec'10, Bolzano, pp. 4:1–4:8 (2010)
13. Massacci, F., Neuhaus, S., Nguyen, V.H.: After-life vulnerabilities: a study on Firefox evolution, its vulnerabilities, and fixes. In: Proceedings of the ESSoS'11, Madrid. Lecture Notes in Computer Science, vol. 6542, pp. 195–208 (2011)
14. Mozilla Foundation: Mozilla-Announce mailing list. https://lists.mozilla.org/listinfo/announce (2012)
15. Neuhaus, S., Zimmermann, T., Holler, C., Zeller, A.: Predicting vulnerable software components. In: Proceedings of the 14th ACM Conference on Computer and Communications Security, Alexandria, pp. 529–540 (2007)
16. Ozment, A., Schechter, S.E.: Milk or wine: does software security improve with age? In: Proceedings of the 15th Usenix Security Symposium, Vancouver, pp. 93–104 (2006)
17. Phipps, G.: Comparing observed bug and productivity rates for Java and C++. Softw. Pract. Exp. **29**, 345–358 (1999)
18. Rescorla, E.: Is finding security holes a good idea? IEEE Secur. Priv. **3**(1), 14–19 (2005)
19. Resnick, S.I.: Heavy tail modeling and teletraffic data. Ann. Stat. **25**(8), 1805–1869 (1997)
20. Rue, R., Pfleeger, S.L.: Making the best use of cybersecurity economic models. IEEE Secur. Priv. **7**, 52–60 (2009)
21. Schryen, G.: Is open source security a myth? What does vulnerability and patch data say? Commun. ACM **54**(5), 130–140 (2011)
22. Śliwerski, J., Zimmermann, T., Zeller, A.: When do changes induce fixes? In: Proceedings of the Second International Workshop on Mining Software Repositories, St. Louis, pp. 24–28 (2005)

Part II
Economics of Information Security

Chapter 5
An Empirical Study on Information Security Behaviors and Awareness

Toshihiko Takemura and Ayako Komatsu

Abstract In this chapter, we investigate some key factors which have effects on employees' behaviors in violating rules which are related to information leaks given the condition that the behaviors are totally prohibited by their organization. By using collected data from a survey that we conducted, and employing a stepwise logit model, we analyze the relationships above. The primary results are as follows: First of all, myopic cognition and hyperopic cognition measured by the CFC scale have effects on the behaviors of violating organizational rules in almost all cases. Next, in many cases, individuals whose information security awareness is higher tend not to violate the rules. Third, the behavior of violating the rules is independent of the size of the organization, and is not related to the degree of workplace satisfaction and the evaluation toward the managers in some cases. Fourth, in an organization in which permanent employment is implemented, individuals tend to violate the rules. It is not easy to control psychological factors such as an individual's attitude toward risk. Conversely, the factors regarded as organizational attributes, such as the degree of workplace satisfaction or the employment system utilized, may be controlled by designing the appropriate organizational environment. Consequently, we consider that it may be effective to improve information security awareness by information security education and training.

T. Takemura (✉)
The Research Institute for Socionetwork Strategies, Kansai University, 3-3-35, Yamate-cho, Suita, Osaka, Japan
e-mail: a084034@kansai-u.ac.jp

A. Komatsu
Security Economics Laboratory, IT Security Center, Information-Technology Promotion Agency, 2-28-8 Hon-Komagome, Bunkyo-ku, Tokyo, Japan
e-mail: a-koma@ipa.go.jp

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_5, © Springer-Verlag Berlin Heidelberg 2013

5.1 Introduction

The primary interests of empirical studies on information security are clarifying the appropriate level of information security investment, and describing effective technical and managerial measures. These studies are useful for managers who introduce and implement organizational information security measures. So, the targets for these research articles are organizations in which the managers introduce and implement security measures. However, many of these articles have missed the important point. They discussed and analyzed only one-sided measures from the managers' standpoints, excluding consideration of the employees (users) who conform to these measures. This one-sided measure sometimes may not work well unless the users understand the meaning of the measures. The reason is that some users have grievances against the managers and the security measures. For instance, as mentioned in previous literature [2, 3, 37], the users sometimes might not comply with the measures or may tend to put their daily activities ahead of the measures even if the policy is enforced in the organization.

To tackle these issues, in recent years some empirical studies on information security take the approach from the users' standpoints. In these types of research, one can either discuss the ability of managers in the organization to implement more effective measures by analyzing the users' information security awareness, or discuss the prevention of undesirable behaviors by analyzing computer abuse problems and insider security contravention. This chapter belongs to the latter type of research. Therefore, we briefly introduce some related literature and then show the significance of this chapter. Much of the literature is supported by behavioral sciences. For example, there is the Theory of Planned Behavior (TPB) and the General Deterrence Theory (GDT).[1] TPB is one of the most widely successful and applied frameworks to explain human behavior and was suggested by Ajzen [1]. TPB shows that the best way to predict an individual's behavior is by examining how that individual intends to behave. In TPB, behavioral intentions (how much effort one is willing to exert to perform a given action) influence a certain actual behavior. The behavioral intentions are formed from three determinants: attitude toward the behavior, subjective norms, and perceived behavioral control. TPB has been explicitly applied to software piracy problems [9, 27], non-work-related computing [28], Internet abuse [10, 16, 42], security policy compliance [5, 7, 15, 44] and insider security contravention [43]. On the other hand, GDT has been widely used in the study of criminal and antisocial behavior, and is a well-established theory within the criminology field. GDT explains how security measures implemented by organizations rely primarily on technology without considering other factors, such as people and processes. Previous computer abuse and misuse studies have been mainly based on GDT [14, 24, 35]. We need to note that in the studies based on TPB or GDT, only a few researchers clarify the relationship between behavioral

[1] These studies provide good reviews about these theories regarding information security [23, 38].

intentions with regard to information security and actual behaviors. Based on the original TPB or GDT, researchers implicitly discuss the assumption that an individual exactly behaves only as he intends to behave. However, Komatsu et al. point out that the behavioral intent does not necessarily lead to the actual behavior, from the analysis of behavior regarding bot measures [22].

In addition, studies with regard to insider threats have been actively conducted with a focus in social psychology. Insider threats strongly get involved in crime. For instance, Pfleeger et al. present a framework for describing insiders and their actions based on four factors: the organization, the environment, the system, and the individual [29–31]. Greitzer et al. build and analyze a predictive model for insider threat mitigation [17,18]. Cappelli et al. model and analyze insider sabotage activity by incorporating both cyber and psychosocial data within an anticipatory decision framework called system dynamics [8].[2]

There are a few studies that approach from the viewpoint of behavioral economics, too. In approaches from TPB and GDT, human behavior is assumed to be rational, but rationality is not necessary in behavioral economics. We will discuss whether or not a certain kind of human behavior, fraud, is rational or irrational in the following section. Takemura models employees' violations of organizational rules related to information leaks in the organization by employing logit regression equations [37]. This model provides straightforward results and possesses a strong ability to predict.

It is needless to say that in many cases, users' behaviors mentioned above are inconsistent with the decisions of their organizations. If an individual commits fraud, he might achieve his individual purpose, but his behavior would be disadvantageous for his organization. Therefore, managers in the organization must pay attention to such individuals. In each study, it is found that the psychological factors, such as attitude toward risk and the individual's working environment, influence their intention and/or behaviors directly or indirectly. The organization is able to change the working environment, but it would still remain an issue that it is difficult to control the individual's psychological factors.

This chapter includes a new breed of behavioral modeling based on Takemura's model [37] incorporating some factors from TPB and other new factors. The purpose of this chapter is to determine key factors which have effects on the violation of rules by employees related to information leaks, given the condition that the behaviors are prohibited totally through organizational measures. This condition enables us to discuss the effectiveness of organizational measures.

This chapter consists of the following sections. In the next section, we explain our behavioral modeling and the survey data. Section 5.3 shows the results of analysis and the implications. Finally, in Sect. 5.4 we summarize our analysis and discuss future work.

[2]System dynamics is a method for modeling and analyzing the holistic behavior of complex problems as they evolve over time. System dynamics has been used to gain insight into some of the most challenging strategy questions facing businesses and government for several decades.

5.2 Framework

5.2.1 Behavioral Modeling

When information security policies or organizational rules are established, almost all employees comply with these rules. However, some employees violate the rules, unfortunately. In addition, it is pointed out that even if the rules have a compelling force, an individual might occasionally regard violation of the rules not as a serious problem for the purpose of completing daily activities [37]. This misjudgment sometimes becomes a trigger for information security accidents such as an information leak. Of course, to some degree, establishing the policy and information security education/training are able to prevent information security accidents [32]. For enhancing the effects of these measures more, we investigate the factors effecting violation of the rule by the employees, including managers, and the need to implement the measure to reduce violation of the rule. Therefore, this chapter focuses on the employee's behavior of violating or complying with the organizational rules.

The primary research question of this chapter is what the determinants of an employee violating organizational rules are. As mentioned in Sect. 5.1, TPB suggests that the behavioral intentions, which influence a certain actual behavior, are formed from three factors, labeled "attitude toward the behavior", "subjective norms", and "perceived behavioral control". The "attitude toward the behavior" is the degree to which the person has a favorable or unfavorable evaluation of the behavior in question, "subjective norm" is the influence of social pressure that is perceived by the individual to perform or not perform a certain behavior, and "perceived behavioral control" is the perceived ease or difficulty of performing the behavior, respectively. In addition to TPB, the theory of fraud triangle suggested by Cressey [13] is one famous theory related to such violation of the rules and fraud in criminal psychology [41]. The theory of fraud triangle consists of three conditions generally present when fraud occurs: incentive/pressure, (perceived) opportunity, and attitude/rationalizations. This implies that anyone may commit a fraud if the three conditions are satisfied at the same time. Each theory has something in common, such as attitude toward the behavior, assessment from persons involved, and individual circumambient environment. Furthermore, human behavior is assumed to be rational in both theories.

Though we support the effect of the psychological factors above, we query the assumption that the human behavior of violating the organizational rule is rational. For example, suppose that an employee would be fired or punished if he violates an organizational rule. Would he violate the rule for the purpose of completing his daily activities at that time? If he is rational, he would not violate the rule because of the risk of being fired. In this case, his behavior may be rational in the short term but not in the long term. That is, the behavior of violating the organizational rule is myopically rational, but the behavior may not be rational from the hyperopic view. From the viewpoint of implementing information security measures, it seems to be

important to make the assumption that human behavior is rational or not, or that they have a myopic or hyperopic view. So, we check whether or not the behavior results from myopic and hyperopic cognition in this chapter.

Based on the factors used in these theories, we incorporate key factors (attitude, motivation toward the behavior information security awareness and workplace environment) into our behavioral model.

Attitude Attitude represents the degree to which the individual has a favorable or unfavorable evaluation of the behavior, e.g., risk attitudes [6,12], or consideration of future consequences (CFC) [34].

Motivation toward the behavior Motivation toward the behavior is the driving force by which an individual achieves his/her goal. As general motivational strategies or specific motivational appeals, there are five factors: monetary rewards, assessment from peers, self-actualization, morality, and pleasantness [40].

Information security awareness Information security awareness represents the measure of an individual's evaluation and/or knowledge of information security. The concept of awareness is an important factor, which enables exogenous control through education or training of the members in the organization effectively [4, 39].

Workplace environment Workplace environment consists of two elements of the organization to which the individual belongs. One is the subjective element, e.g., the degree of workplace satisfaction, or the individual's evaluations of the information security manager and the organizational measures [33]. The other is the objective element, e.g., working pattern, the scale of organization, or the incentive system for employees which is implemented in the organization [37].

To answer this research question, we employ the following logit regression equation.[3]

$$\text{logit}(p_j) = \log \frac{p_j}{1 - p_j} = a + \mathbf{X}_b\mathbf{b} + \mathbf{X}_c\mathbf{c} + \mathbf{X}_d\mathbf{d} + \mathbf{X}_e\mathbf{e} + \mathbf{X}_f\mathbf{f} \qquad (5.1)$$

where p_j represents the probability that an individual violates the rule j. In addition, $\mathbf{X}_b, \mathbf{X}_c, \mathbf{X}_d, \mathbf{X}_e$ and $\mathbf{X}_f$ represent vectors of attitude, motivation toward the behavior, awareness, workplace environment and the individual attributes, respectively.

By using the (binary) logit regression equation in (5.1), we can assess the effects of the explanatory factors on the relative risk of outcome. In this case, the logistic transformation can be interpreted as the logarithm of odds of violating the rule vs. complying with the rule.

Here, we briefly explain the process to estimate the coefficients in (5.1) [19]. We employ a stepwise procedure for deletion of variables from the model (backward

[3]Generally, a behavioral model using a logit regression equation is devoted to explaining and predicting human behavior and has been used in the various fields for a long time.

selection procedure). This procedure is based on a statistical algorithm that checks the importance of the variables and excludes them on the basis of a fixed decision rule. In other words, employing this stepwise selection procedure can provide a fast and effective means to screen a large number of variables and to fit a number of logit regression equations simultaneously. This selection fits the full model of all explanatory variables at the first step and removes the least-significant term, then re-estimates when it is insignificant in subsequent steps. In other words, the variables deleted in the selection process are not significant and are not affecting factors to the explained variable.

5.2.2 Methodology

To test the relationships implied by the model in (5.1), we conducted a Web-based survey for data collection.

We conducted a Web-based survey entitled "Survey on Japanese workers' awareness and behavior to information security measures" in March 2011. This survey focuses on exploring workers' information security awareness and behaviors and has been annually conducted since 2009. Subjects of this survey are Japanese people who have been working for more than 2 years in the same company. The number of survey items is more than 60, including individual attributes such as gender and annual income. For instance, the survey contains questions on whether or not organizational measures are implemented, and questions regarding their information security awareness and behavior. This survey includes 1,800 respondents.

A Web-based survey method inescapably contains certain weakness of data collection. A Web-based survey is well-used in the field of marketing, but has Internet bias. In other words, the data may not guarantee representativeness of the intended population because the survey is not necessarily based on a random sampling. Unfortunately, this problem has not been solved yet [11].[4] Therefore, we interpret and analyze data from the population of Japanese people registered with the Internet survey company. In addition, we presume that this collected data is useful for reasonable analysis.[5]

5.2.2.1 The Behavior of Violating Organizational Rules

There are various organizational measures to prevent information leaks. In this chapter, we pick a few of these behaviors (Behavior-1: Bringing out private

[4] We point out the following characteristics: (1) we can obtain the desired sample size for statistical analysis; (2) imposing conditions on attributes of respondents beforehand has a predilection for a Bayesian approach; and (3) because the Web-based survey is conducted agilely, it is easy to collect data set for analysis.

[5] Of course, we do not intend to ignore this statistical problem. We expect that future studies on the representativeness of data from Web-based surveys will be promoted.

Table 5.1 Cross-tabulation between implementation status and individual experiences

		Individual experiences		
Behaviors	Status	I have experience (I have experience in violating the rule)	I have no experience (I always comply with the rule)	Total
Behavior-1	Totally prohibited	102	685	787
	Unprohibited	103	160	263
Behavior-2	Totally prohibited	55	662	717
	Unprohibited	93	202	295
Behavior-3	Totally prohibited	56	918	974
	Unprohibited	165	181	346
Behavior-4	Totally prohibited	80	578	658
	Unprohibited	278	237	515
Behavior-5	Totally prohibited	54	854	908
	Unprohibited	120	180	300
Behavior-6	Totally prohibited	38	501	539
	Unprohibited	126	162	288

customer data by using portable devices. Behavior-2: Attaching private customer data to e-mail. Behavior-3: Accessing non-work-related websites such as 2channel at the office. Behavior-4: Forwarding office e-mail to private addresses. Behavior-5: Installing software used at home on office computers. Behavior-6: Bringing a company laptop outside) [25,37]. According to the information security white paper in Japan [20], many companies believe that the route of virus infection is through portable devices such as USB memory. Additionally, installing software used at home on office computers is relevant to software piracy. Erroneous sending of e-mail or accessing non-work-related websites is also a trigger of information security accidents. Because these measures enable the prevention of information security accidents such as information leaks, many Japanese companies recently established and implemented some the measures above. It is thought that information security or system managers can forcibly have control over employees by implementing these measures. The question is, though, would the employees comply with these measures?

Table 5.1 shows cross tabulation between implementation status and individual experiences.[6] If the "Behavior" is totally prohibited within the organization by a measure, the implementation status is "Totally prohibited." If there are no rules in the organization, the status is "Unprohibited." Individual experience is whether or not the "Behavior" is experienced. If the measures are implemented, the option "I have experience (resp. I have no experience)" means "I have experience in violating the rule (resp. I always comply with the rule)."

[6]Because some respondents select "I do not know whether or not the measures are prohibited within the organization" or "the measures are prohibited with some conditions within the organization" in the survey, these respondents are excluded.

Irrespective of implementing the organizational measures, more than half of respondents have no experience in violating or always comply with all rule except for forwarding office e-mail to private addresses when implementation status is unprohibited. On the other hand, about 6–13 % of respondents had experience violating rules even if the behaviors are totally prohibited by the organization.

In this chapter, we focus on their behaviors given that the behaviors are totally prohibited by organizational measures. Thus, descriptive statistics are calculated by the subsample of the survey (the sample size is 1,564), not the full sample.

5.2.2.2 Attitude

Attitude relates mainly to the degree to which an individual has a favorable or unfavorable evaluation of a behavior. A positive attitude toward the behavior of violating a rule increases that behavior. Among various concepts of attitude, the concept of risk has been successfully used in theories of decision making in economics, financial engineering, and other sciences. So, we introduce the degree of risk aversion and risk tolerance as risk attitude.

The survey has some questions asking the amount of "certainty equivalent", the cost to gamble for an uncertain profit such as pricing lotteries and/or desired insurances for the damages from a robbery. From the amount of the certainty equivalent that respondents reveal, we can calculate their degree of risk aversion (on lottery and insurance) based on the BMD method. In this chapter, we assume situations where there is a lottery with a 1 % chance of winning 100,000 JY and a 99 % chance of winning nothing, and where there is a 1 % chance of being robbed of 100,000 JY. Figure 5.1 shows the distribution for the degree of risk aversion. The distributions of Fig. 5.1 show that many of the respondents are risk-averse on the lottery because the degree of risk aversion is positive, and that they are adversely risk-loving on the insurance. This implies that their attitudes toward risk vary by conditions, such as the probability and the situation. In some ways, this result is consistent with the Prospect Theory suggested by Kahneman and Tversky [21].

In addition, the survey has one question asking the degree of risk tolerance. The risk tolerance demonstrates the level of risk that the individual can perceive, or the degree of loss that they can receive. Concretely, we ask the following hypothetical question: Now let's assume that your computer at home would be at high risk of becoming infected with computer virus unless you install the latest anti-virus software on the computer. You have the option to purchase and install the latest anti-virus software on your computer or do nothing. Given the probability of virus infection (0.1, 1, 2, 5, 20, 30, 50, 70, 80, 99 and 100 %), the respondents compare option "A" (implementing the measure) with option "B" (do nothing). Then, we can conjecture probability they would prefer to implement the measure.[7] Figure 5.2 shows the distribution for the degree of risk tolerance.

[7]If the respondent selects option "B" when the probability is 99 %, we assume that he tolerates all the risks.

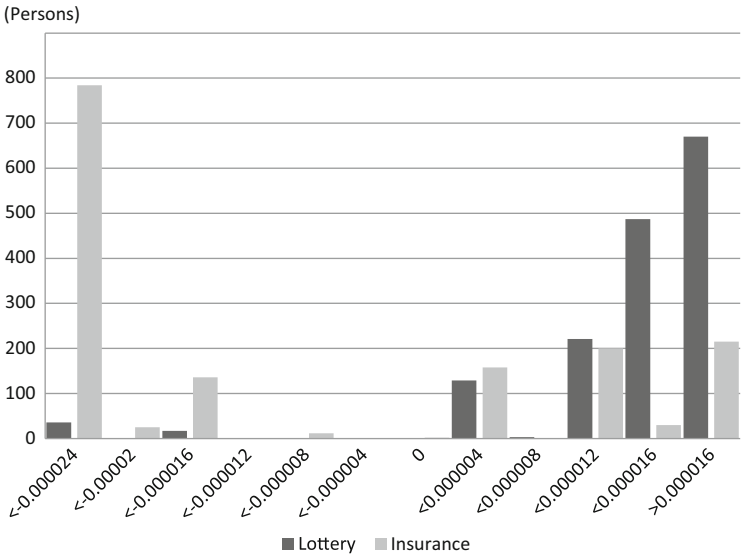

Fig. 5.1 The degree of risk aversion

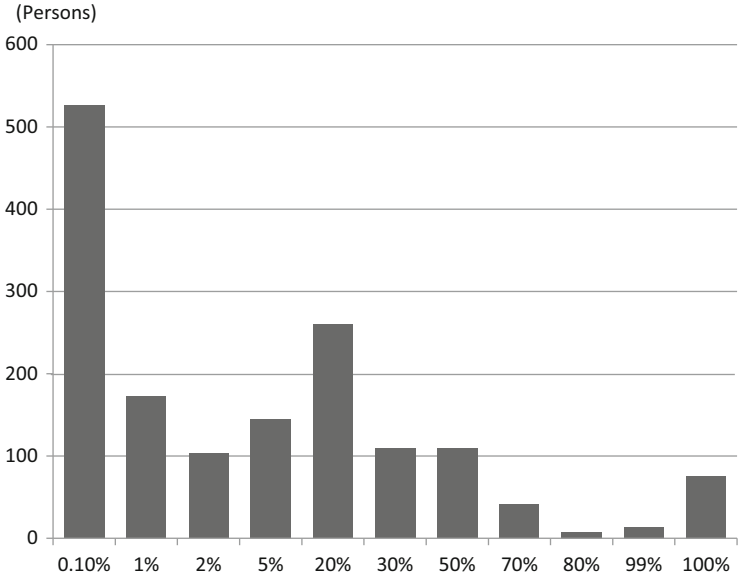

Fig. 5.2 The degree of risk tolerance

If the probability of getting a virus is lower than 1 %, about 44.7 % of respondents answer to implement the measure. On the contrary, about 4.86 % of respondents answer to not implement the measure even if the probability is 99 %. Figure 5.2 shows that most respondents cannot tolerate the risk of virus infection.

As the other concept of attitude, we introduce the CFC scale used in the field of psychology. The CFC scale is scored so that a higher score indicates a greater consideration of future consequences. To create the CFC scale, we include 12 statements in the survey, (for example, "I consider how things might be in the future, and try to influence those things with my day to day behavior") based on the previous study [34], which are measured on a five-point Likert scale. Then, by using factor analysis with promax rotation to the questions, two factors are assumed; myopic and hyperopic cognition.[8]

5.2.2.3 Motivation Toward Behaviors

It is generally agreed that individual performance depends on motivation in addition to ability and working conditions. In order to measure motivation, we introduce an importance indicator on five factors (monetary rewards, assessment from peers, self-actualization, morality, and pleasantness) with regard to activity, which was used in the previous study [40]. Each factor is closely related to the conditions in the theory of fraud triangle.

In the survey, we directly ask the following question: Now let's assume that you do something. How much importance of the following items (1: to gain money, 2: to be assessed by peers or neighbors, 3: to achieve self-actualization, 4: to do right moralistically, and 5: to gain pleasure) do you regard as motivation behind the behavior? Which way of thinking is closest to yours? On a scale of 1–5 with "1" being not important at all, and "5" being very important, please rate your consideration. Figure 5.3 shows the distribution for the importance indicator of five factors.

Over half of respondents answered that any of the items are important motivation behind their behaviors.

5.2.2.4 Information Security Awareness

Many previous studies make the appeal that it is important to improve information security awareness and knowledge. This survey incorporates 11 questions regarding information security awareness and the understanding of the measures used in the previous study [36]. These questions are measured on a five-point Likert scale. By using factor analysis to the questions, one factor is assumed. Cronbach's alpha of

[8]Cronbach's alpha of the scale was 0.691, which showed adequate internal consistency of the scale. The alpha is above the recommended level of 0.6.

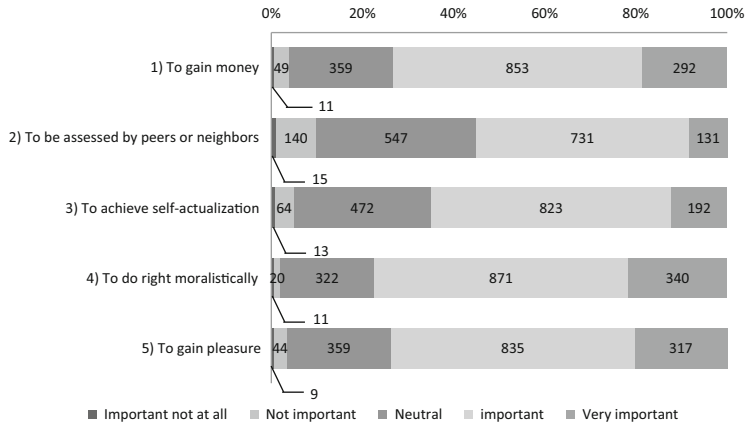

Fig. 5.3 The importance indicator of five factors

the scale was 0.734. The factor is scored so that a higher score indicates higher information security awareness.

5.2.2.5 Workplace Environment

The factors shown above have their roots in an individual's characteristics. On the other hand, workplace environment represents his or her environment. As mentioned above, workplace environment is divided into a subjective evaluation regarding the workplace and objective indicators such as organizational attributes.

The survey has some questions asking the degree of his or her workplace satisfaction and organizational information security measure satisfaction. Each question is scored in the range of 0–10 points. Figure 5.4 shows the distributions for the degrees of workplace satisfaction and information security measure satisfaction.

The average degree of workplace satisfaction is about 6.378 points and the average degree of organizational information security measure satisfaction is about 6.664 points.

According to Albrechtsen and Hovden [3], organizations have a digital divide between employees and information security managers, which arises from employees' dissatisfaction or criticism toward the managers and their measures. In this survey, we directly ask a question regarding the evaluation toward managers in addition to the evaluation of the security measures. Concretely, on a seven-point Likert scale, we ask participants to select the appropriate response to two statements, that the information security manager implements a measure with understanding of the job site and that the information security manager implements a measure which makes the employee's job harder. The factors are scored so that higher numbers indicate a higher evaluation toward the managers. Figure 5.5 shows the evaluations toward the managers.

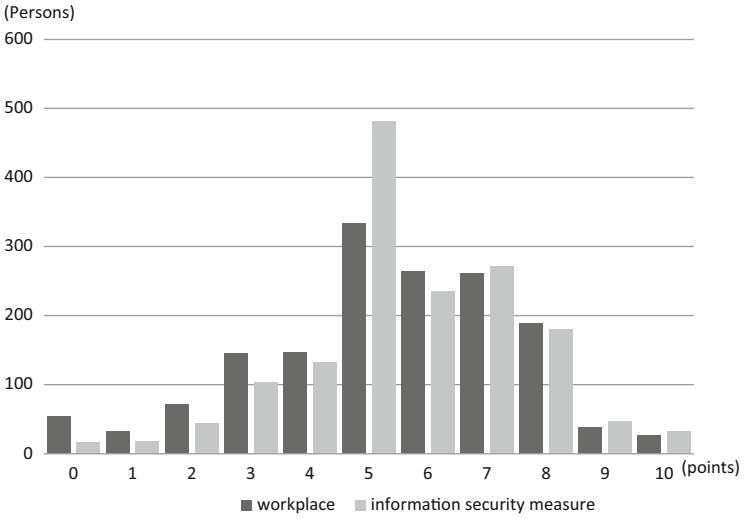

Fig. 5.4 The degrees of workplace satisfaction and information security measure satisfaction

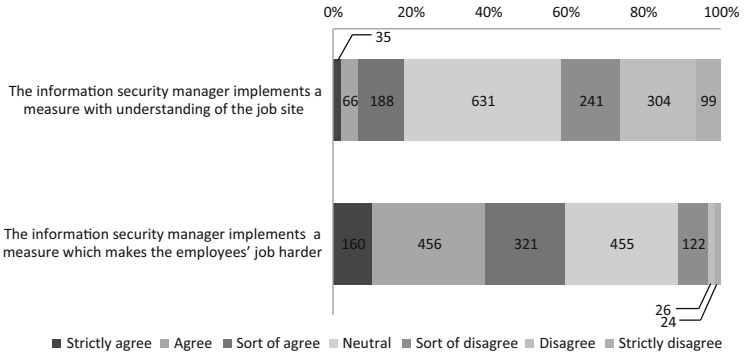

Fig. 5.5 The evaluations toward the managers

On the other hand, as objective indicators, we use the following organizational attributes; listed/non-listed stock, the number of employees, incentive systems for employees, and the employment system which is implemented in the organization. We pick up the same five incentive systems used in the previous study [37]. We use working pattern as the other objective indicator. Table 5.2 shows this demographic data regarding organizational attributes for the respondents of the survey.

5.2.2.6 The Other Individual Attributes

We use gender, age, education and annual income as individual attributes. In addition, in this survey, we ask questions regarding experience in encountering

Table 5.2 Demographic data regarding with the organizational attributes

Items		#
Listed/non-listed option	Listed company	768
	Non-listed company	796
# of employees	<100 (persons)	403
	100–999	439
	1,000–4,999	295
	5,000–9,999	125
	≥10,000	302
Incentive system	(1) Delegating the power from ruling body to lower organization	Yes: 290 / No: 1,274
	(2) A stock option system	Yes: 159 / No: 1,405
	(3) An employee stock ownership program	Yes: 582 / No: 982
	(4) Flexible schedules	Yes: 434 / No: 1,130
	(5) Work reassignment for the purpose of training	Yes: 203 / No: 1,361
Employment system	Permanent employment	Yes: 765 / No: 799
Working pattern	Regular	978
	Non-regular	586

Table 5.3 Demographic data regarding with the individual attributes

Items		#
Gender	Male	1,033
	Female	531
Age	Under 40	549
	40's	654
	50's	306
	Older than 60	55
Education	University degree or higher	968
	Other	596
Annual income	<2 (million JY)	293
	2–6	708
	6–10	400
	≥10	163
Experience in encountering Accidents	Experienced	526
	Not experienced	1,038

information security accidents, for example, "have you experienced virus infection in the past two years?" According to the result of this survey, about 10.7 % of respondents experienced some sort of information security accident. This demographic data regarding individual attributes for the respondents of the survey are shown in Table 5.3.

Table 5.4 Result of factor analysis (CFC scale)

	Factor loadings		
Questionnaire items	Myopic cognition	Hyperopic cognition	Uniqueness
Q1	0.0053	0.6945	0.5174
Q2	−0.0604	0.6935	0.5183
Q3	0.7209	−0.0510	0.4802
Q4	0.6383	0.0676	0.5850
Q5	0.5785	0.2250	0.6058
Q6	0.0171	0.5457	0.7013
Q7	−0.1996	0.6450	0.5530
Q8	−0.0428	0.6160	0.6206
Q9	0.7199	−0.0883	0.4782
Q10	0.7414	−0.1473	0.4361
Q11	0.8184	−0.0755	0.3288
Q12	0.4083	0.3440	0.7054

LR test: independent vs. saturated: chi2(66) = 4,657.53; Prob>chi2 = 0.0000

Table 5.5 Result of factor analysis (information security awareness)

	Factor loadings	
Questionnaire items	Information security awareness	Uniqueness
Q1	0.2910	0.9153
Q2	0.3691	0.8638
Q3	0.7950	0.3679
Q4	0.7707	0.4061
Q5	0.8168	0.3329
Q6	0.7970	0.3648
Q7	0.0853	0.9927
Q8	0.2928	0.9143
Q9	0.0662	0.9956
Q10	0.6253	0.6090
Q11	0.4848	0.7650

LR test: independent vs. saturated: chi2(55) = 4,679.80; Prob>chi2 = 0.0000

5.3 Results of Analysis

5.3.1 Factor Analysis

First of all, we run a factor analysis using the questionnaire items regarding CFC and information security awareness. Next, we calculate the (factor) score for the CFC scale and the information security awareness. Tables 5.4 and 5.5 display the results of the LR test of factor analysis, and the factor loadings and unique variances of the CFC scale and information security factors. Refer to [34] and [36] for the questionnaire items, respectively.

Table 5.6 Results of logit regression analysis

		Coef.	S.E.	z		Coef.	S.E.	z
1.	Myopic	−0.200	0.112	−1.790	Hyperopic	−0.256	0.112	−2.280
	Awareness	−0.330	0.126	−2.620	Satisfaction-WP	−0.079	0.053	−1.510
	Manager-2	−0.257	0.085	−3.030	Incentive-3	−0.466	0.253	−1.850
	Employment Sys.	0.570	0.274	2.080	Working pattern	0.955	0.340	2.810
	Gender	0.524	0.331	1.580	Exp. of accidents	1.122	0.238	4.720
	# of obs = 787, LR chi2(10) = 103.43, Log likelihood = −251.784, Pseudo R2 = 0.1704							
2.	Myopic	−0.361	0.155	−2.330	Hyperopic	−0.301	0.144	−2.090
	Awareness	−0.599	0.155	−3.870	Manager-2	−0.243	0.119	−2.040
	Listed	0.510	0.344	1.480	Incentive-3	−0.712	0.363	−1.960
	Employment Sys.	−0.554	0.353	−1.570	Working pattern	0.954	0.425	2.250
	Income	0.345	0.204	1.700	Exp. of accidents	0.916	0.307	2.980
	# of obs = 717, LR chi2(10) = 71.92, Log likelihood = −158.101, Pseudo R2 = 0.1853							
3.	Myopic	−0.283	0.142	−1.990	Hyperopic	−0.450	0.141	−3.200
	Awareness	−0.508	0.151	−3.370	Manager-1	−0.168	0.105	−1.590
	Incentive-4	−0.579	0.348	−1.660	Working pattern	0.697	0.342	2.030
	Exp. of accidents	0.754	0.295	2.550				
	# of obs = 974, LR chi2(7) = 65.61, Log likelihood = −181.491, Pseudo R2 = 0.1531							
4.	Risk Tolerance	0.065	0.044	1.460	Myopic	−0.326	0.122	−2.680
	Hyperopic	−0.220	0.126	−1.750	Motivation-5	0.264	0.171	1.540
	Manager-1	−0.296	0.092	−3.220	Incentive-2	−0.868	0.470	−1.850
	Employment Sys.	0.430	0.297	1.450	Working pattern	0.705	0.363	1.940
	Income	0.273	0.172	1.590	Exp. of accidents	0.863	0.261	3.300
	# of obs = 658, LR chi2(10) = 69.41, Log likelihood = −208.798, Pseudo R2 = 0.1425							
5.	Myopic	−0.533	0.147	−3.620	Hyperopic	−0.410	0.147	−2.790
	Motivation-2	−0.356	0.189	−1.890	Awareness	−0.290	0.151	−1.920
	Manager-2	−0.254	0.112	−2.270	Incentive-3	−0.639	0.319	−2.000
	Gender	1.145	0.407	2.810	Exp. of accidents	0.806	0.307	2.620
	# of obs = 908, LR chi2(8) = 72.25, Log likelihood = −168.637, Pseudo R2 = 0.1764							
6.	Hyperopic	−0.583	0.162	−3.590	Motivation-2	−0.596	0.286	−2.090
	Motivation-3	0.698	0.309	2.260	Awareness	−0.502	0.172	−2.920
	Incentive-1	0.697	0.434	1.610	Incentive-2	0.786	0.480	1.640
	Working pattern	1.407	0.479	2.940	Income	−0.477	0.275	−1.740
	Exp. of accidents	0.678	0.380	1.780				
	# of obs = 539, LR chi2(9) = 59.86, Log likelihood = −107.480, Pseudo R2 = 0.2178							

5.3.2 Logit Regression Analysis

We need to set a criterion (p-value) for removing insignificant variables in a stepwise logit model [19]. In this study, we set $p = 0.15$ as the criterion. We enter 28 explanatory variables, and eventually 8 variables, such as "Education" and "Age", are removed in the selection process. In this chapter, Stata/MP 12.0 is used as the statistical analysis software. Table 5.6 shows the estimated results.

First of all, the estimated coefficients of the hyperopic cognition (Hyperopic) and experience in information security accidents (Exp. of Accidents) factors are statistically significant in all cases. The former sign is positive and the latter sign is negative in all cases. That is, these factors commonly influence problematic behaviors. Next, in almost all cases, the estimated coefficients of the myopic cognition (Myopic), the information security awareness (Awareness) and working pattern (Working Pattern) are statistically significant. The signs of the myopic and the awareness's coefficients are negative and the sign of the rest is positive. In some cases, the estimated coefficients of at least one of the incentive systems (Incentive-1 to Incentive-4) are statistically significant. The coefficients of Incentive-1 and Incentive-2 in Case (6) are positive whereas the coefficients of the others are negative. In this analysis, the estimated coefficients of some motivations towards behavior (Motivation-2, Motivation-3 and Motivation-5) are statistically significant. The coefficient of Motivation-2 is negative whereas the coefficients of the rest are positive. In addition, with regard to some organizational or individual attributes, some estimated coefficients of the factors, such as evaluation toward the manager (Manager-1 or Manager-2), employment system (Employment Sys.) and annual income (Income) are statistically significant. Finally, with regard to risk attitude, the estimated coefficient of risk tolerance (Risk Tolerance) is statistically significant and the sign is positive only in Case 4.

From these estimated results, we can find out some features of the respondent's behavior in violating organizational rules, and compare with the previous study [37]. According to Takemura [37], the degrees of both risk aversion and risk tolerance have an effect on the behavior of violating organizational rules. However, in this analysis, the degree of risk tolerance has an effect on only the behavior of forwarding office e-mails to private addresses. The risk attitudes do not have an effect on the other behaviors. With regard to the behavior of forwarding e-mail, the more an individual can tolerate the risk, the more he tends to violate the rule. This is consistent with the assertion in the previous study.

With regard to CFC scale, both myopic cognition and hyperopic cognition have an effect on the behaviors of violating organizational rules in almost all cases. The more prevalent either of these cognitions is in the individual, the lesser the tendency to violate the rules. This means that the behavior of violating the rule is related to not only short-term cognition, but also long-term cognition. In addition, these cognitions have the same effect and are important factors to behavior.

The behavior of violating the rules is related to the motivation of assessment from peers, self-actualization, and pleasure, not the motivation of money or morals. Intriguingly, the greater the value of "assessment from peers" the individual places, the smaller the tendency to violate the rule is. On the contrary, an individual who places greater value on self-actualization (or pleasure) tends not to comply with the rule.

With regard to information security awareness, in many cases it is found that the higher the awareness is, the lesser the tendency to violate the rule is.

The behavior of violating the rules is independent of the degree of the information security measure satisfaction, and is not related to the degree of workplace

satisfaction and the evaluation toward the managers in some cases. In addition, the higher the evaluation toward the managers is, the lesser the tendency to violate the rule is.

The number of employees, which represents the scale of the organization, is not related to the behavior of violating the rules. Additionally, some incentive systems shown in Table 5.2 are related to the behavior in some cases. Also, an individual tends to violate the rules if the incentive system of delegating power (Incentive-1) is implemented. On the other hand, by implementing the other incentive systems (Incenive-2, Incentive-3 and Incentive-4), the individual tends not to violate the rules. This result is consistent with the result of the previous study.

Intriguingly, in an organization with permanent employment implemented, individuals tend to violate the rules. The individuals whose working patterns are regular also tend to violate the rules. The fact that these individuals tend to violate the rules is consistent with the result of the previous study. The message from this result might be that individuals violate the organizational rules for the purpose of completing their daily activities because they believe they will not get fired from their job by the employment system.

With regard to the other individual attributes, encountering information security accidents is related to the behavior similar to the previous study, but education is not related to the behavior.

5.4 Summary and Future Work

In this chapter, we determine some key factors which have effects on employees' behaviors in violating rules which are related to information leaks given that these behaviors are totally prohibited by their organizations. As a result, we found out some features of the respondent's behaviors in violating the rules.

First of all, the individual's attitude toward the risk or the cognition of risk (the psychological factors such as risk aversion and risk tolerance) are not related to the behavior of violating organizational rules in many cases. On the other hand, both myopic cognition and hyperopic cognition, measured by the CFC scale, have effects on the behaviors of violating organizational rules in almost all cases.

Next, the behavior of violating the rules is related to the motivations of assessment from peers, self-actualization and pleasure, not the motivations of money and morals.

Third, in many cases, individuals whose information security awareness is higher tend not to violate the rules.

Fourth, the behavior of violating the rules is independent of the number of employees which represents the scale of the organization, and is not related to the degree of workplace satisfaction and the evaluation toward the managers in some cases. Additionally, some of the incentive systems, shown in Table 5.2, are related to their behavior in some cases. Intriguingly, in an organization where permanent employment is implemented, individuals tend to violate the rules. Individuals whose working pattern is regular also tend to violate the rules.

With regard to the other individual attributes, encountering information security accidents is related to the behavior of violating the rules, but education is not related to the behavior.

It is not easy to control psychological factors such as the individual's attitude toward risk, motivations toward their behaviors or consideration of future consequences. Conversely, the factors regarded as organizational attributes such as the degree of workplace satisfaction or the employment system may be controlled by designing the appropriate organizational environment. Consequently, we consider that it may be effective to improve information security awareness by information security education and training which is suggested in some of the previous literature [4, 26]. Actually, as mentioned above, individuals whose information security awareness is higher tend not to violate the rules.

Finally, let us briefly explain the limitation of our work and future work.

The evaluation of "attitude" is somehow difficult to control because the attitude of each person can be changed according to his/her perception of issues regarding risk. Therefore, there is a limitation to investigating the true attitude using Web-based surveys. Of course, through time series comparisons of the relations between attitude and the other factors, we check can robustness of our results.

Although the empirical studies on information security measures have meaningful messages in social science and are essential in business practice, the number of empirical studies is still small. So, there are many yet-to-be-defined information security behaviors and mechanisms. Therefore, individuals' information security behaviors should be deeply analyzed from the perspectives of economics and behavioral science. We will tackle these issues in future work. Though in this chapter we build a behavioral model using the logit model, we will build models based on TPB, GDT or the theory of fraud triangles by using statistical tools such as SEM or PLS in future work.

Furthermore, we expect this chapter will become an academic contribution to this field, and will give an incentive for companies to invest in and implement information security measures.

Acknowledgements This work is supported in part by a Grant-in-Aid from the Zengin Foundation for Studies on Economics and Finance and by the Japan Society for the Promotion of Science: Grant-in-Aid for Young Scientists (B) (22730241).

This chapter incorporates some valuable comments by anonymous reviewers for WEIS 2012. All remaining errors are our own.

References

1. Ajzen, I.: The theory of planned behavior. Organ. Behav. Hum. Decis. Process. **50**, 179–211 (1991)
2. Albrechtsen, E.: A qualitative study of users' view on information security. Comput. Secur. **26**, 276–289 (2007)

3. Albrechtsen, E., Hovden, J.: The information security digital divide between information security managers and users. Comput. Secur. **28**, 476–490 (2009)
4. Albrechtsen, E., Hovden, J.: Improving information security awareness and behaviour through dialogue, participation and collective reflection. An intervention study. Comput. Secur. **29**, 432–445 (2010)
5. Aurigemma, S., Panko, R.: A composite framework for behavioral compliance with information security policies. In: Proceedings of 45th Hawaii International Conference on System Sciences, Maui, pp. 3248–3257 (2012)
6. Becker, G.M., Degroot, M.H., Marschak, J.: Measuring utility by a single response sequential method. Behav. Sci. **9**, 226–232 (1964)
7. Bulgurcu, B., Cavusoglu, H., Benbasat, I.: Roles of information security awareness and percieved fairness in information security policy compliance. In: Proceedings of the Fifteenth Americas Conference on Information Systems, San Francisco, Paper 419, 1-9 (2009)
8. Cappelli, D.M., Desai, A.G., Moore, A.P., Shimeall, T.J., Weaver, E.A., Willke, B.J.: Management and education of the risk of insider threat (MERIT): mitigating the risk of sabotage to employers' information, systems, or networks. Technical Note, Software Engineering Institute, Carnegie Mellon University, CMU/SEI-2006-TN-041 (2007)
9. Chang, K.M.: Predicting unethical behavior: a comparison of the theory of reasoned action and the theory of planned behavior. J. Bus. Ethics **17**(16), 1825–1834 (1998)
10. Chen, J.V., Chen, C.C., Yang, H.H.: An empirical evaluation of key factors contributing to internet abuse in the workplace. Ind. Manage. Data Syst. **108**(1), 87–106 (2008)
11. Couper, M.P.: Web surveys: a review of issues and approaches. Piblic Opin. Q. **64**, 464–494 (2000)
12. Cramer, J.S., Hatog, J., Jonker, N., Van Praag, C.M.: Low risk aversion encourages the choice for entrepreneurship: an empirical test of a truism. J. Econ. Behav. Organ. **48**, 29–36 (2002)
13. Cressey, D.R.: Other People's Money: A Study in the Social Psychology of Embezzlement. Patterson-Smith, Montclair (1973)
14. D'Arcy, J., Hovav, A., Galletta, D.: User awareness of security countermeasures and its impact on information system misuse: a deterrence approach. Inf. Syst. Res. **20**, 1–20 (2008)
15. Foltz, C.B., Schwager, P.H., Anderson, J.E.: Why users (fail to) read computer usage policies. Ind. Manage. Data Syst. **108**(6), 701–712 (2008)
16. Galletta, D.F., Polak, P.: An empirical investigation of antecedents of internet abuse in the workplace. In: Proceedings of the Second Annual Workshop on HCI Research in MIS, Seattle, pp. 47–51 (2003)
17. Greitzer, F.L., Hohimer, R.E.: Modeling human behavior to anticipate insider attacks. J. Strateg. Secur. **IV**(2), 25–48 (2011)
18. Greitzer, F.L., Kangas, L.J., Noonan, C.F., Dalton, A.C., Hohimer, R.E.: Identifying at-risk employees: modeling psychosocial precursors of potential insider threats. In: Proceedings of 45th Hawaii International Conference on System Sciences, Maui, pp. 2392–2401 (2012)
19. Hosmer, D.W., Lemeshow, S.: Applied Logistic Regression, 2nd edn. Wiley-Interscience, New York (2000)
20. Information-Technology Promotion Agency: Information Security White Paper. Information-Technology Promotion Agency, Tokyo (2011)
21. Kahneman, D., Tversky, A.: Prospect theory: an analysis of decision under risk. Econometrica **47**(2), 263–292 (1979)
22. Komatsu, A., Akai, K., Ueda, M., Matsumoto, T.: Is the security measurement situation a social dilemma? Applying Bot measurement operation. IPSK SIG Technical Report, 2009-CSEC-46(40), 1-8 (2009)
23. Lee, J., Lee, Y.: A holistic model of computer abuse within organizations. Inf. Manage. Comput. Secur. **10**(2), 57–63 (2002)
24. Lee, S.M., Lee, S.G., Yoo, S.: An integrative model of computer abuse based on social control and general deterrence theories. Inf. Manage. **40**, 707–718 (2004)
25. Lim, V.K.G.: The IT way of loafing on the job: cyberloafing, neutralizing and organizational justice. J. Organ. Behav. **23**, 675–694 (2002)

26. Mcilwraith, A.: Information Security and Employee Behaviour: How to Reduce Risk Through Employee Education, Training and Awareness. Gower, Hampshire (2006)
27. Peace, A.G., Galletta, D.F., Thong, J.Y.L.: Software piracy in the workplace: a model and empirical test. J. Manage. Inf. Syst. **20**(1), 153–177 (2003)
28. Pee, L.G., Woon, I.M.Y., Kankanhalli, A.: Explaining non-work-related computing in the workplace: a comparison of alternative models. Inf. Manage. **45**, 120–130 (2008)
29. Pfleeger, S.L., Stolfo, S.J.: Addressing the insider threat. IEEE Comput. Reliab. Soc. **7**(6), 10–13 (2009)
30. Pfleeger, S.L., Predd, J.B., Hunker, J., Bulford, C.: Insiders behaving badly: addressing bad actors and their actions. IEEE Trans. Inf. Forensics Secur. **5**(1), 169–179 (2010)
31. Predd, J.B., Pfleeger, S.L., Hunker, J., Bulford, C.: Insiders behaving badly. IEEE Secur. Priv. **6**, 66–70 (2008)
32. Reason, J., Parker, D., Lawton R.: Organizational controls and safety: the varieties of rule-related behaviour. J. Occup. Organ. Psychol. **71**, 289–304 (1998)
33. Riketta, M.: Attitudinal organizational commitment and job performance: a meta-analysis. J. Organ. Behav. **23**, 257–266 (2002)
34. Strathman, A., Gleicher, F., Boninger, D.S., Edwards, C.S.: The consideration of future consequences: weighing immediate and distant outcomes of behavior. J. Personal. Soc. Psychol. **66**(4), 742–752 (1994)
35. Straub, D.: Effective IS security: an empirical study. Inf. Syst. Res. **1**(3), 255–276 (1990)
36. Takemura, T.: A quantitative study on Japanese workers' awareness to information security using the data collected by web-based survey. Am. J. Econ. Bus. Adm. **2**(1), 20–26 (2010)
37. Takemura, T.: Empirical analysis of behavior on information security. In: The Proceeding of 2011 IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing, Dalian, pp. 358–363 (2011)
38. Theoharidou, M., Kokolakis, S., Karyda, M., Kiountouzis, E.: The insider threat to information systems and the effectiveness of ISO17799. Comput. Secur. **24**, 472–484 (2005)
39. Thomson, M.E., Solms R.: Information security awreness: educating your users effectively. Inf. Manage. Comput. Secur. **6**(4), 167–173 (1998)
40. Tsukahara, Y.: Human motivating behavior and employee's behavior. In: Chida, R., Tsukahara, Y., Yamamoto, M. (eds.) Behavioral Eonomics: Theory and Practice, pp. 50–71. Keiso-shobo, Tokyo (2010)
41. Wells, J.T.: Principles of Fraud Examination, 3rd edn. Wiley, Hoboken (2010)
42. Woon, I.M.Y., Pee, L.G.: Behavioral factors affecting internet abuse in the workplace: an empirical investigation. In: Proceedings of the Third Annual Workshop on HCI Research in MIS, Washington, DC, pp. 80–84 (2004)
43. Workman, M., Gathegi, J.: Punishment and ethics deterrents: a study of insider security contravention. J. Am. Soc. Inf. Sci. Technol. **58**(2), 212–222 (2007)
44. Zhang, J., Reithel, B.J., Li, H.: Impact of perceived technical protection on security behaviors. Inf. Manage. Comput. Secur. **17**(4), 330–340 (2009)

Chapter 6
Sectoral and Regional Interdependency of Japanese Firms Under the Influence of Information Security Risks

Bongkot Jenjarrussakul, Hideyuki Tanaka, and Kanta Matsuura

Abstract Although there are some studies on inter-sectoral information security interdependency, the lack of regional interdependency analysis is one of their limitations. In this empirical study, we used an inter-regional input–output table in order to analyze both sectoral and regional interdependencies under the influence of information technology and the information security of Japanese firms. Our analysis showed that the economic scale of a region has a great influence on the characteristics of the interdependency. Furthermore, we found that the demand-side sectors can be classified into five classes based on the characteristics. Among them, the groups with high self-dependency get more benefits from simultaneous understanding of regional characteristics; for the sectors in these classes, investment advice obtained from sectoral characteristics only is very limited, whereas they can obtain much more from regional characteristics. Since these classes include a majority of the sectors, we can recognize the importance of regional interdependency analysis. In the above basic study, what we see is the situation before the Great East Japan Earthquake on March 11, 2011.

As an extended study, we estimated the impact of the earthquake on the interdependency. Our main finding from the regional perspective is that the interdependency characteristics of the most damaged region (Tohoku) and of the economically largest region (Kanto) are impacted most significantly. This feature is not changed by the limitation of damage through prior security investment.

Both in the basic study and in the extended study, we can see that considering not only sectoral but also regional characteristics is an effective approach to the task

B. Jenjarrussakul (✉) · K. Matsuura
Institute of Industrial Science, The University of Tokyo, 4-6-1 Komaba, Meguro-ku, Tokyo, Japan
e-mail: bongkot@iis.u-tokyo.ac.jp; kanta@iis.u-tokyo.ac.jp

H. Tanaka
Graduate School of Interdisciplinary Information Studies, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
e-mail: tanaka@iii.u-tokyo.ac.jp

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_6,

of empirically deriving implications related to the interdependency. There are many possibilities of more extended studies based on our methodology.

6.1 Introduction

6.1.1 Interdependency

Interdependency of information security is one of the main concerns in security economics. Empirical studies on interdependency of information security require two main groups of knowledge: one is from the viewpoint of economic transactions, and the other is from the viewpoint of security efforts or investments.

Regarding the interdependency of economic activities, information technology (IT) becomes one of the role players in supply chains [24]. Many firms use IT systems to interact with supply chain participants via applications such as supply chain management (SCM) and electronic data interchange (EDI) systems. Thus, IT brings interdependency into many industrial sectors such as *automotive* [17], *computer* [7], *financial services* [4, 10, 15], and *retail and logistics* [2, 9, 18].

In the area of security economics, interdependency is very important particularly in the context of externalities where the security of firms depends not only on their own efforts but also the efforts by other firms [1]. Kunreuther and Heal applied Nash equilibria to assess the interdependent security [19]. The impacts of network security vulnerabilities and supply chain integration on firms' incentives to their investments in information security were studied by Bandyopadhyay et al. [3]. They showed that the degree of network vulnerability or the degree of supply chain integration has relations to security investments. Hausken provided a framework in which two interdependent firms will be impacted both by security investment and by attacks if their interdependency increases [16]. Ogut et al. showed that the interdependency reduces firms' incentives towards investments in security technologies as well as towards insurance coverage [23]. Tanaka studied economic interdependency between sectors under the influence of IT systems [28]; he assumed that a malfunctioning IT system in a firm will affect not only the economic activities of the firm but also those of its business partners. He then introduced the concept of ISBL (information security backward linkage) and analyzed interdependencies between firms in different sectors. Although he empirically assessed the influence of business locations on information security efforts [27], he did not analyze regional interdependencies in his ISBL study.

6.1.2 Our Contributions

In this chapter, we analyze the interdependency of information security from both sectoral and regional perspectives by using Japanese official datasets. Showing

how regional perspective is helpful in systematic analyses of interdependency is our main contribution. In other words, this chapter broadens the concept of the measurement methodology of interdependency by considering both sectoral and regional interdependencies of information security.

After the above analysis, the Great East Japan Earthquake occurred on March 11, 2011. This unfortunately reduced the empirical significance of each particular interdependency characteristic observed. However, rather than being disappointed in the empirical analysis, we proceeded to extended analyses on the impact of the earthquake. Thus we suggest a wide variety of possibilities regarding extended studies based on the proposed methodology. This suggestion and some empirical findings in the earthquake analysis is our second contribution.

In the rest of this chapter, we will first summarize more related works in Sect. 6.2. Our basic analysis methodologies will be described in Sect. 6.3. Then our datasets will be introduced in Sect. 6.4, and an extended study on the earthquake impact will be explained in Sect. 6.5. After showing and discussing the results in Sect. 6.6, we will give conclusions in Sect. 6.7.

6.2 Related Literature

Let us start from the Inoperability Input–output Model (IIM). IIM is a Leontief-based infrastructure input–output model introduced by Haimes and Jiang [12] in 2001. In particular, IIM can be used to quantify and address the risks from the intra- and inter-connectedness of infrastructures [29].

Inoperability in IIM is defined as "the inability of the system to perform its intended natural or engineered functions" [14]. It can be referred to as the level of the system's dysfunction. The main objective of their model is to assess the impact of interdependencies between infrastructures on the system. The use of IIM in [14] focused on the *industry-by-industry* viewpoint, and interdependencies between locations were not considered. Haimes et al. also introduced "Dynamic IIM" in order to test interdependency with temporal dynamic behaviors of industry recoveries after damage. In another work [13], they used high-altitude electromagnetic pulse attack scenarios to evaluate their model.

The IIM framework can be used to integrate analyses of systems from a hierarchical viewpoint where economic interdependency and physical interdependency are considered [29]. Here a hierarchical pyramid is used to show how economic and physical systems interact. Likewise, IIM can be used in the analysis of interdependencies under the influence of information security where several interactions may be considered. In fact, the framework for linking hierarchies of cybersecurity metrics is used to show consequent risks in the case of a cyber attack in an industrial sector such as Oil and Gas [22].

There are two main limitations of the existing works based on IIM. First, IIM does not distinguish between the demand-driven perspective and the supply-driven

perspective. Another limitation is the lack of data regarding the level of IT dependency and information-security measures.

6.3 Methodology

Let us use a two-step approach to introduce our analysis methodology regarding Information Security Backward Dependency (ISBD). The first step is about the analysis of cross-sectoral/regional interdependency as a basic economic analysis; we can conduct a sensitivity analysis by supposing complete damage in a particular part of the input–output table. The second step is about the analysis of the interdependency under the influences of IT and information security (IS); we can conduct a similar but different analysis by supposing that the damage depends on the level of IT dependency and the level of IS efforts. This view as a sensitivity analysis helps an intuitive understanding of our methodology.

6.3.1 Structural Interdependency

From an economic viewpoint, structural interdependency can be assessed from two perspectives: demand-driven perspective and supply-driven perspective. In the case of demand-driven perspective, the assessment is done from the purchaser's viewpoint. On the other hand, in the case of supply-driven perspective, the assessment is done from the producer's viewpoint.

The assessment methodology from demand-driven and supply-driven perspectives was initially proposed by Dietzenbacher and van der Linder in 1997 [8]. Their method was used to measure the inter-industry linkages in a multi-sectoral framework. They analyzed the value of absolute *Backward Linkage* (BL) which reflects sectors' dependency on its *inputs* that they produced within the production processes. Another analyzed value is the absolute *Forward Linkage* (FL), which, by contrast, reflects a sector's dependency on its *outputs* that were sold by a particular industry to other production sectors as well as to itself.

In our work, we aim to find interdependency from the demand-driven perspective. Hence, we focus on BL. As another important feature of our work, we extend the basic definitions in the Dietzenbacher and van der Linder work so that we can handle both sectoral and regional interdependencies.

6.3.1.1 Observed Values

In [8], the input–output table is used to show relationships between industrial sectors. We extend their definitions by considering additional indices to indicate different regions. In other words, we consider an *inter-regional input–output table*

$Z = \left(z_{q,i,r,j}\right)$ where each intersection $z_{q,i,r,j}$ is the economic transaction of goods and services purchased by demand-side companies of sector j in region r from supply-side companies of sector i in region q. Each transaction is valued at producers' prices. The combination of a region and a sector is called a *group*. When we talk about firms in a particular sector in a particular region on the demand side, we call the corresponding group a *demand-side group*. Likewise, we define a *supply-side group*. In terms of the matrix structure, the four indices are used as follows:

$$
Z = \begin{pmatrix}
z_{1,1,1,1} & z_{1,1,1,2} & \cdots & z_{1,1,1,n} & z_{1,1,2,1} & z_{1,1,2,2} & \cdots & z_{1,1,2,n} & \cdots & z_{1,1,d,1} & z_{1,1,d,2} & \cdots & z_{1,1,d,n} \\
z_{1,2,1,1} & z_{1,2,1,2} & \cdots & z_{1,2,1,n} & z_{1,2,2,1} & z_{1,2,2,2} & \cdots & z_{1,2,2,n} & \cdots & z_{1,2,d,1} & z_{1,2,d,2} & \cdots & z_{1,2,d,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
z_{1,n,1,1} & z_{1,n,1,2} & \cdots & z_{1,n,1,n} & z_{1,n,2,1} & z_{1,n,2,2} & \cdots & z_{1,n,2,n} & \cdots & z_{1,n,d,1} & z_{1,n,d,2} & \cdots & z_{1,n,d,n} \\
z_{2,1,1,1} & z_{2,1,1,2} & \cdots & z_{2,1,1,n} & z_{2,1,2,1} & z_{2,1,2,2} & \cdots & z_{2,1,2,n} & \cdots & z_{2,1,d,1} & z_{2,1,d,2} & \cdots & z_{2,1,d,n} \\
z_{2,2,1,1} & z_{2,2,1,2} & \cdots & z_{2,2,1,n} & z_{2,2,2,1} & z_{2,2,2,2} & \cdots & z_{2,2,2,n} & \cdots & z_{2,2,d,1} & z_{2,2,d,2} & \cdots & z_{2,2,d,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
z_{2,n,1,1} & z_{2,n,1,2} & \cdots & z_{2,n,1,n} & z_{2,n,2,1} & z_{2,n,2,2} & \cdots & z_{2,n,2,n} & \cdots & z_{2,n,d,1} & z_{2,n,d,2} & \cdots & z_{2,n,d,n} \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
z_{d,1,1,1} & z_{d,1,1,2} & \cdots & z_{d,1,1,n} & z_{d,1,2,1} & z_{d,1,2,2} & \cdots & z_{d,1,2,n} & \cdots & z_{d,1,d,1} & z_{d,1,d,2} & \cdots & z_{d,1,d,n} \\
z_{d,2,1,1} & z_{d,2,1,2} & \cdots & z_{d,2,1,n} & z_{d,2,2,1} & z_{d,2,2,2} & \cdots & z_{d,2,2,n} & \cdots & z_{d,2,d,1} & z_{d,2,d,2} & \cdots & z_{d,2,d,n} \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots & \vdots & \ddots & \vdots \\
z_{d,n,1,1} & z_{d,n,1,2} & \cdots & z_{d,n,1,n} & z_{d,n,2,1} & z_{d,n,2,2} & \cdots & z_{d,n,2,n} & \cdots & z_{d,n,d,1} & z_{d,n,d,2} & \cdots & z_{d,n,d,n}
\end{pmatrix}
$$

where we denote the number of regions by d and the number of sectors by n.

In addition to Z, the following values are directly observed from [20]:

Final demand: Final demand is denoted by matrix $F = \left(f_{q,i,r}\right)$. From F, we obtain the following two vectors:

Regional final demand: $f^* = \left(f^*_{q,i}\right)$ where $f^*_{q,i} = f_{q,i,q}$.

Accumulated final demand: $\hat{f} = \left(\hat{f}_{q,i}\right)$ where $\hat{f}_{q,i} = \sum_{r=1}^{d} f_{q,i,r}$.

Import: Import is denoted by vector $m = \left(m_{r,j}\right)$ where each element represents the absolute value of the import by each demand-side group. Normalization of the import vector m by the regional final demand gives the *import coefficient* matrix $B = \left(b_{q,i,r,j}\right)$ where

$$
b_{q,i,r,j} = \begin{cases} m_{q,i}/f^*_{q,i} & \text{if } r = q \text{ and } j = i \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}
$$

Export: Export is denoted by vector $e = \left(e_{q,i}\right)$ where each element represents the value of the export by each supply-side group.

Value added: Value added is denoted by vector $c = \left(c_{r,j}\right)$ where each element represents the value or tax added to the purchase by each demand-side group. From Z and c, we compute the gross output vector $g = \left(g_{r,j}\right)$ where

$$g_{r,j} = \sum_{q=1}^{d} \sum_{i=1}^{n} z_{q,i,r,j} + c_{r,j} \tag{6.2}$$

represents the gross output to each demand-side group. Normalization of Z by the gross output gives the *input coefficient* which is denoted by matrix $A = \left(a_{q,i,r,j}\right)$ where

$$a_{q,i,r,j} = z_{q,i,r,j} / g_{r,j}. \tag{6.3}$$

In order to extract the input coefficients inside each region, we define a matrix $A^* = \left(a^*_{q,i,r,j}\right)$ by

$$a^*_{q,i,r,j} = \begin{cases} a_{q,i,r,j} & \text{if } q = r \\ 0 & \text{otherwise.} \end{cases} \tag{6.4}$$

6.3.1.2 Backward Dependency

If all the deliveries to a demand-side group $(\overline{r}, \overline{j})$ are reduced to be zero by a disastrous event, the output from the group will be reduced. We compute such output reductions in order to study absolute backward linkages. The output reductions are given by $h - \overline{h}\left(\overline{r}, \overline{j}\right)$ where

$$h = \left\{I - \left[A - BA^*\right]\right\}^{-1} \left(\hat{f} - Bf^* + e\right), \tag{6.5}$$

$$\overline{h}\left(\overline{r}, \overline{j}\right) = \left\{I - \left[\overline{A}\left(\overline{r}, \overline{j}\right) - B\overline{A}^*\left(\overline{r}, \overline{j}\right)\right]\right\}^{-1} \left(\hat{f} - Bf^* + e\right), \tag{6.6}$$

and I is the identity matrix of the corresponding size. The matrices $\overline{A}\left(\overline{r}, \overline{j}\right) = \left(\overline{a}\left(\overline{r}, \overline{j}\right)_{q,i,r,j}\right)$ and $\overline{A}^*\left(\overline{r}, \overline{j}\right) = \left(\overline{a}^*\left(\overline{r}, \overline{j}\right)_{q,i,r,j}\right)$ are calculated from A and A^* as follows:

$$\overline{a}(\overline{r}, \overline{j})_{q,i,r,j} = \begin{cases} 0 & \text{if } r = \overline{r} \text{ and } j = \overline{j} \\ a_{q,i,r,j} & \text{otherwise} \end{cases} \tag{6.7}$$

and

$$\overline{a}^*(\overline{r}, \overline{j})_{q,i,r,j} = \begin{cases} 0 & \text{if } r = \overline{r} \text{ and } j = \overline{j} \\ a^*_{q,i,r,j} & \text{otherwise.} \end{cases} \tag{6.8}$$

Let vector $u(\overline{r}, \overline{j}) = \left(u(\overline{r}, \overline{j})_{q,i}\right)$ denote the backward dependency (BD) of a demand-side group $(\overline{r}, \overline{j})$ on the supply-side groups. We can obtain $u(\overline{r}, \overline{j})$ in terms of percentage by

$$u(\overline{r}, \overline{j})_{q,i} = 100 \frac{h_{q,i} - \overline{h}(\overline{r}, \overline{j})_{q,i}}{g_{\overline{r},\overline{j}}}. \tag{6.9}$$

6.3.2 Interdependency Under the Influence of Information Security

In [28], the ISBD vector of a demand-side group $(\overline{r}, \overline{j})$ is defined as the BD vector computed by replacing (6.7) and (6.8) with

$$\overline{a}(\overline{r}, \overline{j})_{q,i,r,j} = \begin{cases} (1 - s_i s_j) a_{q,i,r,j} & \text{if } r = \overline{r} \text{ and } j = \overline{j} \\ a_{q,i,r,j} & \text{otherwise} \end{cases} \tag{6.10}$$

and

$$\overline{a}^*(\overline{r}, \overline{j})_{q,i,r,j} = \begin{cases} (1 - s_i s_j) a^*_{q,i,r,j} & \text{if } r = \overline{r} \text{ and } j = \overline{j} \\ a^*_{q,i,r,j} & \text{otherwise} \end{cases} \tag{6.11}$$

where s_i represents the security risk level of sector i. The values of security risk levels are obtained from additional datasets [21, 25].

6.4 Data for Sectoral and Regional Interdependency

6.4.1 Inter-regional Input–Output Table for 2005

In this chapter, we mainly use the dataset of 12 sectors. The dataset of 53 sectors is used for further analyses on some sectors.

In this dataset [20], Japan is divided into nine regions: Hokkaido, Tohoku, Kanto, Chubu, Kinki, Chugoku, Shikoku, Kyushu, and Okinawa. These regions are indexed by A, B, C, …, and I, respectively. Regarding the economic scale, Kanto (C), Kinki (E), and Chubu (D) are the top three regions with high production values. On the other hand, Okinawa (I), Shikoku (G), and Hokkaido (A) are the bottom three regions with low production values.

From the sectoral perspective, the top three sectors with high production values are Services (12), Commerce & logistics (09), and Manufacturing Machinery (05),

Table 6.1 Regional production values in Japan

Region name	Region ID	Output (billion US$[a])
Kanto	C	8,175.19
Kinki	E	3,042.11
Chubu	D	2,341.25
Kyushu	H	1,576.64
Chugoku	F	1,176.51
Tohoku	B	1,136.39
Hokkaido	A	684.96
Shikoku	G	508.69
Okinawa	I	116.78

Source: Inter-regional input–output table for 2005

[a] 1 US($) = 76.75 JYP(¥). Rate on Oct 19, 2011

Table 6.2 Japanese sectoral production values for 12 industrial sectors

Sector name	Sector ID	Output (billion US$[a])
Services	12	3,090.26
Commerce & Logistics	09	1,916.02
Manufacturing-Machinery	05	1,696.06
Financial, Insurance, and Real Estate	10	1,404.47
Manufacturing-Other	06	1,229.48
Construction	07	823.94
ICT	11	598.51
Manufacturing-Metal	04	593.76
Manufacturing-Food & Beverage	03	468.23
Utilities	08	349.05
Agriculture	01	171.40
Mining	02	13.14

Source: Inter-regional input–output table for 2005

[a] 1 US($) = 76.75 JYP(¥). Rate on Oct 19, 2011

whereas Mining (02), Agriculture (01), and Utilities (08) are the bottom three sectors with low production values.

Tables 6.1 and 6.2 show Japanese production values from regional and sectoral (12 sectors) perspectives, respectively.

6.4.2 The 2006 Survey of Information Technology

The Survey of Information Technology is a popular periodical in Japan. Its 2006 version contains reliable data of 3,647 firms from 27 industries [21]. We use the average number of information security (IS) measures deployed by the firms in each

Table 6.3 List of information security measures

Category	Information security measures
Implementation of organizational measures	- Risk analysis - Security policy - Examination of specific measures based on security policy - Creation of information security report - Creation of Business Continuity Plan (BCP) - Deployment of an corporate-wide security management - Sectoral deployment of security management - Information security training for employees - Confirmation on information security measures of trading partners (including outsourcing)
Implementation of technical solutions/defense measures	- Access control of important computer rooms - Access control of important systems - Data encryption (including Public Key Infrastructure (PKI)) - Firewall installation against external connection - Installation of ISO/IEC15408 certified product
System monitoring	- Installation of security monitoring software - Full-time monitoring by external professionals
Assessment	- Use of information security benchmark - Regular system auditing by external professionals - Regular system auditing by internal experts - Regular information security auditing by external professionals - Regular information security auditing by internal experts - Obtaining certification of information security management system (ISO/IEC27001)

sector as a proxy of the level of IS in each sector. The IS measures are classified into four categories shown in Table 6.3.

We compute *IS multiplier* (denoted by m_i) which represents the normalized level of IS measures. This variable is defined by

$$m_i = \mathrm{M}^*/\mathrm{M}_i \tag{6.12}$$

where M^* is the average number of deployed IS measures across all the sectors and M_i is the average number of deployed IS measures in sector i.

Although there are some similar surveys in other countries (e.g., 2005 CSI/FBI Computer Crime and Security Survey [11]), our dataset is more reliable and usable for empirical studies. First, let us recall the sample size and the coverage of industries of our dataset (3,647 firms from 27 industries). In contrast, [11] has approximately 700 samples, and its coverage of industries is questionable. Second, our dataset is more usable since we can see more detailed statistics regarding the deployment of IS measures. In particular, we can obtain not only the average number

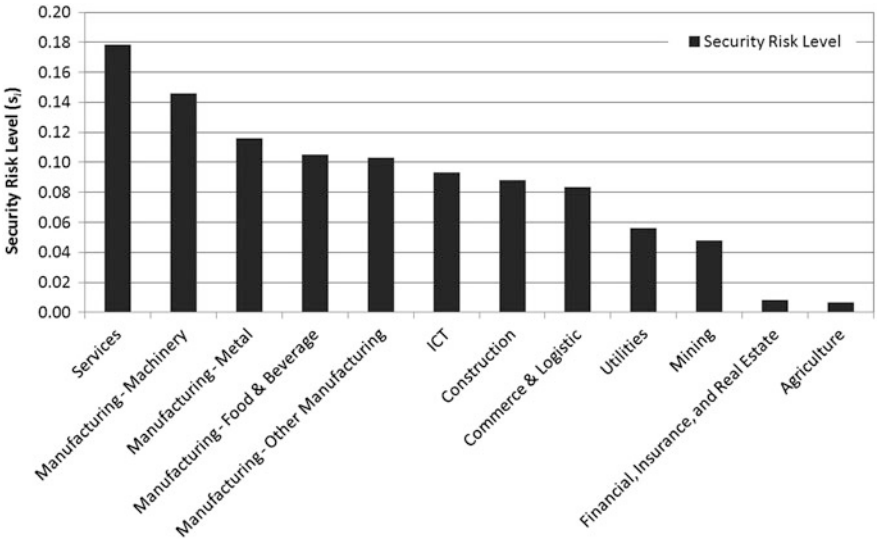

Fig. 6.1 Security risk levels of 12 industrial sectors

of deployed IS measures across all the sectors but also the average number of deployed IS measures in each sector.

6.4.3 Japan Industrial Productivity Database 2008

We use the data of *IT Capital Stock* and *non-IT Capital Stock* reported in [25] in order to estimate *the level of IT dependency* of each sector. Let t_i denote the level of IT dependency of sector i. We estimate the level of IT dependency by

$$t_i = \mathrm{IT}_i / (\mathrm{IT}_i + \mathrm{nIT}_i) \tag{6.13}$$

where IT_i denotes the IT capital stock of sector i and nIT_i denotes the non-IT capital stock of sector i. We then use

$$s_i = t_i m_i \tag{6.14}$$

as a proxy for the security risk level of sector i.

Figure 6.1 shows the security risk levels of 12 sectors. The levels of IT dependency, the levels of IS measure, and the IS multipliers computed from our dataset are shown in Figs. 6.2 and 6.3.

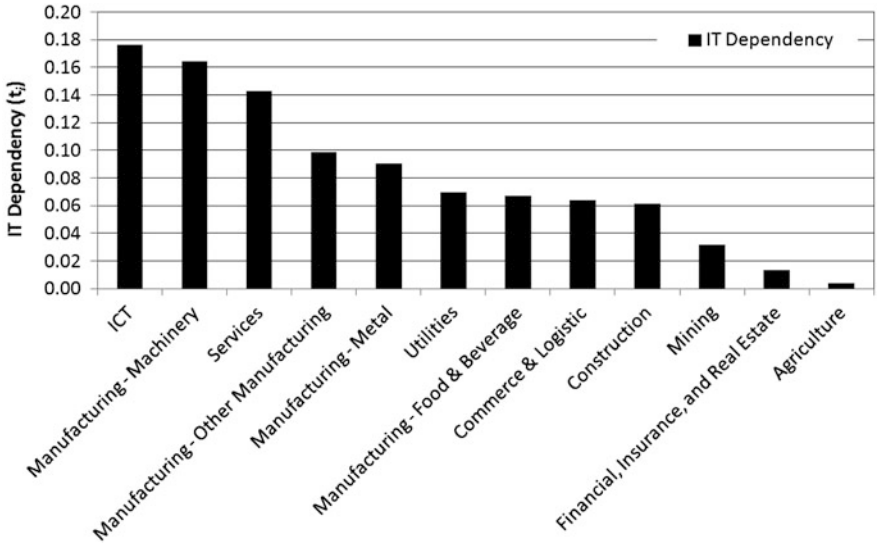

Fig. 6.2 Levels of IT dependency of 12 industrial sectors

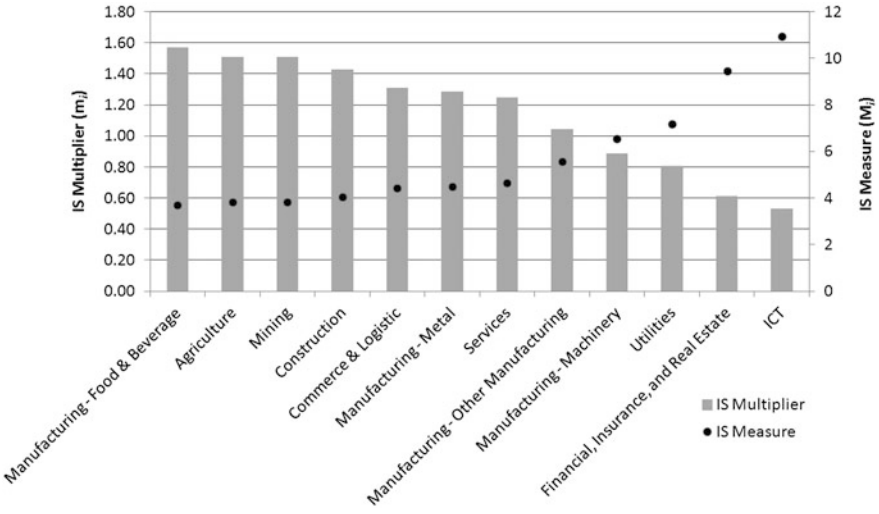

Fig. 6.3 Levels of IS measures and IS multipliers of 12 industrial sectors

6.5 Extended Analysis on the Impact of the Earthquake

At 14:46 p.m. on March 11, 2011, the Great East Japan Earthquake hit the Tohoku region with magnitude 9.0. This massive earthquake also triggered tremendous and powerful tsunami waves which left dreadful damage. The Cabinet Office of

the Government of Japan defined seven prefectures as disaster areas regarding this earthquake [6]. Among them, the three most significantly damaged prefectures are in the Tohoku region. The Cabinet Office defined the following two types of damage.

Case 1: refers to the damage directly by the earthquake, and
Case 2: refers to the damage by the earthquake and the subsequent tsunami.

Shinozaki et al. estimated the impact on ICT-related private capital stock due to the Great East Japan Earthquake [26]. Their result shows that the damage is around 2.5–4.4 trillion yen in total.

We study the impact of the Great East Japan Earthquake by using the methodology in Sect. 6.3 with a modification based on the following two additional datasets:

1. **Special cabinet meeting material on the monthly economic report due to the earthquake** [6]
   This report was provided a few weeks after the earthquake by the government. We obtain the overall damage on capital stock, D^{all}, from this dataset.
2. **Gross Capital Stock by Industry** [5]
   We use the values of the gross capital stock of year 2009, which was the newest at the time we estimated the impact of the earthquake. We obtain the nationwide capital stock, C_n, from this dataset.

Now let us describe the extended analysis. First, in the analysis regarding the structural interdependency, we use

$$\bar{z}_{q,i,r,j} = \begin{cases} (1 - R_r) z_{q,i,r,j} & \text{if } r = \text{Tohoku} \\ z_{q,i,r,j} & \text{otherwise} \end{cases} \tag{6.15}$$

instead of $z_{q,i,r,j}$. R_r is a "regional ratio of damage" of region r defined by

$$R_r = D^{all}/C_r \tag{6.16}$$

where C_r represents the capital stock of region r estimated by

$$C_r = \frac{P_r}{P_{total}} \cdot C_n \tag{6.17}$$

P_r is the production value of region r, and P_{total} is the total production value of all regions. P_r and P_{total} are observed from the inter-regional input–output table [20].

Second, in the analysis regarding ISBD, we estimate the damage on IT systems, D^{IT}, by

$$D^{IT} = D^{all} t_{\text{total}} \tag{6.18}$$

where t_{total} is the ratio of IT capital stock given by

$$t_{\text{total}} = \text{IT}_{\text{total}}/(\text{IT}_{\text{total}} + \text{nIT}_{\text{total}}) \tag{6.19}$$

where IT_{total} denotes the total amount of IT capital stock, and $\text{nIT}_{\text{total}}$ denotes the total amount of non-IT capital stock.

To further investigate the effects from investment in information security, we assume that the investment will reduce the damage from disasters such as earthquakes. In particular, we assume that a pre-disaster investment in information security can damage avoid much of the disaster by a certain *degree of improvement*, Deg. So we replace R_r in (6.15) with

$$\tilde{R}_r = (1 - \text{Deg}) D^{IT}/C_r \tag{6.20}$$

in our analysis. We set the degree of improvement as 10 % (therefore, Deg = 0.1) as a first estimation, but the same methodology can be used for more detailed analysis with different degrees.

6.6 Results and Discussions

6.6.1 Sectoral and Regional Interdependency

First, we analyze the sectoral and regional interdependencies before the earthquake. The dataset with 53 industrial sectors is used to analyze more details for Agriculture (01) and Financial, Insurance, and Real Estate (10) because these two sectors showed very low values of ISBD in the analysis based on the 12-sector dataset.

Suppose that we want to see the BD between a *pair* of groups (a supply-side group and a demand-side group). By using a heuristic threshold ISBD = 0.01 %,[1] we say "dependent" if ISBD is larger than or equals this threshold, and "not dependent" otherwise. We count the number of dependent pairs to see regional and sectoral interdependencies.

6.6.1.1 Sectoral Interdependency

The results regarding sectoral interdependency can be summarized by Table 6.4.

In Table 6.4, different symbols indicate different levels of interdependency as follows. For example, let us look at the sixth row of Table 6.4. The i-th element of this row shows the level of interdependency between the demand-side sector Manufacturing-Other (06) and the supply-side sector i. When we evaluate the

[1] In our raw result, the average mean value of ISBD is 0.00754 %. By considering this mean value and the standard deviation, we set the threshold.

Table 6.4 Summary of sectoral interdependency of information security

Demand-side sector name (ID)	Sector ID of supply-side sector (Sector ID)											
	01	02	03	04	05	06	07	08	09	10	11	12
Agriculture (01)	–	–	–	–	–	–	–	–	–	–	–	–
Mining (02)	–	–	–	○	–	○○	–	○	○○	○	●	○○
Manufacturing-Food & Beverage (03)	○	–	○○	○	–	○○	–	○	○○	○	○	○○
Manufacturing-Metal (04)	–	–	–	○○	–	○○	○	○	○○	○	○	○
Manufacturing-Machinery (05)	–	–	–	○○	○○	○○	●	○	○○	○	○	○○
Manufacturing-Other (06)	–	●	–	○	–	○○	●	○	○○	○	○	○
Construction (07)	–	–	–	○○	○	○○	–	○	○	○	○	○
Utilities (08)	–	–	–	–	–	○○	○	○	○○	○	○	○○
Commerce & Logistics (09)	–	–	–	–	○	○○	●	○	○○	○	○○	○○
Financial, Insurance, and Real Estate (10)	–	–	–	–	–	–	–	–	–	–	–	○○
ICT (11)	–	–	–	–	–	○○	●	○	○○	○	○○	○○
Services (12)	●	–	○	○	○○	○○	○	○	○○	○	○	○○

interdependency level of this element, we compute the ISBD for each of the $9 \times 9 = 81$ pairs of the demand-side group in Sector 06, supply-side group in Sector i, and count the number of "dependent" pairs. The result of this counting is shown in the last row of Table 6.5. The largest element in this last row is the sixth row, and its value is 56. Then we compute the ratio of "the value of each element of this row" to this highest value. If the ratio is larger than or equals 50 %, we use the sign "○○" in the corresponding element in Table 6.4. Likewise, we use "○" if the ratio is between 10 and 50 %. We use "●" if the ratio is non-zero but less than 10 %. Finally, we use "–" if the ratio is zero. Since $(0/56, 2/56, 0/56, 8/56, 0/56, 56/56, 2/56, 9/56, 28/56, 11/56, 10/56, 22/56) = (0, 0.036, 0, 0.143, 0, 1, 0.036, 0.161, 0.500, 0.196, 0.179, 0.393)$, the sixth row of Table 6.4 is

$$(–, ●, –, ○, –, ○○, ●, ○, ○○, ○, ○, ○).$$

Table 6.4 shows that supply-side sectors of Manufacturing-Other (06), Commerce & Logistics (09), and Services (12) are the sectors highly dependent on demand-side sectors. We call these three sectors *critical sectors* or *influential sectors*. Demand-side sectors have high likelihood to be affected by security

Table 6.5 Number of dependent pairs for demand-side sector of Manufacturing-Other (06)

Demand-side region name (ID)	Number of dependent pairs for each supply-side sector (Sector ID)											
	01	02	03	04	05	06	07	08	09	10	11	12
Hokkaido (A)	0	1	0	0	0	6	0	1	2	1	1	2
Tohoku (B)	0	0	0	1	0	6	0	1	3	2	1	2
Kanto (C)	0	0	0	1	0	7	1	1	2	1	1	2
Chubu (D)	0	0	0	3	0	6	0	1	3	2	2	3
Kinki (E)	0	0	0	1	0	7	1	1	3	1	2	2
Chugoku (F)	0	0	0	1	0	6	0	1	4	1	1	3
Shikoku (G)	0	0	0	0	0	6	0	1	5	1	1	3
Kyushu (H)	0	0	0	1	0	6	0	1	4	1	1	3
Okinawa (I)	0	1	0	0	0	6	0	1	2	1	0	2
Total	0	2	0	8	0	56	2	9	28	11	10	22

incidents in the critical sectors. Likewise, Table 6.4 shows that the demand-side sectors of Machinery (05) and Services (12) are the most *influenced sectors*.

By observing Table 6.4 in more detail, we can classify demand-side sectors into the following five classes.

Class 1: *Sectors which show high interdependency* ***when and only when*** *tested with the critical sectors*. Mining (02) and Utilities (08) belong to this group.

Class 2: *Sectors which show high interdependency when tested with its own sector and* ***all*** *of the critical sectors*. Manufacturing-Food & Beverage (03), Manufacturing-Machinery (05), Commerce & Logistic (09), ICT (11), and Services (12) belong to this group.

Class 3: *Sectors that show high interdependency when tested with their own sectors and* ***not all but some*** *of the critical sectors*. Manufacturing-Metal (04) and Manufacturing-other (06) belong to this group.

Class 4: *Sectors which shows little interdependency when tested with supply-side sectors*. Although Financial, Insurance, and real estate (10) belong to this group, our detailed analysis by using the 53-sector dataset shows that the sub-sector Financial and Insurance (0400) shows characteristics similar to those of Class 3.

Class 5: *The other demand-side sectors.* Agriculture (01) and Construction (07) belong to this group. These two sectors show no interdependency when tested with their own sectors.

We can see that the demand-side sectors with high self-dependency (i.e., the sectors in Class 2 and Class 3) do not show high interdependency with non-critical sectors. Since investment advice regarding self-dependency and critical sectors are trivial, they need to learn from the analysis of regional interdependencies. Paying attention to the fact that the majority of sectors belong to these two classes, we notice the importance of regional interdependency analysis.

Table 6.6 Summary of regional interdependency of information security

Demand-side region name (ID)	Region ID of supply-side region (Region ID)								
	A	B	C	D	E	F	G	H	I
Hokkaido (A)	○○	○	○○	○	○	○	●	●	—
Tohoku (B)	●	○○	○○	○	○	●	—	●	—
Kanto (C)	●	●	○○	○	○	●	●	●	—
Chubu (D)	●	●	○○	○○	○	○	●	○	—
Kinki (E)	●	●	○○	○	○○	○	●	●	—
Chugoku (F)	—	●	○○	○	○	○○	●	○	—
Shikoku (G)	—	●	○○	○	○○	○	○○	○	—
Kyushu (H)	—	●	○○	○	○	○	●	○○	—
Okinawa (I)	●	●	○○	○	○	○	●	○	○○

6.6.1.2 Regional Interdependency

The results regarding regional interdependency can be summarized in Table 6.6 where different symbols indicate the different levels of interdependency in the same way as in the sectoral interdependency analysis. In Table 6.6, we can see the economic scale of a region has a great influence on the characteristics of the interdependency, and most of the results are intuitively easy to accept; for example, on the supply-side, Kanto (the economically largest region) is the most influential.

As a remarkable (somewhat counter–intuitive) point, on the demand-side, Tohoku (economically middle-sized) has the same features (i.e., less influenced) as Kanto. Also, the features regarding the highly influenced sectors are quite different from those of the highly influenced regions. From the regional perspective, we found that the highly influenced regions likely have small economic scales. By contrast, from the sectoral perspective, the two highly influenced sectors, Machinery (05) and Services (12), have large economic scales.

6.6.2 Impact of the Earthquake

Based on the government's announcement about the damage mentioned in Sect. 6.5, we set the following four testing scenarios:

Case 1a: Full damage from the earthquake. The full amount of 9 trillion yen is used as the damage value.

Case 1b: Damage from the earthquake with some reduction by investment in information security. The amount of 9 trillion yen with 10 %-reduction is used as the damage value.

Table 6.7 ISBD reduction (in terms of the number of *missing* dependent pairs) from the sectoral perspective in the investigation of the impact of the Great East Japan Earthquake

	Supply-side Sector ID												
	01	02	03	04	05	06	07	08	09	10	11	12	Total
Case 1a	0	0	1	0	1	5	0	1	3	9	2	0	22
Case 1b	0	0	1	0	1	5	0	1	3	9	2	0	22
Case 2a	0	0	1	2	2	6	1	4	7	10	5	0	38
Case 2b	0	0	1	2	2	6	1	4	7	10	5	0	38

Case 2a: Full damage from the earthquake and the consequent tsunami. The full subsequent amount of 16 trillion yen is used as the damage value.

Case 2b: Damage from the earthquake and the subsequent tsunami with some reduction by investment in information security. The amount of 16 trillion yen with 10 %-reduction is used as the damage value.

In each of the four cases, we did the following:

1. Count the number of dependent pairs (demand-side group in Tohoku and supply-side group in Sector i) before the earthquake, N_i.
2. Count this number after the earthquake, N_i'.
3. Compute the reduction of this number (i.e., $N_i - N_i'$). We refer to this reduction as the number of *missing* dependent pairs.

We obtained Table 6.7 using the above procedure. The reduction of interdependency is more likely with the following sectors: Financial, Insurance, and Real Estate (10), Manufacturing-Other (06), and Commerce & Logistics (09). It should be noted that Manufacturing-Other (06) and Commerce & Logistics (09) are critical sectors identified by the basic analysis in Sect. 6.6.1.1 but Financial, Insurance, and Real Estate (10) is not a critical sector. The above characteristics are not changed by the reduction of damages through prior security investment.

Likewise, in each of the four cases, we did the following.

1. Count the number of dependent pairs (demand-side group in Tohoku and supply-side group in Region q) before the earthquake.
2. Count this number after the earthquake.
3. Compute the reduction of this number. We refer to this reduction as the number of *missing* dependent pairs.

We obtained Table 6.8 using the above procedure. The reduction of interdependency is concentrated in two patterns: one is between sectors inside Tohoku (B), and the other is between sectors in Tohoku (B) and those in Kanto (C). Thus the earthquake impacted the most damaged region (Tohoku) and the economically largest region (Kanto) most significantly. This feature is not changed by the reduction of damages through prior security investment.

Table 6.8 ISBD reduction (in terms of the number of *missing* dependent pairs) from the regional perspective in the investigation of the impact of the Great East Japan Earthquake

	Supply-side Region ID									
	A	B	C	D	E	F	G	H	I	Total
Case 1a	3	8	5	3	2	0	0	1	0	22
Case 1b	3	8	5	3	2	0	0	1	0	22
Case 2a	3	14	11	3	6	0	0	1	0	38
Case 2b	3	14	11	3	6	0	0	1	0	38

6.7 Conclusion

In this chapter, we have presented our empirical study on sectoral and regional interdependencies under the influence of information security in Japan from the demand-side perspective.

In our main study, first the economic scale of a region has a great influence on the characteristics of the interdependency. For example, the security problems of economically larger supply-side regions tend to affect demand-side firms more significantly. Second, we observed that there are three supply-side sectors which are *critical* in the sense that information security problems in the three sectors can highly affect the demand-side sectors. Another common features of the three critical sectors (Manufacturing-Other, Commerce & Logistics, and Services) is that they have high self-dependencies.

As an extended study, we investigated the impact of the Great East Japan Earthquake by evaluating interdependency reductions caused by the earthquake. The results are consistent with the results of our main study; the role of the critical sectors is very important in Japan. We also found that the earthquake impacted the most damaged region (Tohoku) and the economically largest region (Kanto) most significantly. These features are not changed through the reduction of damages by prior security investment.

Both in the basic study and in the extended study, we can see that considering not only sectoral perspective but also regional perspective is very helpful in empirical analyses related to the interdependency under the influence of information security. By analyzing the sensitivity of interdependency to changes in an inter-regional input–output table in a wide variety of scenarios, there are many possibilities of more extended studies based on our methodology. For instance, an analysis regarding a large-scale earthquake in Kanto expected in the near future would bring important implications and suggestions since there are many predictions about such earthquakes. A limitation of our study so far is the use of some heuristic parameters. Our future work would include additional analyses to overcome this limitation. For instance, we could suggest time-series analysis for setting a proper estimated degree of improvement of pre-disaster investment in information security.

Acknowledgements We would like to express our sincere gratitude to the Volkswagen Foundation for their kind support with a travel grant to attend the Workshop on the Economics of Information Security (WEIS 2012) in Berlin. We would like to also thank the anonymous reviewers for their valuable comments.

References

1. Anderson, R., Moore, T.: The economics of information security. Science **314**(5799), 610–613 (2006)
2. Aoyama, Y., Ratick, S.J.: Trust, transactions, and information technologies in the U.S. logistics industry. Econ. Geogr. **83**(2), 159–180 (2007)
3. Bandyopadhyay, T., Jacob, V., Raghunathan, S.: Information security in networked supply chains: impact of network vulnerability and supply chain integration on incentives to invest. Inf. Technol. Manage. **11**(1), 7–23 (2010)
4. Boot, A.W.A., Marinc, M.: The evolving landscape of banking. Ind. Corpor. Change **17**(6), 1173–1203 (2008)
5. Cabineet Office, Government of Japan: Gross capital stock by industry. Available via DIALOG. http://www.esri.cao.go.jp/jp/sna/sonota/minkan/kekka/20110107/h21y_stock_all.xls (2009)
6. Cabinet Office, Government of Japan: Special Cabinet Meeting Material on Monthly Economic Report due to the Earthquake. Available via DIALOG. http://www5.cao.go.jp/keizai3/getsurei-s/1103.pdf (2011)
7. Dedrick, J., Kraemer, K.L.: The impacts of IT on firm and industry structure: the personal computer industry. Calif. Manage. Rev. **47**(3), 122–142 (2005)
8. Dietzenbacher, E., Linder, J.A.: Sectoral and spatial linkages in the EC production structure. J. Reg. Sci. **37**(2), 235–257 (1997)
9. Fearon, C., Philip, G.: An empirical study of the use of EDI in supermarket chains using a new conceptual framework. J. Inf. Technol. **14**(1), 3–21 (1999)
10. Fearon, C., Philip, G.: Measuring success of electronic trading in the insurance industry: operationalising the disconfirmation of expectations paradigm. Behav. Inf. Technol. **27**(6), 483–493 (2008)
11. Gordon, L.A., Loeb, M.P., Lucyshyn, W., Richardson, R.: 2005 CSI/FBI Computer crime and security survey. Available via DIALOG. http://www.cpppe.umd.edu/Bookstore/Documents/2005CSISurvey.pdf (2005)
12. Haimes, Y.Y., Jiang, P.: Leontief-based model of risk in complex interconnected infrastructures. Int. J. Netw. Virtual Organ. **4**(3), 130–144 (2001)
13. Haimes, Y.Y., Horowitz, B.M., Lambert, J.H., Santos, J.R., Crowther, K., Lian, C.: Inoperability input–output model for interdependent infrastructure sectors. II: Case studies. J. Infrastruct. Syst. **11**(2), 80–92 (2005)
14. Haimes, Y.Y., Horowitz, B.M., Lambert, J.H., Santos, J.R., Lian, C., Crowther, K.G.: Inoperability input–output model for interdependent infrastructure sectors. I: Theory and methodology. J. Infrastruct. Syst. **11**(2), 67–79 (2005)
15. Han, K., Kauffman, R.J., Nault, B.R.: Information exploitation and interorganizational systems ownership. J. Manage. Inf. Syst. **21**(2), 109–135 (2004)
16. Hausken, K.: Income, Interdependence, and substitution effects affecting incentives for security investment. J. Account. Public Policy **25**(6), 629–665 (2006)
17. King, J.L., Lyytinen, K.: Automotive Informatics: Information Technology and Enterprise Transformation in the Automobile Industry. Transforming Enterprise: The Economic and Social Implications of Information Technology, pp. 283–312. MIT, Cambridge (2005)
18. Klein, R., Rai, A.: Interfirm Strategic Information Flows in Logistics Supply Chain Relationships. Manage. Inf. Syst. Q. **33**(4), 735–762 (2009)
19. Kunreuther, H., Heal, G.: Interdependent security. J. Risk Uncertain. **26**(2–3), 231–249 (2003)

20. Ministry of Economic, Trade and Industry: Inter-regional input–output tables 2005. Available via DIALOG. http://www.meti.go.jp/statistics/tyo/tiikiio/result/result_02.html (2005)
21. Ministry of Economic, Trade and Industry: The 2006 Survey of Information Technology. Available via DIALOG. http://www.meti.go.jp/statistics/zyo/zyouhou/result-2/h18jyojitsu.html (2007)
22. Office of the Manager, National Communications System: Supervisory Control and Data Acquisition (SCADA) Systems: Technical information bulletin 04-1, National communications system. Available via DIALOG. http://www.ncs.gov/library/tech_bulletins/2004/tib_04-1.pdf (2004)
23. Ogut, H., Menon, N., Raghunathan, S.: Cyber insurance and IT security investment: impact of interdependent risk. In: 4th Workshop on the Economics of Information Security (WEIS), Cambridge (2005)
24. Pierre, R., Timothy, D.: Modular strategies: B2B technology and architectural knowledge. Calif. Manage. Rev. **47**(4), 86–113 (2005)
25. Research Institute of Economy, Trade and Industry: Japan industrial productivity database 2008. Available via DIALOG. http://www.rieti.go.jp/jp/database/JIP2008/index.html (2008)
26. Shinozaki, A., Yamamoto, Y., Yamazaki, S.: Technical papers on ICT related economics. No. 11-1: Estimation on the amount of damage on ICT-related capital stock. Available via DIALOG. http://www.icr.co.jp/ICT/report/TP_201106.pdf (2011)
27. Tanaka, H.: Geography and information security: does location affect information security effort? In: Fourth Forum on Financial Systems and Cyber Security: A Public Policy Perspective, Smith School of Business, University of Maryland (2007)
28. Tanaka, H.: Quantitative analysis of information security interdependency between industrial sectors. In: Proceedings of the 3rd International Symposium on Empirical Software Engineering and Measurement, Lake Buena Vista, pp. 574–583 (2009)
29. The Institute for Information Infrastructure Protection (I3P): Security Solution for the Oil and Gas Industry, Technology Fact Sheet. Inoperability input–output model (IIM). Available via DIALOG. http://www.dartmouth.edu/~i3p/docs/publications/IIM-factsheet-Feb2007.pdf (2007)

Chapter 7
Can We Afford Integrity by Proof-of-Work? Scenarios Inspired by the Bitcoin Currency

Jörg Becker, Dominic Breuker, Tobias Heide, Justus Holler, Hans Peter Rauer, and Rainer Böhme

Abstract Proof-of-Work (PoW), a well-known principle to ration resource access in client-server relations, is about to experience a renaissance as a mechanism to protect the integrity of a global state in distributed transaction systems under decentralized control. Most prominently, the Bitcoin cryptographic currency protocol leverages PoW to (1) prevent double spending and (2) establish scarcity, two essential properties of any electronic currency. This chapter asks the important question whether this approach is generally viable. Citing actual data, it provides a first cut of an answer by estimating the resource requirements, in terms of operating cost and ecological footprint, of a suitably dimensioned PoW infrastructure and comparing them to three attack scenarios. The analysis is inspired by Bitcoin, but generalizes to potential successors, which fix Bitcoin's technical and economic teething troubles discussed in the literature.

7.1 Introduction

Proof-of-Work (PoW) is a principle to artificially impose transaction costs in the absence of a payment system. The idea is to "charge" the requester of a service with the effort to present a solution to a problem that is much harder to solve than to verify. This way, PoW can help to ration access to services which would otherwise be abused. Originally presented by Dwork and Naor [14] as a mechanism to combat junk email, it has been proposed as a solution for numerous other situations in which the goal is to prevent some sort of fraudulent use, e.g., when measuring the number

J. Becker · D. Breuker (✉) · T. Heide · J. Holler · H.P. Rauer · R. Böhme
European Research Center for Information Systems (ERCIS), University of Münster, Leonardo-Campus 3, 48149 Münster, Germany
e-mail: becker@ercis.de; breuker@ercis.de; heide@ercis.de; holler@ercis.de; rauer@ercis.de; rainer.boehme@uni-muenster.de

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_7,

of visitors a website has [20]. Whether PoW actually has any practical relevance has been debated in the literature, mainly for the case of spam prevention (see [28] and [30]).

A particularly innovative use of this principle has been proposed by Nakamoto [34], who used it as a core component in designing a fully decentralized peer-to-peer electronic currency system called Bitcoin. In this system, users constantly participate in a lottery, and each user's chance of winning is proportional to the computing power he is willing to invest. The task is to modify a document until its hash is of a particular structure. In parallel, users publicly announce transactions of Bitcoins, thereby expressing their intention to transfer a certain amount of this currency to another user. The lottery is designed in such a way that every win, as a side-effect, returns a timestamp of all transactions in an atomic operation. Furthermore, the user who won the lottery is rewarded for his effort by receiving new Bitcoins as well as fees for timestamped transactions. Other users accept this timestamp after validating that the PoW has been delivered. Then, the next round of the lottery starts. The sequence of timestamps forms a history of transactions. In case that competing histories emerge (which is easily possible in a peer-to-peer network), users believe in the history for which the most PoW has been delivered. This quickly resolves the conflict. Altering any transaction timestamped in the past requires redoing all work that has been done afterwards. As this becomes more unlikely the longer a transaction has been timestamped, manipulation is prevented and users collectively agree on a single history of transactions, thereby determining Bitcoin ownership.

As an advantage of such a decentralized currency, Nakamoto points out that electronic payment nowadays heavily relies on central authorities processing them. Widely trusted (but not necessarily trustworthy) financial institutions handle electronic payments and ensure the integrity of the system's global state. In return, they charge society for this service. The goal of Bitcoin is to replace trust in financial institutions with trust in PoW, thereby eliminating the need for financial institutions and with it the fees they charge. However, running the Bitcoin system is not for free either. As computational power is required for the PoW, hardware has to be acquired, powered, and maintained.

The Bitcoin system is secure as long as no single party controls more than 50 % of the network's computing power. If a party would, it could redo work of the past and ultimately outperform the honest part of the network. While digital signatures on transactions prevent arbitrary manipulations, the attackers would have the power to double-spend their own Bitcoins, thereby generating profits for themselves, or to prevent transactions from being timestamped, thereby undermining the trust in this system. The viability of the system depends on the assumption that nobody will ever gain this power.

These considerations suggest that the overall computational power required to run a system like Bitcoin depends on security considerations. If computational power was low enough to allow a single party to gain control over 50 % of it, no one could trust in PoW anymore. Consequently, the cost of running the system depends on security requirements, as computing power drives costs. The aim of this

chapter is to propose an estimation of what the cost-saving potential of an electronic currency relying on PoW could be. To accomplish this, we present three different attack scenarios. In each of them, an individual or a group acquires control over computing power with the purpose of compromising the integrity of the system. Estimating their computational power delivers an idea of what we would need to defend the system. As a point of reference, we also estimate the transaction costs that a central electronic payment system would produce if used globally, and imagine a distributed network for PoW generation of equal costs. Comparing this network with the attackers allows us to estimate by how much the network could be downscaled securely.

Another interesting aspect is the environmental impact of large-scale PoW application. As electricity costs constitute a large part of the total costs of providing computing power, running a PoW-based currency system would consume a considerable amount of power. In turn, this means it could be responsible for a substantial amount of carbon dioxide (CO_2) emissions and may contribute to global warming. Therefore, we make an interesting detour in estimating the CO_2 footprint of a global PoW-based electronic currency.

The remainder of this chapter is structured as follows. Section 7.2 discusses details of the Bitcoin system, in particular the role of PoW, and explains how an attack on a currency relying on PoW can be accomplished. In Sect. 7.3, previous research regarding Bitcoin is surveyed. Section 7.4 then presents our proposal for estimating potential cost savings as well as the environmental impact of PoW-based currencies. Results from this estimation as well as its limitations are discussed in Sect. 7.5. Finally, Sect. 7.6 concludes and provides an outlook on future research.

7.2 The Bitcoin System

Possessing a certain amount of a currency means possessing the promise that one can collect favors from others in the future whose value is equal to those one had to give to others to acquire the amount. Thus, scarcity is a necessary property of anything that is used as a currency. If it could be produced at low cost, devaluation would destroy the trust in this promise. Building upon this property, it is also necessary that the currency can be transferred from one person to another, ideally in any quantity. The transaction mechanism must make sure the currency is correctly transferred in the sense that the overall amount of currency is preserved and that the transaction cannot be reversed later on. We will call this integrity.

For ordinary physical cash, both scarcity and integrity are enforced by the laws of nature, i.e., usage of physical tokens (also, legislation imposes constraints on counterfeit money, theft, ...). With electronic currencies, however, the amount of currency one possesses is nothing but an account balance. Thus, it requires a financial institution to manage these accounts and enforce scarcity and integrity. Bitcoin [34] has been proposed as a distributed peer-to-peer accounting system accomplishing this without relying on trust in a central authority.

A Bitcoin is represented by a chain of publicly announced transactions every participant of the network is aware of. Each transaction contains a public key signifying the owner of the Bitcoin. It further contains a hash of both the previous transaction and this public key, thereby entangling the previous transaction with the next one. The previous owner signs this hash using his private key. Any user can now validate whether the signature of the transaction matches the public key of the previous owner. He will only accept the transaction if it does. Thus, knowledge of the private key enables a user to spend a Bitcoin.

As long as users keep their private keys secret, this mechanism prevents potential attackers from spending Bitcoins they do not own. Nothing however stops them from spending those they once received twice. What is required to prevent double spending is a consensus on a temporal ordering of transactions among all users. This way, the current owner of a Bitcoin can always be determined. Attempts of previous owners to spend it again can be detected.

The temporal order is established by what is called a block chain. Blocks collect transactions combined with a hash of the previous block, thereby creating a chain. A block is valid only if it exhibits a special property which proves that a certain amount of work has been put into its creation. In particular, blocks contain a nonce value. It must be chosen by the creator in such a way that, when hashing the block, the hash starts with a certain number of zeros. As for an ideal hash function, this can only be achieved by randomly trying many different nonce values. The probability of finding a block depends on the number of trials. It can be adjusted globally by changing the number of leading zeros the hash must have. Regular adjustments ensure that new blocks are found on average each 10 min.

Via the block chain users collectively agree on the temporal order of transactions as defined by the order of the blocks containing them. Users always believe in the longest valid block chain they are aware of. As long as the majority of users are honest, no single attacker will be in the position to deliver the PoW necessary to change the temporal order to his advantage. Removing a transaction one did an hour ago (i.e., about 5–6 blocks in the past) from the block containing it would not only require recreating this block but also all subsequent ones, since the hash of the altered block is part of the next, and so forth. Thus, it quickly becomes computationally intractable to alter blocks deep within the chain.

To motivate users to participate in creating blocks, optional transaction fees can be paid by users creating transactions to users ironing these into a new block. Furthermore, a special transaction rewards the creator with a certain amount of newly created Bitcoins (in the form of a special initial transaction preceding every valid Bitcoin). This rewarding mechanism – called Bitcoin mining – constantly issues new currency. The amount is reduced at regular intervals and will eventually stop by convention. Once stopped, all Bitcoins are in circulation and the system will solely be in the transaction phase (as opposed to the mining phase in which new Bitcoins are created). Consequently, the scarcity of Bitcoins is ensured by eventually stopping the Bitcoin supply, and integrity is ensured by digitally signed transactions timestamped in a publicly visible block chain receiving credibility through the work invested in its creation.

As pointed out before, the underlying assumption is that the majority of users (in terms of computing power) are honest, which appears reasonable in a large distributed peer-to-peer network. If however a single user gains control over the majority of the (distributed) computing power, he is able to manipulate the temporal order of transactions to his advantage. The attack is to spend a Bitcoin, which then gets incorporated into the honest block chain. This makes the recipient believe he now possesses the coin. The attacker secretly computes a second chain not containing this transaction and outpaces the honest one. Once he publishes it, users will believe in the attacker's chain and he can spend the coin again, as it now appears to be still his. Scròfina [37] discusses several possibilities of double-spending and estimates their payoffs based on the Bitcoin system as it currently is.

It is important to notice that the motivation for attacking the system is not necessarily to profit from the attack. It might also be to simply destroy the system, e.g., as a form of terrorism. In case a destructive attack is launched, an attacker could for instance prevent any transaction from being timestamped. Effectively, no user could be sure anymore that he actually received a Bitcoin in a transaction, which would quickly destroy trust in the currency and devalue it. Any of the attack scenarios we analyze in this chapter is potentially destructive, i.e., we do not require the attack to be profitable.

7.3 Related Work

Although the idea of cryptographic electronic currencies came up more than two decades ago [29], it took until today for one of them to be widely discussed in media and research. With Digicash [9], a digital currency was presented in 1990. A cryptographic protocol allowed for anonymous payments and copying money was impossible. In contrast to the Bitcoin system, it relied on a central authority. While gaining quite some attention, it never had a significant breakthrough. Several other attempts to establish electronic currencies, e.g., E-Gold [48],[1] met a similar fate.

In 2008, the blueprint for the Bitcoin system, a decentralized cryptographic currency, was published [34]. The system draws on the principles of b-money [12] and enhances the security of previous electronic currencies with the PoW-based timestamping mechanism. Apart from presumably low transaction costs, a heavily advertised advantage of Bitcoin is that no financial institution has the power to exercise control over money creation or transactions.

While Bitcoin is "hyped" by some, there are several downsides to it as well. Most consumers in e-commerce prefer prices in their local currency [23]. With the heavily fluctuating exchange rates currently observed, it is hard to price goods [22]. Concerns regarding massive deflation have been expressed as well [27]. Yet

[1] Ultimately the E-Gold company pled guilty to money laundering and to operating an unlicensed money-transmitting business [39] and was shut down immediately.

this deflation also creates incentives for early adopters to promote the currency in order to profit from changes in exchange rates later on. Hence, some critics describe Bitcoin as a Ponzi scheme which over-rewards early adopters [4,22]. Bitcoin is also a topic for legal scholars, who ask the question how law should deal with such a currency (see e.g., [23]). Being a cryptographic currency using a public/private key mechanism, anonymity is a feature often promised, but contested in the literature [23,36]. In fact, unlike Digicash, Bitcoin has never been designed to be anonymous.

As for security, there are numerous threats to Bitcoin users other than double-spending. For instance, losing one's private keys due to computer malware means losing one's Bitcoins [10,13,23]. Also, denial-of-service attacks on the underlying peer-to-peer communication infrastructure or the Bitcoin exchanges (websites which allow users to exchange Bitcoins for other currencies) pose a threat [23,34].

Not only is the current technical implementation of the system subject to critique, but also the protocol itself. Babaioff et al. [1] claim that nodes in the Bitcoin network have an incentive not to propagate transactions. They also propose a modification to solve this problem. Several other problems have been identified [3].

However, any threats or problems discussed above are specific to the current implementation of the Bitcoin protocol, not to the idea of using a PoW-based timestamping mechanism to ensure the integrity of an electronic currency. To the best of our knowledge, no scholar has yet analyzed the economic or ecologic implications of a large-scale application of this mechanism. Therefore, we present a hypothetical scenario of a PoW-based currency system in the next chapter. The only threat we consider is the fundamental assumption that no single party can gain control over 50 % of the total computation power.

7.4 Costs of a PoW-Based Currency

7.4.1 General Scenario

To compare the two systems, a PoW-based, Bitcoin-like currency and a centralized solution, we now consider the following scenario. First, we conduct the analysis as if the two systems were in use today. Thus, the scenario is based on technology currently available. We do not worry about transition problems or speculate about future states of the world, but compare steady states which are, albeit imaginary, inspired by the world as it is. Consequently, we assume that the temporary mining phase is over (or never existed, to counter the Ponzi critique), i.e., there is no fixed reward for creating a new block of the chain. The only reward stems from fees paid by those who carry out transactions.

Second, we assume that payments are handled using a single currency for all transactions worldwide. As a PoW-based currency obviously benefits from economies of scale, we chose to consider the largest conceivable network.

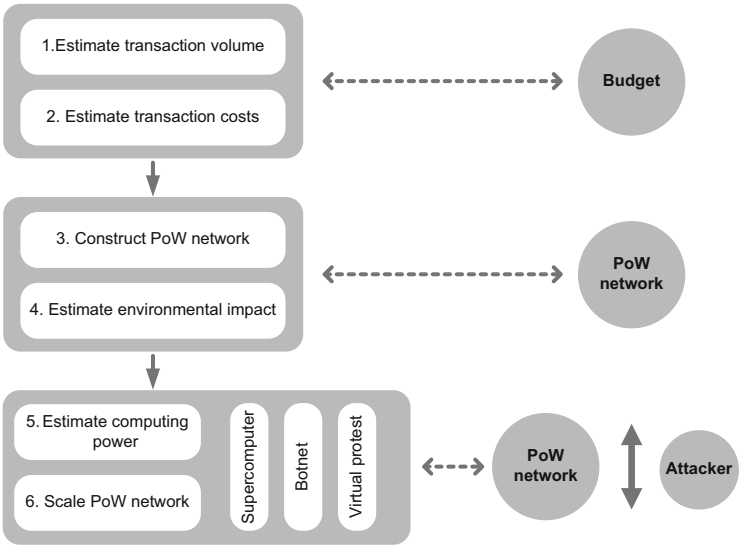

Fig. 7.1 Outline of the analysis

The most important assumption in our analysis is the system to which we compare the PoW network. The main goal of our analysis is to explore the effect of the replacement of a centralized institution by a PoW network. For this reason, it is necessary to compare the PoW network to a centralized system that is, apart from being centralized, as similar as possible. We have chosen a card payment system as it processes payments electronically, just as the PoW network does. Other costs, for instance for printing and distributing cash money, are certainly incurred in today's financial system, yet they would distort the analysis as they are not the result of using a centralized system but of demand for non-electronic payments. We also do not assume that financial intermediation vanishes completely as Bitcoin is no replacement for capital markets. It is just an instrument for payments.

In summary, we compare a rolled-out PoW network to a card payment system, assume that both systems process all ordinary payment transactions of the entire world, and we do not rely on forecasts but on financial estimates and the technology of the present.

7.4.2 Analysis Procedure

The idea of our analysis is outlined in Fig. 7.1 and structured as follows. First, we will estimate the volume of all transactions taking place anywhere in the world. Second, we estimate the transaction costs incurred in a system using a centralized currency as a fraction of the volume. Combining these two estimates, we end up

Table 7.1 Total value of transactions by payment instrument in US dollars, 2010

Payment instrument	Transaction volume (USD)
Credit transfers	$3.74 * 10^{14}$
Direct debit	$4.37 * 10^{13}$
Check	$9.36 * 10^{13}$
E-Money	$1.80 * 10^{10}$
Card payments (debit/credit)	$9.45 * 10^{12}$

Source: Bank for International Settlements [2], comparative Table 9

with the worldwide total transaction costs of a centralized currency. It will serve as a reference budget for the remainder of the calculation. Third, we will design a PoW network using this reference budget, thereby dimensioning a PoW network that costs exactly as much as the centralized system. We will describe this network in terms of the computing power it could achieve. To keep it running, electricity is required. The generation of this electricity has an environmental impact, which is what we will estimate in a fourth step. In step 5 the computing power of potential attackers is estimated. We consider three different attack scenarios. In the sixth step, given the computing power of both the PoW network and the attacker, we can see if and to which extent the size of the PoW-network could be decreased such that it can still fight off the attack. Finally, this allows us to estimate the potential for decreasing transaction costs compared to a centralized solution. Additionally, the environmental impact can be estimated.[2]

7.4.3 Estimate of Transaction Volume

The Bank for International Settlements (BIS), located in Switzerland, publishes annual statistics on payment, clearing, and settlement systems. The reports provide data for 23 countries, among them the United States, Germany, China, Japan, France, the United Kingdom, Russia, Italy, Canada. We will use the transactions taking place in these countries as a proxy for those of the entire world. This appears justified as the economically strongest countries of the world are included. In the latest report [2], the BIS reports the total yearly transaction volume for all 23 countries, categorized by method of payment. Figures for 2010 can be found in Table 7.1.

Recall that the centralized system to which we compare the PoW network is an electronic card payment system, whose transaction costs will be estimated in step 2 as a certain fraction of the volume. Such systems are used to process

[2]All assumptions of the analysis are debatable. Therefore we provide the spreadsheet of our estimation online and invite readers to come up with their refined scenarios. It can be accessed via: http://dl.dropbox.com/u/1168860/DoesProofOfWorkPayOff.xlsx

small, cash-like transactions. Therefore, large credit transfers between corporations, professional investors, or nations should not be included in the estimate, for applying transaction fees of an electronic card payment system to them would be unreasonable. Consequently, we exclude the item *credit transfers* of Table 7.1 from our estimate. *Direct debit* on the other hand will usually represent cash-like transactions and is therefore included.

The next important item in the list is *check,* which is also a famous payment instrument. Looking at the data, one can see that more than 75 % of the check transaction volume stems from China and the United States ([2], comparative Table 9). To identify if these transactions are cash-like we have a look at the average volume per transaction ([2], comparative Table 9c). For China, it amounts to more than 52,000 USD per transaction, which indicates that the volume stems in large parts from very huge transactions. Thus, we exclude China's volume from our estimate. For the United States however, the average value amounts to only 3,000 USD. This indicates that, while there will surely be large transactions, there will also be a number of small, cash-like ones. Therefore, we include 50 % of the check transaction volume of the United States.

All other check transactions are included. This also applies to all *E-Money* transactions as well as *card payments,* delivering a transaction volume of $9.03 * 10^{13}$ USD in 2010. Transactions for which actual cash is being used are not included yet. For these, we do not know any reliable data source. However, statistics from the BIS report the total volume of ATM cash withdrawals for 2010, which amounts to $4.09 * 10^{12}$ USD ([2], comparative Table 13). Assuming that most of this cash is spent only once and then deposited into a bank account before being withdrawn again, we use this as an estimate for cash transactions. Adding it, we end up with a *total transaction volume of* $9.44 * 10^{13}$ USD.

7.4.4 Estimation of the Total Transaction Costs of Centralized Payment

Given an estimate of the transaction volume of all cash-like transactions taking place anywhere on the world, we now proceed with estimating the cost of these transactions charged by the financial system to the real economy. As a reference, we will use an electronic payment system, as for such systems market prices are known. In general, there are two different types of them, the first of which is a credit card system. Typically, credit card fees paid by merchants in different countries can vary around 1–3 % of the transaction volume ([40], Fig. 7.3). However, pricing of credit cards is subject to considerable suspicion. Retailers have filed numerous lawsuits in the past as they believe institutions issuing these cards exaggerate their costs [8]. Thus, a credit card system is not the ideal reference for comparison with a PoW network. In addition, credit cards are not only used for ordinary payments, but also offer a credit function, i.e., the card holder buys products on credit and pays for

them later. This is an additional service not offered by a PoW network like Bitcoin. Naturally, these services charge a credit risk premium. This further strengthens the belief that costs of a credit card system are inadequately high for the purpose of this analysis.

The second type of electronic payment system is a debit card system. In contrast to credit cards, payments made with a debit card are transferred immediately (within days) from the user's bank account to the recipient. It does not provide a credit function. Also, the transaction costs are considerably lower as compared to credit cards. Therefore, a debit card system is used for comparison to the PoW network, as it comes closest to a frictionless centralized payment system free of bells and whistles that have no counterpart in the PoW network. The debit card system used in Germany, called *electronic cash,* charges merchants 0.3 % of the transaction volume as a fee [17]. Applying this to the transaction volume estimated in step 1, we end up *with total transaction costs of* $2.83 * 10^{11}$ USD for an imaginary world in which all transactions are processed with a centralized system comparable to debit cards.

7.4.5 Size of a PoW Network

In the following step, we use the transaction cost of a centralized system as a budget to estimate the size of a corresponding PoW network. The first and most important question is what the components of that network should be. Today's Bitcoin network is run to a large extent by individuals. They use their PCs and other equipment to create new blocks. However, once block creation becomes a commercial activity, it is unlikely that any kind of equipment owned by individuals could compete with specialized hardware as it is found in data centers. Thus, creating blocks would quickly become unprofitable for them, leaving the market to professional players.

On the other extreme, economies of scale could result in only a very few data centers, owned by a small number of organizations that cover the entire market of PoW delivery. Such a setting would clearly not be the decentralized currency envisioned by the Bitcoin pioneers. However, given that the fate of the global financial system rests on the security of this system, the risk that somebody gained control over these data centers could under no circumstances be acceptable. Moreover, as the organizations running the data centers might collude to exploit the system, one would be forced to trust them. For these reasons, we assume for our scenario a scale between these two extremes. More precisely, we assume a large number of independent data centers distributed all over the world, each small enough to implement a governance regime with effective checks and balances.

Having defined the main building block of the PoW network, we now analyze the typical cost structure of a data center. According to Belady [5], there are three main components.

- *Acquisition cost:* The cost of acquiring hardware. Typically amounts to about 25 % of the total cost.

Table 7.2 Exemplar electricity prices in various countries

Country	Electricity price (USD/kWh)	Source
Germany	0.36	[18]
China	0.16	[47]
USA	0.11	[15]
Russia	0.10	[33]

- *Energy cost:* The cost of electricity required to run the data center. Typically amounts to about 30 % of the total cost.
- *Infrastructure cost*: The cost of providing the environment in which the data center is run (e.g., the rent for the building). Typically amounts to about 45 % of the total cost.

Of particular importance for this analysis are the costs for electricity, as these will be the main driver for pollution. We will therefore pursue the following course in the construction of the PoW network. An estimate of computing power is achieved with respect to consumed electricity, under the assumption that particularly energy efficient hardware is used. At the same time, we disregard the fact that such hardware might have higher acquisition cost. This means the data centers of the network are energy efficient but still have the above-mentioned cost structure.

It is worth noticing that we do not take into account the costs of the communication infrastructure required to operate the system. For instance, Kaminsky [26] expresses doubt that the decentralized architecture of the current Bitcoin network could scale up to the size of our PoW network. Other authors already investigate the aspect of scalability and provide suggestions for how a large-scale system could be designed [3]. For our calculation, we assume that smart protocols could overcome all problems. However, as we do not know how such a solution might look, we do not estimate any communication costs for now and focus solemnly on effort for delivering PoW.

With a total budget of $2.83 * 10^{11}$ USD, of which 30 % is being spent on electricity, we obtain an electricity budget of $8.49 * 10^{10}$ USD. The next task is to determine the amount of electricity that can be produced. Again, we use current market prices. Typical prices for different countries can be found in Table 7.2, a broad overview in the Wikipedia [44].

Several countries have very low electricity prices but are rather unimportant with respect to global energy production. Considering only countries with a major output of electricity, Russia has the lowest (0.10 USD/kWh). Consequently, this value is used as a lower bound for the cost of electricity. Given the electricity budget of $8.49 * 10^{10}$ USD, a total of $8.49 * 10^{11}$ kWh is available for computation. This equals $3.06 * 10^{18}$ Ws (Watt-seconds) after a unit conversion. Note that this is the amount of electricity available per year, since we have calculated it from annual transaction costs.

We will now estimate the computing power that could be sustained over the year. An important aspect is how computing power should be measured for this purpose. Typically, high-performance computing power is measured in floating

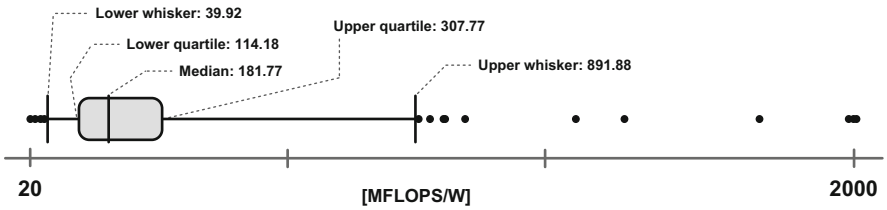

Fig. 7.2 Boxplot of the Green 500 November 2011 list

point operations per second (FLOPS). For instance, the Top 500 list, ranking the 500 most powerful supercomputers in the world, exclusively relies on this measure [38]. Delivering PoW in the Bitcoin system however heavily relies on computing hashes, for which integer operations are required.[3] Technically, the FLOPS measure should be replaced with a more appropriate metric such as megahash per second. Unfortunately, due to the widespread adoption of FLOPS, data on other metrics is not directly available. Even the Bitcoin community itself reports estimates of the network's computing power in terms of FLOPS [6]. Therefore, we adopt it for this analysis and discuss this as one limitation in Sect. 7.5.

Similar to the Top 500 list, the Green 500 list reports the 500 most energy-efficient supercomputers in the world [21] as measured in Mega-FLOPS per watt (MFLOPS/W). In the most recent ranking (November 2011), the most efficient one is an experimental computer from the IBM Blue Gene/Q project at IBM Rochester, achieving 2,026.48 MFLOPS/W. The performance of an "average" green supercomputer however is not even close to this value. In Fig. 7.2, a boxplot of the performance distribution in MFLOPS/W is presented, with the upper (lower) whisker being the 97.5 % (2.5 %) quantile.

This distribution indicates that most of the computers achieve a performance between 100 and 300 MFLOPS/W. For the "average" energy-efficient computer that we use to construct our hypothetical PoW network, we believe that a value of 181.77 MFLOPS/W is appropriate. It equals the median of the Green 500 supercomputers. Note that MFLOPS/W means million floating point operations per second per watt, i.e., million floating point operations per watt-second. To keep the notation clear, we will express the efficiency from now on in terms of (floating point) operations per watt-second (Ops/Ws).

Multiplying our estimate for energy-efficiency ($1.82 * 10^8$ Ops/Ws) with the available electricity delivers a total of $5.56 * 10^{26}$ Ops that can be achieved per year. Divided by the $3.15 * 10^7$ s of a year, the *computing power of the PoW network is* $1.76 * 10^{19}$ Ops/s.

[3] This is specific to the Bitcoin system as it currently is. It is also conceivable to design PoW functions that are better aligned with the optimization criteria of microprocessor architectures.

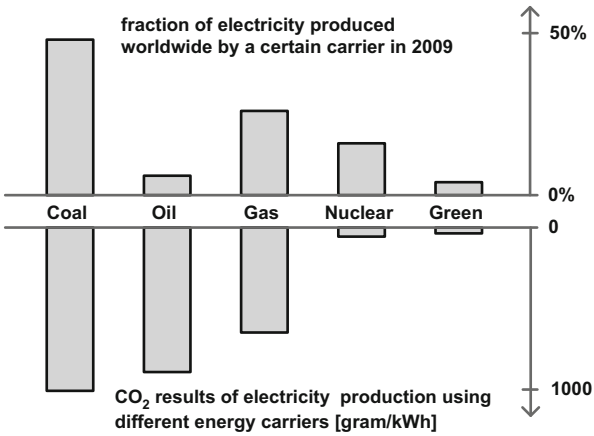

Fig. 7.3 Energy carriers: relative importance in 2009 vs. CO_2 emissions

7.4.6 Estimate of Environmental Impact

Given the annual power consumption of $3.06 * 10^{18}$ Ws, it is now straightforward to estimate the environmental impact of the PoW network. We will measure environmental impact in terms of CO_2 emissions. To start with this, information on how energy is being produced is required. Such data can be obtained from the International Energy Agency, which provides statistics on the worldwide production of electricity in 2009 with respect to the energy carrier [25]. The fraction of the total energy production each energy carrier is responsible for can be seen in the upper part of Fig. 7.3.[4]

Knowing the relative importance of each energy carrier, we can combine this information with corresponding average CO_2 emissions. Figures for that are provided by Lübbert [31] and can be seen in the lower part of Fig. 7.3.[5] Computing an average of the CO_2 emissions, weighted by the relative importance of each carrier, delivers an average CO_2 emission of about 718 g/kWh or, after a unit conversion, of $1.99 * 10^{-7}$ kg/Ws.

Multiplying this estimate with the power consumption of $3.06 * 10^{18}$ Ws, the PoW network is responsible for a total of $6.10 * 10^{11}$ kg of CO_2 per year. Compared to the total man-made CO_2 emissions due to fuel combustion in 2009, which have been $2.90 * 10^{13}$ kg [24], the PoW network at this scale would increase CO_2 emissions by more than 2.1 %. This is about the share of global commercial air traffic. We

[4]The list provided by the IEA (2012) includes an item "hydro" which consists to a large extent of energy produced by pumped storage plants. As electricity stored in these plants has been generated by other means, we exclude this item. We further exclude the items "waste" and "other sources" as their impact on the result is negligible.

[5]For some carriers, Lübbert [31] provides minimum and maximum estimates for CO_2 emissions. We use their averages in Fig. 7.3.

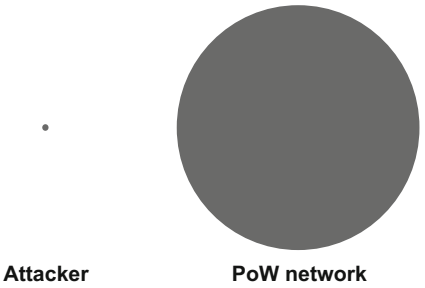

Fig. 7.4 Computing power of supercomputer attacker and PoW network. The area of the dots corresponds to the operations per second

assume the CO_2 emissions of the centralized system to be negligible compared to global emissions, making the effect of the PoW network a net increase.

7.4.7 Attack by Supercomputer

Whether the size of the constructed PoW network is adequate, too large, or too small can only be judged with respect to the computing power of potential attackers. Once known, the PoW network can be rescaled to the smallest size (with some headroom) such that it is still safe.

As a first naïve attack, we compare the network to a large supercomputer. In an extreme scenario, today's largest supercomputer on earth, namely the K computer at Riken in Japan, may attack the system. According to the Top 500 list of November 2011, it has a computing power of $1.05 * 10^{16}$ Ops/s [38]. Compared to the $1.76 * 10^{19}$ Ops/s of the PoW network, the attacker would only control about 0.06 % of the total computing power. Consequently, there is much potential to reduce the size of the network. More precisely, if it would be reduced to 0.12 % of its original size, the attacker was on par with the network. Figure 7.4 illustrates the computing powers of the PoW network and attacker.

7.4.8 Attack by Botnet

As a next idea, consider an attack in which a large botnet is leveraged to compete with the PoW network. There are mainly two variables that need to be estimated: the size of the botnet and the computing power of an average bot.

Measuring the size of botnets is a topic of active research. In [49], several different methods are being reviewed. Rajab et al. [35] point out the difficulties in generating reliable estimates. For instance, the botnet *Storm,* which was infiltrated by a security analyst of University of California, San Diego in 2007, has been estimated to consist of up to 50 million bots [32], yet the analysis of the security analyst revealed about 200,000 bots being online at a given time and a total of

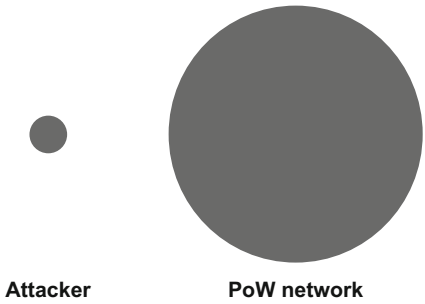

Fig. 7.5 Computing power of botnet attacker and PoW network. The area of the dots corresponds to the operations per second

1.5 million bots per day, thus indicating a considerably smaller size [16]. The article *Botnet* on Wikipedia contains a list of the largest botnets currently known, together with estimates of their size [43]. As we want to create a worst-case attack scenario, we assume the size of the attacking botnet equals the largest currently known botnet (called *BredoLab*), regardless of the fact that it might be an overestimation. In numbers, we assume the botnet controls 30 million bots.

Regarding the computing power of an individual bot, we draw an analogy. The Berkeley Open Infrastructure for Network Computing (BOINC) is a software platform designed to enable distributed scientific computing. It is used by numerous projects, among them the famous SETI@home initiative, and publishes statistics on the number of active users as well as total computing power [7]. At the time of writing, a total of 456,876 machines generate about $5.64 * 10^{15}$ Ops/s, which is an average of $1.23 * 10^{10}$ Ops/s per machine.

Assuming the same average computation power for any of the 30 million bots, the botnet can achieve a total of $3.70 * 10^{17}$ Ops/s. Compared to the PoW network, the botnet controls approximately 2.06 % of the total computing power. Figure 7.5 illustrates this ratio. The PoW network had to be shrunk to 4 % of its original size to allow the botnet to reach 50 % of the total power.

7.4.9 Attack by Virtual Protest ("Occupy Bitcoin")

In the original Bitcoin article, Nakamoto [34] uses the expression "one-CPU-one-vote" to describe the philosophy of PoW (p. 3). In a world in which all financial transactions – and with them all the world's economies – depend on a PoW network, we want to raise the question: What if a large number of CPU-owners vote against the system?

In particular, consider an attack scenario in which Internet activists acquire a large number of participants via social networks for a virtual protest. In the recent past, the Occupy Wall Street movement found millions of followers. Still, numerous protests are regularly being held all over the world [46]. October 15, on 2011, global protests, partly inspired by this movement, were held in more than 950 cities with

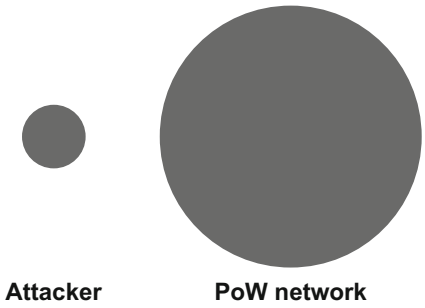

Fig. 7.6 Computing power of protesters and PoW network. The area of the dots corresponds to the operations per second

an estimated number of participants between 1 and 2 million people (for the 112 locations for which Wikipedia [42] provides data). In the protests against the Iraq war in 2003, the total number of participants is estimated to be around 36 million (between January 3 and April 12, [45]), with estimates for the peak value ranging between 6 and 30 million (on February 15, 2003 [41]).

Such figures demonstrate that a popular cause can easily unite several million people all over the world and make them cooperate to demonstrate and emphasize their collective opinion. We believe the existence of a PoW network would provide an opportunity for a new form of online virtual protest. If a group of activists provides easy-to-use software and enough participants willing to use it are found, protesters could try to collaboratively launch a destructive attack against the network. Effectively, this could undermine the trust to an extent such that no transactions are accepted anymore. Unlike ordinary protests, the economic consequences of a successful virtual protest could be devastating. This may be a reason for them being particularly attractive.

Apart from this, virtual protests also lower the effort a protester has to invest for participation. While ordinary protests require traveling to particular locations and spending a considerable amount of time there, virtual protests require only downloading and running software. With social networks such as Facebook, the infrastructure to coordinate protesters and distribute software is already in place. For these reasons, we believe that virtual protests could be much bigger than today's ordinary protests ever have been.

To construct a simple scenario, we assume that of the 845 million monthly active Facebook users [19], 10 % are motivated to participate in a virtual protest, e.g., against a war. They all agree to run a suitable piece of software at a particular day. Let the software be designed such that it prevents any transaction from being validated. Further assuming that each participant is the owner of a computer running at the same speed as the bots of our botnet ($1.23 * 10^{10}$ Ops/s), the attackers' computational power would be $1.04 * 10^{18}$ Ops/s. This is approximately 5.59 % of the total computing power when compared to the PoW network, which is visualized in Fig. 7.6. In this scenario, attackers and the network would be on par if the network was reduced to 12 % of its original size.

7.5 Discussion and Limitations

7.5.1 Results

Citing actual data and abstracting from frictions specific to the Bitcoin protocol, our analysis sheds light on the true cost of running a decentralized cryptographic currency secured by PoW. Our approach in this chapter is to estimate the size of the largest PoW network that could be constructed subject to a cost constraint given by a comparable centralized system, such as one that can process all electronic payments of the developed world today. In addition, we sketch three worst-case attack scenarios, each bounding the potential computing power of attackers from above. This is reasonable because an economy relying on the PoW-based currency as a primary means to process payments could not afford to be vulnerable to attack. As a side-effect, our analysis allows us to estimate the ecological footprint of such a PoW-based system.

The results of this estimation exercise suggest that the cost-saving potential might be smaller than claimed by Nakamoto [34] and hoped for by the Bitcoin proponents. Even with the large PoW network we considered in our estimation, cutting cost by only one order of (decimal) magnitude would already allow the attackers of the third scenario to overpower the network. Given that one would probably require a decent safety margin over what an attacker might achieve, costs may be cut at best to a fifth of the original amount; notwithstanding that a safety margin of factor two is by no means comparable to typical safety margins in cryptographic security, where the attacker's effort to break a system is calibrated to be at least $1.2 * 10^{24}$ times the defender's effort to use it. Furthermore, a range of political and social obstacles to implementing a PoW-based currency globally further limits the cost-saving potential. In particular, a realistic adoption scenario for a PoW network would unlikely be a "big bang" (think of a digital Bretton Woods). Thus, the transition is either very expensive, because the large-scale PoW network must run in parallel to conventional systems, or very risky, as smaller-scale networks remain vulnerable.

In addition, the cost savings would be bought dearly through a substantial increase in global CO_2 emissions. Even if the 2.1 % increase of the baseline network was reduced to a fifth, it would still be an increase of about 0.4 % only for direct power consumption. Not yet considered is the environmental impact of producing the required hardware as well as the impact of the communication infrastructure (which is assumed at zero cost and zero emissions in this analysis). In times of growing interest to reduce global emissions, it is questionable if the merits of a PoW-based system justify its environmental cost.

Table 7.3 Discussion of assumptions

Calculation step	Assumption	Comment
Transaction volume	Upper bound	The largest possible size for the system is that all transactions of the world are processed
Transaction cost	Upper bound	Using current market prices of centralized electronic payment systems bounds the costs from above as there is no incentive to pay more
Cost of electricity	Lower bound	Electricity is assumed to be produced at the lowest market price among all countries with major electricity output
Available electricity	Upper bound	Producing electricity at a lower-bounded price with an upper-bounded budget delivers an upper bound for the available amount
Cost of computing power	Lower bound	We calculate the cost of computing power in terms of power consumption and assume particularly energy-efficient hardware
Available computing power	Upper bound	Producing computing power at a lower-bounded price with an upper-bounded budget delivers an upper bound for the available amount
Environmental impact	Realistic	With respect to the network's power consumption, we believe that CO_2 emissions are fairly realistic given the data that has been used
Attackers' computing power	Upper bound	We consider worst-case attack scenarios with attackers that do not necessarily act economically rationally

7.5.2 Robustness

We see our calculation as a first step to understand the longer-term implications of PoW deployment. Of course our results depend on the validity of a number of assumptions. We are the first to admit that some of them are debatable. However, great care has been taken to ensure that our approximations are conservative. To back this up, we list all relevant assumptions in Table 7.3, state whether they represent upper or lower bounds, and argue why we chose so.

7.5.3 Limitations

Despite the care we have taken in the estimation, some limitations remain. First of all, the measure we use for computing power is FLOPS, while Bitcoin's PoW function is based on hashes. Integer operations or hashes per second would be more appropriate measures. Given that attack and defense are only one magnitude apart, even a small scaling factor could change our conclusions quite a bit. However, alternatives to Bitcoin's hash-based PoW function exist. And once PoW is going to be rolled out in a large infrastructure, demand for hardware optimized to calculate hashes will stimulate R&D and reduce its cost.

To curb the amount of speculation in the analysis, our scenarios are built as counterfactuals assuming that the PoW network is in use today. However, it actually may (or may not) be in use at some time in the future, when technological innovation might have changed the parameters of the scenario. Specifically in the last two of our attack scenarios, a specialized PoW network is compared to attackers who compose their networks largely out of desktop computing hardware. With growing interest in green high-performance computing, more energy-efficient specialized hardware might be developed. This would shift the relation between specialized and desktop hardware to the advantage of the PoW network. Moreover, current trends in end-user computing indicate that future consumer devices will be tablets and smartphones instead of desktop computers. Also, functionality is more and more provided through Web-based services. Thus, computational power will not be that important for these devices in the future, making it harder to assemble a competing network from them.

Lastly, our analysis ignores the cost of communication because this is more associated with achieving availability rather than integrity. Depending on the design of the underlying peer-to-peer communication network, this cost can be substantial or even prohibitive. Whether it is an advantage or disadvantage for the PoW network is hard to tell, because the two most threatening attack scenarios need extensive communications as well. Therefore we deem it tenable to ignore the costs of communication on both sides for a first and crude analysis.

A string of other limitations stands to reason. We have commented on most of them in the description of the analysis.

7.6 Conclusions and Future Research

The main goal of this research is to scrutinize the claim of Bitcoin proponents that a decentralized PoW-based currency charges society fewer transaction costs than a centralized electronic payment systems. This is connected to the more general question whether society can afford to use PoW to enforce the integrity of the global state in a distributed system. The answer to this question has implications on the design and governance of future information infrastructures.

We calculated a conservative cost estimate, accounting for both economic and ecological costs. If we take our results at face value, they indicate that a PoW network as costly as today's electronic payment systems could easily withstand attacks by a single supercomputer, and most likely defend against a very successful botnet or a social disobedience attack. However, its ecological footprint would be about the share of global commercial air traffic.

The system would still be secure if scaled down somewhat, but it is striking to see that attack and defense are only one (decimal) magnitude apart, although both estimates draw on completely independent inputs and involve many factors. We conclude that although our analysis does not discard PoW networks right away, the

question we asked is valid and, given the error margins, it is by no means certain that PoW networks are worth their price.

But this is not the end of the game. Future technological innovations could change the cost-benefit ratio of a PoW network completely, also in favor of PoW. For example, it is conceivable to reuse byproducts of PoW functions. More specifically, it could be of interest to develop PoW mechanisms that compute something useful. Right now, the activity of finding a nonce value such that the resulting hash satisfies a particular structure is in itself a waste of resources. It is an open research question to formulate relevant problems (e.g., complicated scientific computations, genome sequencing, protein folding) in a form such that a distributed network could solve them and the solution would be easily verifiable. Then anyone who wants a problem solved could formulate it as a PoW function and post a reward on it, thereby financing the PoW network. This could significantly reduce or even nullify the cost of the network and thus contest our result.

Another idea is to reuse the energy instead of (or in addition to) the computation result. Every computation converts electricity into heat pretty efficiently. Therefore, PoW could very well be used to help heat buildings. As decentralization is a key security principle behind PoW networks, a new kind of local compute-heat cogeneration appears much more feasible than transporting the waste heat of large datacenters to the places where it is needed. However, since the overall efficiency of electricity-based heating as compared to direct fuel combustion (such as natural gas heating) is rather low, heating with PoW cannot be expected to make up for all the cost.

Yet another aspect is that the timestamping service provided by the PoW network could serve purposes other than ensuring the integrity of a currency system. For instance, Clark and Essex [11] suggest using it to timestamp commitments in a cryptographic commitment scheme. If enough application scenarios are being found, it might be justified to finance a very large POW-based timestamping service as an infrastructure for all these services. Further research in the above-mentioned areas might eventually allow for a PoW network of an even larger scale than that considered in this chapter.

To conclude, we provided a first and very rough proposal to investigate the economic potential of PoW applied to ensure the integrity of electronic currency systems. While our results indicate potential for a moderate reduction of economic transaction costs, the ecological impact is substantial and would surely arouse public resistance if a PoW network were to be established. However, numerous possibilities have been discussed that could turn this result around and make a decentralized PoW timestamping service a valuable infrastructure for the future IT landscape. We hope that this chapter stimulates both discussions about and further research on these aspects.

References

1. Babaioff, M., et al.: On Bitcoin and red balloons. In: 13th ACM Conference on Electronic Commerce, Valencia (2011)
2. Bank for International Settlements: Statistics on payment, clearing and settlement systems in the CPSS countries – figures for 2010. http://www.bis.org/publ/cpss99.htm (2011)
3. Barber, S., et al.: Bitter to better – how to make Bitcoin a better currency. In: 16th International Conference on Financial Cryptography and Data Security, Bonaire (2012)
4. Barok, D.: Bitcoin: censorship-resistant currency and domain system for the people. http://pzwart3.wdka.hro.nl/mediawiki/images/6/64/Barok.bitcoin.pdf (2011)
5. Belady, C.L.: In the data center, power and cooling costs more than the IT equipment it supports. Electron. Cool. **13**(1), 24–27 (2007)
6. Bitcoinwatch.com: Bitcoinwatch. http://bitcoinwatch.com/
7. Boincstats.com: BOINC combined project statistics. http://boincstats.com/stats/project_graph.php?pr=bo
8. Bradford, T., Hayashi, F.: Developments in interchange fees in the United States and abroad. http://www.kc.frb.org/Publicat/PSR/Briefings/PSR-BriefingApr08.pdf (2008)
9. Chaum, D., et al.: Untraceable electronic cash. In: Goldwasser, S. (ed.) Advances in Cryptology – CRYPTO'88, Santa Barbara, pp. 319–327. Springer, New York (1988)
10. Chirgwin, R.: Bitcoin collapses on malicious trade – Mt.Gox scrambling to raise the Titanic. http://www.theregister.co.uk/2011/06/19/bitcoin_values_collapse_again/
11. Clark, J., Essex, A.: CommitCoin: carbon dating commitments with Bitcoin. In: 16th International Conference on Financial Cryptography and Data Security, Bonaire (2012)
12. Dai, W.: b-money. http://www.weidai.com/bmoney.txt (1998)
13. Doherty, S.: All your Bitcoins are ours... – Symantec connect community. http://www.symantec.com/connect/blogs/all-your-bitcoins-are-ours
14. Dwork, C., Naor, M.: Pricing via processing or combatting junk mail. In: Brickell, E.F. (ed.) Advances in Cryptology – CRYPTO'92, Santa Barbara, pp. 139–147. Springer, New York (1992)
15. EIA: Electric power monthly. http://www.eia.gov/electricity/monthly/
16. Enright, B.: Exposing stormworm. http://www.scribd.com/doc/2674816/exposing-storm (2007)
17. EURO Kartensysteme GmbH: Händlerbedingungen – Bedingungen für die Teilnahme am electronic cash-System der deutschen Kreditwirtschaft. http://www.electronic-cash.de/media/pdf/haendlerbedingungen.pdf (2008)
18. Europe's Energy Portal: Retail (end-user) energy prices for households. http://www.energy.eu/#domestic
19. Facebook: Fact sheet. http://newsroom.fb.com/content/default.aspx?NewsAreaId=22
20. Franklin, M.K., Malkhi, D.: Auditable metering with lightweight security. In: 1st International Conference on Financial Cryptography, Anguilla, pp. 151–160 (1997)
21. Green500.org: Green500. http://www.green500.org/
22. Grigg, I.: Financial cryptography: Bitcoin – the bad news. http://financialcryptography.com/mt/archives/001327.html
23. Grinberg, R.: Bitcoin: an innovative alternative digital currency. Hastings Sci. Technol. Law J. **4**, 160–208 (2011)
24. IEA: CO_2 Emissions from Fuel Combustion. http://www.iea.org/media/statistics/CO2highlights.pdf (2011). Last Access: Oct 2012
25. IEA: Electricity/heat in the world in 2009. http://www.iea.org/stats/electricitydata.asp?COUNTRY_CODE=29
26. Kaminsky, D.: Some thoughts on Bitcoins. http://www.slideshare.net/dakami/bitcoin-8776098
27. Krugman, P.: Golden cyberfetters. http://krugman.blogs.nytimes.com/2011/09/07/golden-cyberfetters/

28. Laurie, B., Clayton, R.: Proof-of-Work proves not to work. In: 3rd Annual Workshop on the Economics of Information Security, Minnesota (2004)
29. Levy, S.: E-Money (That's what I want). http://www.wired.com/wired/archive/2.12/emoney.html
30. Liu, D., Camp, L.J.: Proof of work can work. In: 5th Annual Workshop on the Economics of Information Security, Cambridge (2006)
31. Lübbert, D.: CO_2-Bilanzen verschiedener Energieträger im Vergleich – Zur Klimafreundlichkeit von fossilen Energien, Kernenergie und erneuerbaren Energien (2007)
32. McMillan, R.: Storm worm now just a squall. http://www.pcworld.com/article/138721/storm_worm_now_just_a_squall.html (2007)
33. Mosenergosbyt: Electricity tariffs for the population of the city of Moscow in 2012. http://www.mosenergosbyt.ru/portal/page/portal/site/personal/tarif/msk
34. Nakamoto, S.: Bitcoin: a peer-to-peer electronic cash system. https://www.crfdl.org/bitstream/handle/10838/959/bitcoin.pdf?sequence=1 (2008)
35. Rajab, M.A., et al.: My botnet is bigger than yours (maybe, better than yours): why size estimates remain challenging. In: Provos, N. (ed.) 1st Workshop on Hot Topics in Understanding Botnets, Cambridge, p. 5 (2007)
36. Reid, F., Harrigan, M.: An analysis of anonymity in the Bitcoin system. http://arxiv.org/abs/1107.4524 (2011)
37. Scròfina, S.: How can Bitcoin be hacked? https://www.quora.com/How-can-Bitcoin-be-hacked
38. TOP500.Org: Top500 http://top500.org/
39. US District Court for the District of Columbia: E-Gold indictment. http://www.justice.gov/criminal/pr/2007/04/CRM_07-301_042707_egold_indict.pdf (2006). Last Access: Oct 2012
40. Weiner, S., Wrigth, J.: Interchange fees in various countries: developments and determinants. Rev. Netw. Econ. **4**(4), 290–323 (2005)
41. Wikipedia: 15 February 2003 anti-war protest. http://en.wikipedia.org/wiki/February_15,_2003_anti-war_protest
42. Wikipedia: 15 October 2011 global protests. http://en.wikipedia.org/wiki/15_October_2011_global_protests#cite_note-atlantic-10
43. Wikipedia: Botnet. http://en.wikipedia.org/wiki/Botnet#cite_note-19
44. Wikipedia: Electricity pricing. http://en.wikipedia.org/wiki/Electricity_pricing#Global_electricity_price_comparison
45. Wikipedia: Protests against the Iraq war. http://en.wikipedia.org/wiki/Protests_against_the_Iraq_War
46. Wikipedia: Timeline of Occupy Wall Street. http://en.wikipedia.org/wiki/Timeline_of_Occupy_Wall_Street
47. Yang, J.: China to encourage solar use. http://webcache.googleusercontent.com/search?q=cache:deiCSQwS69cJ:online.wsj.com/article/SB124397202782578277.html+China+to+Encourage+Solar+Use&cd=1&hl=de&ct=clnk&gl=de
48. Zetter, K.: Bullion and bandits: the improbable rise and fall of E-Gold. http://www.wired.com/threatlevel/2009/06/e-gold/
49. Zhu, Z., et al.: Botnet research survey. In: Computer Software and Applications (COMPSAC'08), Turku, pp. 967–972 (2008)

Chapter 8
Online Promiscuity: Prophylactic Patching and the Spread of Computer Transmitted Infections

Timothy Kelley and L. Jean Camp

Abstract There is a long history of studying the epidemiology of computer malware. Much of this work has focused on the behaviors of specific viruses, worms, or botnets. In contrast, we seek to utilize an extension of the simple SIS model to examine the efficacy of various aggregate patching and recovery behaviors. We use the SIS model because we are interested in the global prevalence of malware, rather than the dynamics, such as recovery, covered in previous work. We consider four populations: vigilant and non-vigilant with infected or not for both sets. Using our model we show that small increases in patch rates and recovery speed are the most effective approaches to reduce system-wide vulnerabilities due to unprotected computers. Our results illustrate that a public health approach may be feasible, requiring a subpopulation adopt prophylactic actions rather than near-universal immunization.

8.1 Introduction

Studying the spread of computer malware through the use of epidemiological models is a useful tool in understanding the dynamics of individual outbreaks of malware, and provides some insight into possible mitigation policies. Kephart and White's early work on system-wide prevalence examined effects of topology on virus spread as well as the possibility of a social response to infection [18, 19]. Other work has focused on describing the dynamics of individual types of viruses, worms, or botnets.

In Kephart and White's examination of the social response, even a small social response was able to reduce significantly the total level of infection in the system. However, this result depends on a system where the recovered population could not

T. Kelley (✉) · L.J. Camp
School of Informatics and Computing, Indiana University, Bloomington, IN, USA
e-mail: kelleyt@indiana.edu; ljcamp@indiana.edu

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_8,

become infected. For this simulation we wanted to examine the effects of social response when it led to recovery, but did not fully protect the user.

We use the results from these individual models, as well as larger data on websites hosting phishing sites to model system-wide properties of malware spread. We use these system-wide properties to draw analogies from public health research regarding the spread of sexually transmitted infections (STIs) to examine organizational patching policies. From these results, we argue that thinking of security problems in terms of public health policy is a good addition to more traditional mental models of security.

8.2 Background and Related Work

In early work adapting epidemiological models to computer viruses, the local nature of data transfer had to be taken into account. Computer viruses, in general, were spread very locally, and certain assumptions such as homogeneous population and the probability that an infected individual could infect any other individual in the susceptible population did not hold [18]. In this environment Kephart and White (KW) adapted the Susceptible-Infected-Susceptible (SIS) model to account for the non-homogeneous behavior of program sharing.

In their model, each computer is a vertex in a graph and an arc connects another computer in a program-exchange relationship. The arcs are associated with individual rates of infection and represent the set of vertices that can be infected by a given vertex, while each vertex is given an individual rate of recovery. Once a vertex has recovered, it is immediately capable of being reinfected. This, as the authors state, represents a very simple assumption that users will not become more vigilant after being infected. While this is a simple assumption, it seems to be a fairly good approximation for real world data [26].

Their deterministic calculations correspond to early results in prevalence driven epidemiological models [20], but failed to capture the social or organizational aspects of dealing with virus spreads. They modified their model to include a social response, or, as they call it, a kill switch model. That is, each computer, upon discovery and cleaning, alerts all other computers it is connected to alert them of possible infection [19].

This extension (Fig. 8.1) assumes that recovery corresponds to a temporary immunization from the virus [19]. Based on these model extensions, KW showed that central reporting and response to an incident is important to containing the incident. With central reporting and response, even if an organization is above the epidemic threshold, an incident can be limited in size and duration [19].

Other early work in modeling computer viruses overcame the limitations of the well-mixed assumption by incorporating specific characteristics of individual malware, such as scanning behavior or interesting topological structures. Knight, Elder, and Wang analyzed networks in hierarchical and cluster topologies to study the effects of immunization from viruses in theoretical email networks [41].

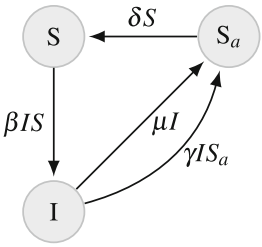

Fig. 8.1 SIS model with recovery and social response. β represents the effect contact rate, μ represents recovery rate, and γ represents social response rate. δ represents the return from risk-averse to susceptible population

Newman, Forrest, and Balthrop expanded Knight et al.'s work by incorporating actual email network data and studies of network structure from the realm of statistical physics [30].

Both Knight et al. and Newman et al. demonstrated that targeted immunization could have a drastic effect on the spreading of viruses spread by emails by drawing heavily on studies in graph theory and network science found in Albert, Jeong, and Barabási's work on describing the network topology of the Internet [7, 42] and the effects of that topology on disease spread found in Pastor-Satorras and Vespignani's work [31, 32].

Zou et al.'s work on modeling the Code-Red worm using the description and data provided by Moore et al. modified the standard SIS model by incorporating a variation of Kephart and White's social response model, incorporating scanning rate, and allowing for infection rates to fluctuate in time [27,44]. Including the social response in their model allowed them to take into account human responses to the onset of an infection [44].

Zou, Gong, and Towsley also included a model that allowed systems to become quarantined, removing them from the susceptible and infectious populations [45]. They demonstrated that removing computers from both populations for some amount time was an effective mitigating factor [45]. However, as Serazzi and Zanero pointed out in their later work on Sapphire, quarantines would be difficult to implement, as infected hosts cannot be trusted to quarantine themselves [36]. Zou and Towsley revisited their earlier work to demonstrate that the increased range of addresses in IPv6 would effectively reduce the total prevalence of routing worms such as Sapphire. They showed this reduction is due to scanning worms' inability to access significant parts of the IPv6 address space in a reasonable amount of time [43].

Moore et al.'s data collection and description of the explosive growth of the Sapphire worm required further modifications to earlier models [28]. While Code-Red generally followed standard models, Sapphire spread fast enough to become bandwidth-limited, which, in turn, limited its total ability to spread [28]. Serazzi and Zanero designed a model that encoded network resources. Utilizing incoming and outgoing traffic rates into their model, they were able to capture Sapphire's aggressive scanning. This scanning choked the Internet and greatly impeded Sapphire's rate of growth [36]. Serazzi and Zanero also pointed out the difficulty in implementing global security policies such as quarantines and hub immunizations.

Staniford, Paxson, and Weaver contributed an excellent summary of many of the modeling attempts and call for a CDC for computer malware [39]. We agree with this model of thinking, and the data collected via their suggested sensors and analysis would be useful for further mitigation of online pathogens. However, this chapter focuses more substantially upon the effects of risk takers on the total prevalence of contagion. Thus, we hope to show that a small group of users engaged in risky behavior creates a threat to the risk-averse population.

To this end we look primarily at August and Tunca's work on allowing users with illegal copies of software to patch [4] and Choi, Fershtman, and Gandal's work on cost of patching [13]. While August and Tunca focus primarily on whether or not firms should allow users of illegal copies to patch, Choi, Fershtman, and Gandal look at the costs associated with different users and their willingness to patch. We combine both the pirates in August and Tunca's work with the non-patching populations of Choi, Fershtman, and Gandal to show that limitations on user ability to maintain a secure system is dangerous to the risk-averse population.

Models of sexually transmitted diseases have become very complicated to deal with multiple population interactions [11, 17]. However, most multiple population models do not couple the behavioral changes that occur to individuals' perceptions of disease spread [33]. We build on Perra et al.'s work to create a two-population model with a social response that represents the ability of users to change behavior, and, thus, their population group. This differentiates our model from more complicated models of STIs that use different characteristics of infection for individual population groups, but do not include behavioral responses to infection [5, 10, 34, 37].

8.3 Methodology

We first develop a simple model based on Kephart and White's initial social response model and Wang et al.'s user vigilance model. We then use our model to examine the long-term global prevalence of malware. Then we analyze the various parameters within this model to identify which parameters are most effective at controlling systemic infection. We also attempt to answer questions about feasible responses to malware diffusion that could result in reduction of botnet prevalence.

8.3.1 Model Creation

Kephart and White's social response model (KW) demonstrates the effectiveness of social responses to computer infection. We extend their model to allow the possibility of infection in the inoculated population. This extension includes aspects of Wang et al.'s vigilance model [40]. Similar to their approach, we view user

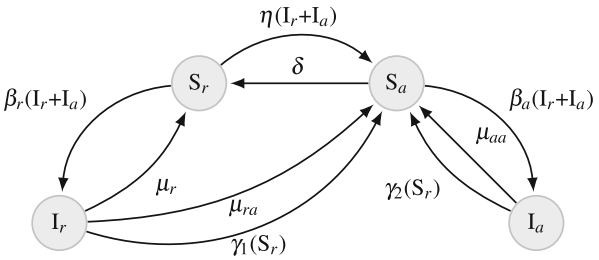

Fig. 8.2 Allowable transitions in our two-population SIS model with recovery and social response

vigilance as a prevalence-based response to the infectious population, with vigilant users returning to the more susceptible population at a constant rate.

Since we are modeling the global diffusion of all malware, individual behaviors are of limited use. We are more interested in equilibrium states. Thus, the more common Susceptible-Infectious-Recovered (SIR) type models previously used for models of specific malware infection are not useful for our purposes. Moreover, we find that not assuming total recovery better captures the long-term behavior seen in malware such as the Blaster worm [6] as well as the persistent insecurities found in Web servers that allow them to be reinfected [26].

$$\begin{aligned}
\frac{\mathrm{d}S_r}{dt} &= -\beta_r(I_r + I_a)S_r - \eta(I_r + I_a)S_r + \delta S_a + \mu_{rr} I_r \\
\frac{\mathrm{d}I_r}{dt} &= \beta_r(I_r + I_a)S_r - \mu_{rr} I_r - \mu_{ra} I_r - \gamma_{ra} I_r S_a \\
\frac{\mathrm{d}S_a}{dt} &= -\beta_a(I_r + I_a)S_a - \delta S_a + \mu_{ra} I_r + \mu_{aa} I_a + \\
&\quad \gamma_{ra} I_r S_a + \gamma_{aa} I_a S_a + \eta(I_r + I_a)S_r \\
\frac{\mathrm{d}I_a}{dt} &= \beta_a(I_r + I_a)S_a - \mu_{aa} I_a - \gamma_{aa} I_a S_a
\end{aligned} \tag{8.1}$$

The model we propose (Fig. 8.2 and Eq. 8.1) is a modified version of an SIS model with two interacting subpopulations. Our model does not assume immunity in the S_a population. This represents the fact that no security system is 100 % effective at stopping all vulnerabilities. We do not find, in the course of our analysis, that the rates of infection in the resistant population are so low that they may be ignored.

Our model also assumes a well-mixed, homogeneous population. This is, in many ways, an unrealistic assumption, given the patterns of connection displayed by social networks and browsing behavior [25]. Moreover, it distracts from our metaphor of STIs, in that it assumes that all users are equally likely to interact with one another, rather than rely on contact patterns [15]. However, the dissemination of many online attacks is based on random scanning, which creates a scaled version of a well-mixed, homogeneous population [18]. Thus, this is a useful simplifying assumption, but it can be expanded upon in future work.

Table 8.1 Table defining included symbols

Notation	Definition
S_r	Susceptible non-vigilant population
S_a	Susceptible vigilant population
I_r	Infected non-vigilant population
I_a	Infected vigilant population
η	Non-vigilant response to infection
δ	Rate to return to non-vigilant population
β_r	Infection rate in non-vigilant population
β_a	Infection rate in vigilant population
μ_{rr}	Non-vigilant to non-vigilant recovery rate
μ_{ra}	Non-vigilant to vigilant recovery rate
μ_{aa}	Vigilant recovery rate
γ_{ra}	Non-vigilant to vigilant social response rate
γ_{aa}	Vigilant social response rate
R_∞	Equilibrium infected population
R_{∞_a}	Equilibrium infected vigilant population
R_{∞_r}	Equilibrium infected non-vigilant population

8.3.1.1 Parameter Definitions

Table 8.1 briefly summarizes the various symbols we use in our model and analysis, which we describe here. S_r represents the susceptible population of non-vigilant (risk taking) users. These are systems that do not have a form of malware and can be infected. S_a represents susceptible systems within vigilant (risk-averse) users. When an S_r or S_a system is infected, it transitions to the infected populations I_r or I_a, respectively.

η and δ govern the transitions between the two population groups. η represents the response of non-vigilant users to a given level of global infection. The higher η is, the faster non-vigilant users secure their systems. δ governs the response to the cost of maintaining a secure system. This is a constant rate, and the higher δ is, the less accepting users become of the cost, driving them to become insecure at a faster rate.

β_r, μ_{rr}, β_a, and μ_{aa} are the infection spread parameters for the non-vigilant and vigilant populations respectively. β_r and β_a govern how fast an infection spreads, while μ_{rr} and μ_{aa} dictate how quickly a user recovers. Recovery could be a simple as deleting an infected file, or as complex as reinstalling an OS. We assume that $\beta_r > \beta_a$ and $\mu_{rr} < \mu_{aa}$ to represent the fact that users that are maintaining a secure system will be less likely to become infected and more likely to recover.

γ_{ra} and γ_{aa} embed the response to social pressure to recover in the non-vigilant and vigilant populations, respectively. Users responding to these parameters, but not to μ_{rr} or μ_{aa}, do not scan their systems for potential threats, but respond when an entity they know alerts them to a possible threat. For example, a user may respond to a Firefox reminder to update their browser or the exhortation of a friend. A specific instance of this situation was Google's effort to alert users to possible infections

in 2011 [21]. This method of updating is less than ideal for maintaining a secure system, as, with limited occasions for reminders, infections can persist.

μ_{ra} defines the non-vigilant user's ability to clean or recover their system to a more secure state. This requires that non-vigilant users have access to the necessary patches and other up-to-date software to maintain a secure computer, at least until the cost of maintainence, δ, drives them back to the non-vigilant population.

8.3.2 Parameter Analysis

This model can be made equivalent to Kephart and White's kill switch model (Fig. 8.1) by setting $\delta = 0.01$, $\beta = 0.5$, $\mu_{ra} = 0.1$, and $\gamma_1 = 0.05$ and all other parameters to 0. We use both the Kephart and White (KW) model and a standard SIS model to compare our model under different parameter conditions. This allows us to evaluate which parameters may be realistic and useful.

8.3.2.1 Parameter Analysis in Risk-Averse Population Only

We first analyze the various effects of adjusting the parameters on the vigilant population to identify the most important parameters in controlling infection in that population. From there we move to analyzing the whole system, individually adjusting certain parameters to identify the key components of the system as a whole. For these simulations we vary one parameter and keep others constant. For each of the parameters we hold constant: β_a, μ_{aa}, and γ_{aa}, we set them to $0.5, 0.1, 0.01$ respectively.

These parameters are taken directly from KW and varied in later simulations. This sets a fixed social response at 1/10 the level of the recovery or cleaning response. This allows us to maintain consistency with our system-wide analysis below. We then vary the parameters of interest for each simulation from 0 to 1 by 0.01. Because we are only working with S_a, we initialize the populations to: $S_r = 0, S_a = 0.99, I = 0, I_a = 0.01$. Without an infected population I_r or I_a, no infections are possible in this model.

8.3.2.2 Parameter Analysis with Both Populations

The system parameter analysis keeps the infection rate and cleaning rate in the non-security-aware population at the same level as the standard SIS model used by KW ($\beta = 0.5$ and $\mu_{ra} = 0.1$). Fixing these parameters reduces the number of variables we must examine and provides us with a reasonable worst-case scenario of 80 % of non-vigilant computers infected. However, we adjust the security-aware population to reflect a greater vigilance.

We set the infection rate of S_a to half of the non-vigilant population's rate. Similarly, the cleaning rate of S_a is twice that found in the non-vigilant populations. In the vigilant population, there is a social response, but this is 1/10th the cleaning rate. This leads to the following parameter values: ($\beta_a = 0.25, \mu_{aa} = 0.2$, and $\gamma_{aa} = 0.02$). We normalize the initial populations to $S_r = 0.99, S_a = 0, I = 0.01, I_a = 0$.

Additionally, these parameter values are a reasonable estimation of actual global prevalence. Our initial parameter values in isolated populations lead to roughly 80 % of the population falling into the non-vigilant population, and roughly 80 % of that population infected. Within the vigilant population, the initial parameter values lead to roughly 13 % of that population infected. With no interactions between the populations, this leads to a global prevalence of roughly 77 %. These results correspond to the estimates of global prevalence below.

In their report to the House of Lords in 2007, the Science and Technology Committee reported on results from an earlier study that showed that roughly 80 % computers lacked neccesary security measures, and roughly 72 % of sampled systems had some type of malware [35]. However, the committee noted that this study only sampled 354 computers, so it probably was not an accurate portrayal of the actual prevalence of malware. For example, for 2010 and the 1st half of 2011, The Anti-phishing Working Group (APWG) found that an average of 48 % (sd = 6.53) of their observed computers were infected with some sort of malware [1–3]. Thus, our initial parameter values align very closely with the earlier study, and represent approximately a 60 % increase over the APWG's results.

8.3.2.3 Sensitivity Analysis

The first two sets of analysis represent a very crude sensitivity analysis given the number of parameters. We analyze each parameter in light of a fixed system. This analysis reduces the problem from a nine-dimensional problem to a one-dimensional one, but it is not informative in terms of how the parameters interact with one another. To address this, we performed sensitivity analysis under two different sets of conditions.

We performed a sensitivity analysis by using Latin Hypercube Sampling (LHS) on the set of all parameters against the measured value of total infectious computers [9]. LHS first samples from prior distributions of parameter values and generates sampled output for the number of samples. In our case, we used 1,000 samples. From there LHS uses a rank-transform correlation coefficient to measure the sensitivity of each parameter as it pertains to the measured output [9].

We used uniform priors for all parameter values, given our own uncertainty of acceptable distributions. Our initial test was performed over all parameters and used to identify the key bifurcation parameters [24]. These bifurcation parameters are key to differentiating the major equilibrium behavior in the system; mainly, whether a contagion is maintained or dies out. After we identified the key bifurcation parame-

ters, we set them to ensure continued prevalence and performed the analysis again. This allowed us to identify key parameters associated with reducing prevalence.

8.4 Results

We examined the effects of parameter variation in different phases. We first wanted to see if there was a way to reduce total infection prevalence by adjusting only the parameters associated with the vigilant population. After considering only the vigilant population, we investigated the effects of making the non-vigilant population respondent to the vigilant population.

We conducted this investigation by adjusting the μ_{ra} and γ_{ra} parameters to investigate the effect of a user's ability to recover to more secure behavior. We then adjusted the parameters that determined the speed of transition to and from the vigilant population (η and δ) in uninfected users. When the infection is less potent in vigilant users, we can reduce the total infected population by having more users become vigilant and having vigilant users stay vigilant longer. For these simulations we do not adjust the infection or clean rates, but keep the non-vigilant and vigilant population parameters at their fixed rates discussed above.

8.4.1 Effects of Adjusting Parameters in Risk-Averse Population Only

In KW's model of social response, they add a prevalence-driven recovery effect on top of the standard, constant-rate recovery. In order to investigate the effects of this recovery in the population we varied the social response in the vigilant population only, to see if it would lead to significant reduction in the equilibrium of total infected. Since we split the model into two subpopulations, we could also examine the source of the infections.

8.4.1.1 Simulations 1–3

In Kephart and White's examination of the social response, even a small social response was able to reduce the total level of infection in the system. However, this relied on a system where the recovered population could not become infected. For these simulations we wanted to examine the effects of social response when it led to recovery, but the recovery did not protect the user from reinfection. We set the infection characteristics in the vigilant population to correspond to KW's model ($\beta_a = 0.5$ and $\mu_2 = 0.1$) and varied γ_{aa}.

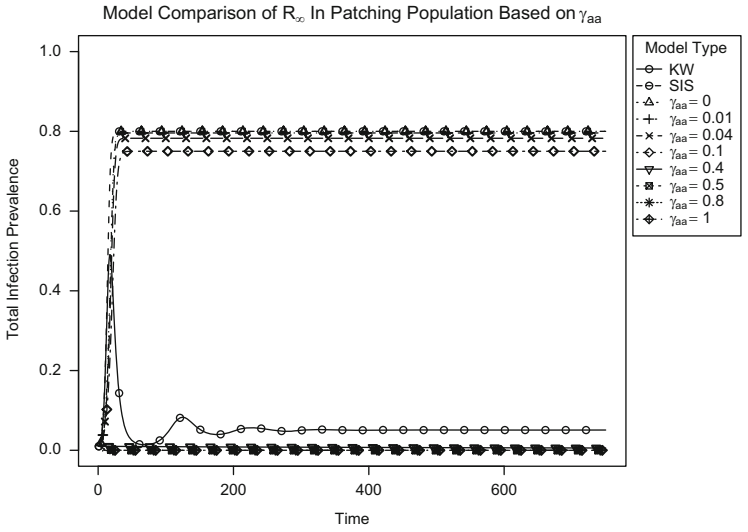

Fig. 8.3 Comparison between SIS, KW, and our model, γ_{aa} variable. Only when γ_{aa} reaches high rates relative to β_a does R_{∞_a} fall

Looking at the results (Fig. 8.3) we find that only when the combination of social response and cleaning rate is greater than the infection rate does the infection die off. That is, $\mu_{aa}/\beta_a - \gamma_{aa} = 1$ is the bifurcation point in this single population. When $\mu_{aa}/\beta_a - \gamma_{aa} < 1$ then $R_\infty = 1 - \mu_{aa}/\beta_a - \gamma_{aa}$, and when $\mu_{aa}/\beta_a - \gamma_{aa} > 1$, the infection disappears. Moreover, with this mathematical analysis of the single population, we quickly identify the effects of each parameter. β_a and μ_{aa} have a multiplicative effect on R_∞, whereas γ_{aa} has only an additive effect.

These results mean that the total social response and the cleaning rate must affect the network at the same rate as the malware to be effective at eliminating its spread. For a single population then, social response is a useful measure in reducing the total infection rate, but is unlikely to be able to reduce the infection from a pandemic unless it is unreasonably high.

8.4.2 Effects of Adjusting a Single Parameter in Both Populations

For this set of analyses, the populations become coupled in multiple ways making analytical analysis difficult. Thus, we examine the effects of parameters by adjusting a single parameter in both vigilant and non-vigilant populations and investigate how it affects the R_∞ for the entire population. The first three simulations examine the effects of behavior changes in the vigilant population. The final simulations study

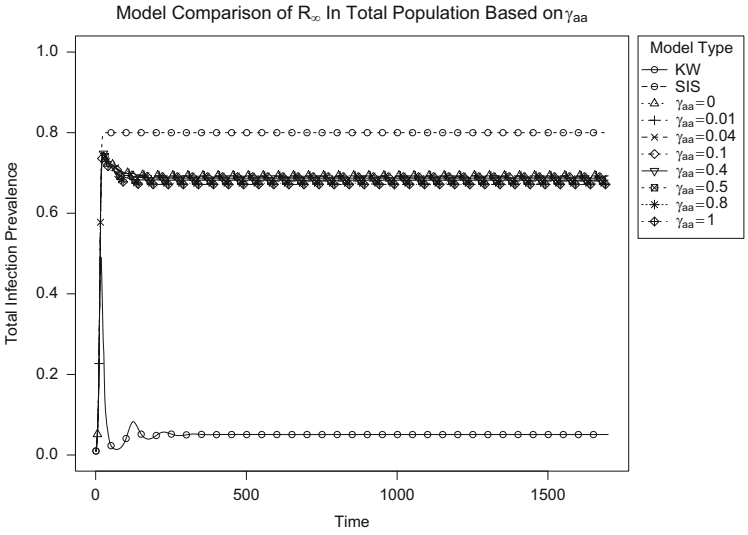

Fig. 8.4 Comparison between SIS, KW, and our model, γ_{aa} variable. Increasing the social response rate does reduce R_∞, but even at extreme values, γ_{aa} is unable to significantly reduce R_∞ due to the effects of the large infected non-vigilant population

the interactions between the two uninfected populations governed by η and δ. We find that the principal way to reduce R_∞ is to allow infected individuals to recover to the vigilant population.

8.4.2.1 Simulation 4

In this simulation we adjust the social response parameter in the vigilant population. This allows us to see what effect increasing the parameters in the vigilant population has on the system-wide R_∞ (Fig. 8.4). We notice, as in the following two simulations, increasing the responses in the vigilant population does little to reduce the total R_∞.

The dynamic relationship between γ_{aa} and R_∞ is a bit more complicated in this simulation, as this simulation contains an I_r value that is non-zero and transitions between the populations. We are holding $\mu_{ra} = 0$ and $\gamma_1 = 0$, so we know that the relationship between $\eta = 0.05$ and $\delta = 0.04$ gives us an approximate 80–20 split between security-conscious users and those that are unable or unwilling to engage in more secure behaviors. We also know that when $\gamma_{aa} + \mu_{aa} > \beta_a$, $R_{\infty_a} = 0$ in an isolated situation.

However, even when $\gamma_{aa} = 1$, we still end up with an infected vigilant population. In this case, $R_{\infty_a} \approx 0.072$, while $R_{\infty_r} \approx 0.6$. $R_{\infty_a} = 0.072$ represents approximately 31 % of the vigilant population, while $R_{\infty_r} \approx 0.6$ is approximately 77 % of the non-vigilant population. Recall Fig. 8.3 that illustrated that with only

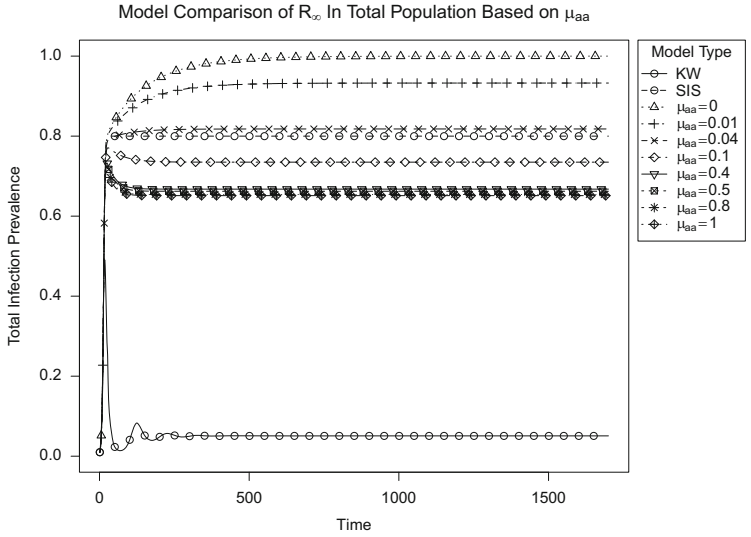

Fig. 8.5 Comparison between SIS, KW, and our model, μ_{aa} variable. Reductions in μ_{aa} lead to increases in R_∞, as the global prevalence tends towards the behavior of the least secure population. Increases in μ_{aa} are capable in reducing R_∞, but the reductions in R_∞ are mitigated by the behavior in the non-vigilant population

a vigilant population $\gamma_{aa} = 1$ should remove all contagion within the vigilant population. Thus, the infections within the vigilant population are being driven by the non-vigilant population.

8.4.2.2 Simulation 5

In this simulation we adjusted the cleaning rate within the vigilant population. The key result here is if $R_{\infty_a} > R_{\infty_r}$, R_{∞_a} drives the total R_∞ (Fig. 8.5). However, this is unlikely, as it is improbable that vigilant users will become infected at a greater rate than non-vigilant users. While, if $R_{\infty_a} < R_{\infty_r}$, but $\gamma_{aa} + \mu_{aa} < \beta_a$, the infection is driven by both vigilant and non-vigilant populations. In the case where $R_{\infty_a} < R_{\infty_r}$, and $\gamma_{aa} + \mu_{aa} > \beta_a$, the infections in the security-aware population are due to the prevalence of infectious non-vigilant systems.

For example, when $\mu_{aa} = 0$, there is no cleaning and the social response cannot reduce the spread of the infection within the vigilant population. Thus, $R_{\infty_a} = 1$. At the end of the 1,700 time steps in our simulation, $R_\infty = 1$, with most of it (99.998 %) being made up of the "vigilant" population. This suggests that the vigilant population cannot rely merely on protective measures to avoid infection, but must also be diligent in actively monitoring and maintaining their systems.

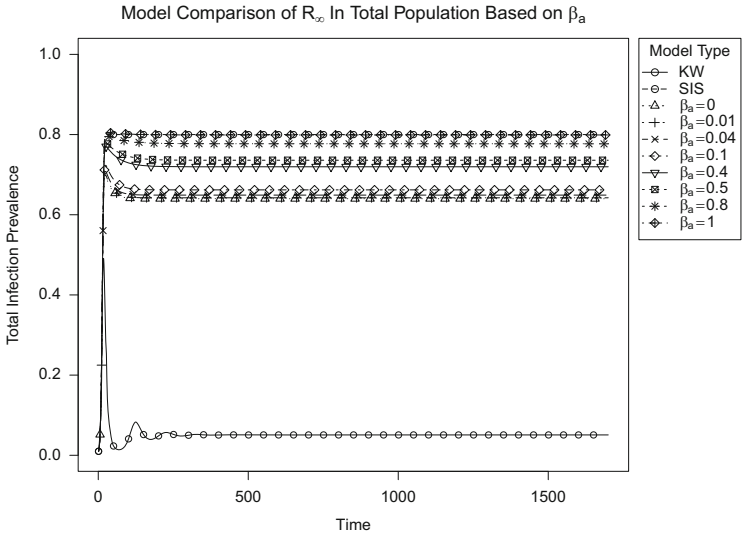

Fig. 8.6 Comparison between SIS, KW, and our model, β_a variable. Increasing β_a increases R_∞, but due to the values of γ_{aa} and μ_{aa}, R_∞ performs as R_{∞_r}. If γ_{aa} or μ_{aa} are reduced, R_∞ would also increase, as the vigilant population would be the least secure population

8.4.2.3 Simulation 6

This simulation adjusted the β_a parameter to investigate how allowing the vigilant population to reduce, or increase its infection rate would affect the system-wide R_∞. Given the parameter values for μ_{aa}, as β_a increases to 1, the vigilant populations dynamics approach those of the non-vigilant population. Hence the convergence to $R_\infty = 0.8$, as β_a goes to 1 (Fig. 8.6).

This simulation suggests that if our recovery and social response parameters were such that the reproduction rate of the vigilant population were greater than the non-security aware population, it would pull the system to a total R_{∞_a}, as users would flee the infectious environment of the non-vigilant group, to the even more infectious vigilant group.

8.4.2.4 Simulation 7

For this simulation, we varied η to see how increases in the response rate of non-security users in the face of infection impacted R_∞. Recall that η represents a user's ability to transition from non-vigilant to vigilant in the face of an impending infection in order to reduce the likelihood of infection. Obviously, when $\eta = 0$, there is no transition to the security-aware population, and the model behaves as a standard SIS model as shown in Fig. 8.7.

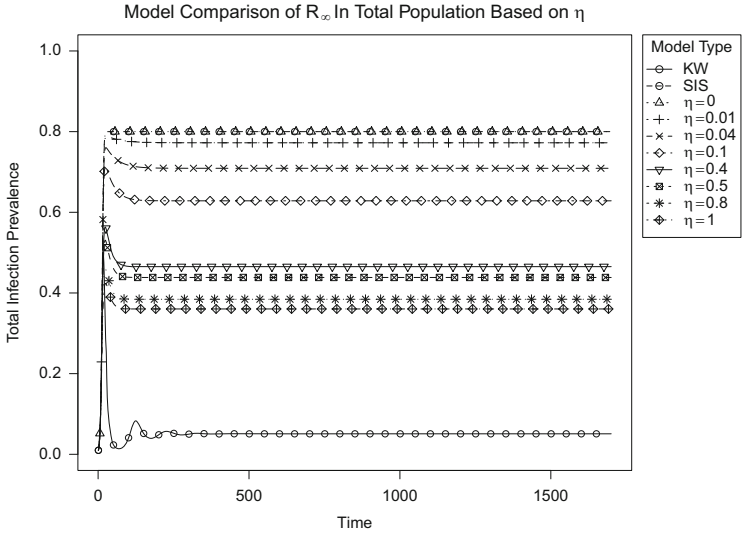

Fig. 8.7 Comparison between SIS, KW, and our model, η variable. Increases in the ability for the susceptible non-vigilant population to become vigilant reduces R_∞

However, when $\eta > 0$, the model behaves in an interesting manner, where it is possible to see that increasing η reduces the system-wide R_∞ (Fig. 8.7), but at the same time highlighting a complex relationship between η and R_{∞_a}. As seen in Table 8.2 and Fig. 8.7, while the total R_∞ is decreasing, the R_{∞_a} increases until $0.6 < \eta < 0.7$, when it begins to decrease. It is also possible to see that R_{∞_a}, while increasing in those intervals, is always decreasing as a percentage of the vigilant population.

The transition speed η pulls more of the total population into the vigilant population, but, until it is able to overcome the increasingly small non-vigilant population, that population still exerts a growing cost on the vigilant population. This result is important, since it indicates that even a small population engaged in risk behavior, with limited opportunity to reduce their risk, threatens a larger, risk-averse population.

When $\eta \gg \delta$, it is unable to pull R_∞ to R_{∞_a} in the isolated system case. Yet even an η as low as 0.1, is capable of reducing R_∞ more than any of the test values of β_a, μ_{aa}, or γ_{aa}. This suggests that if modifying η is feasible, it would have a significant impact on global malware presence.

8.4.2.5 Simulation 8

In this simulation we varied the other part of the transitions from non-vigilant to vigilant. δ represents the constant rate of relapse where users view the costs

Table 8.2 Interaction between η and R_{∞_a}

η	% Population$_a$	% Population$_a$ infected	R_{∞_a}
0	0	–	0
0.1	40.1	43.4	0.174
0.2	54.9	39.9	0.219
0.3	63.1	37.5	0.237
0.4	68.5	35.7	0.245
0.5	72.3	34.3	0.248
0.6	75.2	33.1	0.249
0.7	77.4	32.1	0.249
0.8	79.2	31.3	0.248
0.9	80.8	30.5	0.246

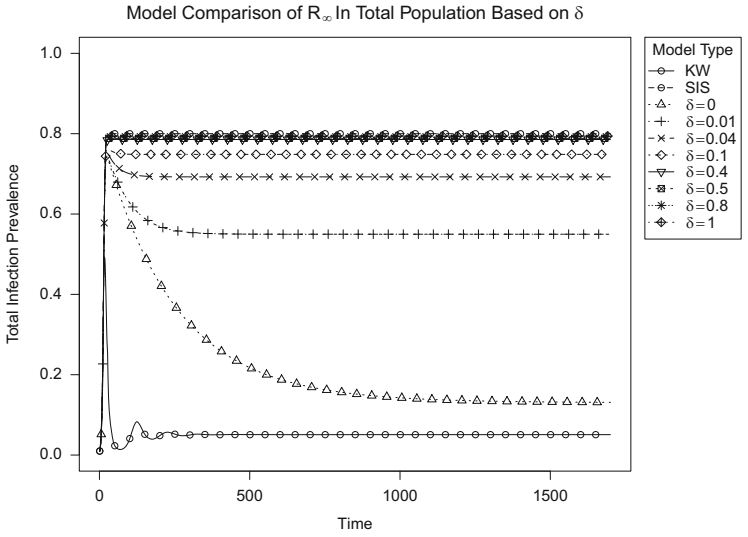

Fig. 8.8 Comparison between SIS, KW, and our model, δ variable. Reducing users' willingness to become non-vigilant reduces R_∞

of maintaining security as impractical or too expensive. Reducing δ represents increasing a users willingness to engage in more secure behavior, while increasing δ represents users that are only willing to be vigilant in the face of large outbreaks.

What becomes immediately apparent is that when $\delta = 0$, R_∞ approaches R_{∞_a} (Fig. 8.8). However, $\delta = 0$, while ideal, is unlikely. It represents a population that is fully vigilant, irrespective of cost. We can see that reducing δ from 0.04 to 0.01, results in $R_\infty \approx 0.549$, which is lower than the R_∞ achievable by extreme values in β_a, μ_{aa}, or γ_{aa}. It is unlikely that such lack of sensitivity is realistically achievable, though it is probably reasonable to assume that η and δ are of the same order of magnitude.

Reducing the costs of risk-averse behavior requires many aspects outside of mere security and patching. For example, while use of automated patching may be effective in reducing the cost of maintaining an up-to-date system, a user's behavior on the Internet is also important. For example, sites that offer illegal copies of software or music are known vectors of malware [14]. In the instance of downloading illegal copies there are cultural and economic factors, such as the price of legitimate goods, which affect the level of participation and thus level of exposure of these vectors [29]. Thus, in order to reduce the cost of risk-averse behavior in terms of accessing black market digital goods, economic and social strategies must be used, rather than purely technical security solutions.

8.4.2.6 Simulation 9

In this simulation we investigate the ability of users to recover to a vigilant population through social response, rather than merely recovering to the standard susceptible population. γ_{ra} represents non-vigilant users' ability to respond to social pressure applied by non-infected vigilant users, not just to clean their machines, but to also, at least for some time, to become vigilant users.

In KW's social response model, γ_{ra} is kept to 1/10 of standard cleaning rate, but is effective at reducing R_∞ due to the lack of infection rate in the recovered population, and the inability to recover directly back to the susceptible population [19]. We concur with their assumption, in terms of limiting the social response rate. However, it is important to note how effective increases in γ_{ra} are at controlling width of the infection peak curve, and mitigating R_∞. Arguably, γ_{ra} should be limited in regards to β and μ, but it may be, that given certain network topologies, even a relatively low γ_{ra} will still be effective at reducing R_∞ (Fig. 8.9).

8.4.2.7 Simulation 10

In our final simulation, we vary μ_{ra}, the parameter representing cleaning a computer and adapting vigilant behavior. For example, a user reinstalling an OS and applying patches and installing AV software, rather than just removing malware and hoping to avoid infection in the future. When $\mu_{ra} = 0$, users are unable to become vigilant users until they clean their computers and respond to the infection through η. $\mu_{ra} > 0$ means that users have some method to recover directly to vigilant behavior (Fig. 8.10).

μ_{ra} is not as effective as γ_{ra} at limiting the duration of the infection peak, but it does limit the peak's height, limiting the total infections. Moreover, μ_{ra} is effective at low parameter values. When $\mu_{ra} = 0.05$, and no other parameters are changed, $R_\infty = 0.451$, a roughly 44 % reduction of R_∞ in the standard SIS model. This suggests that providing users with the ability to recover to updated and secured software/machines, should be a key component in any campaign to limit global prevalence of malware.

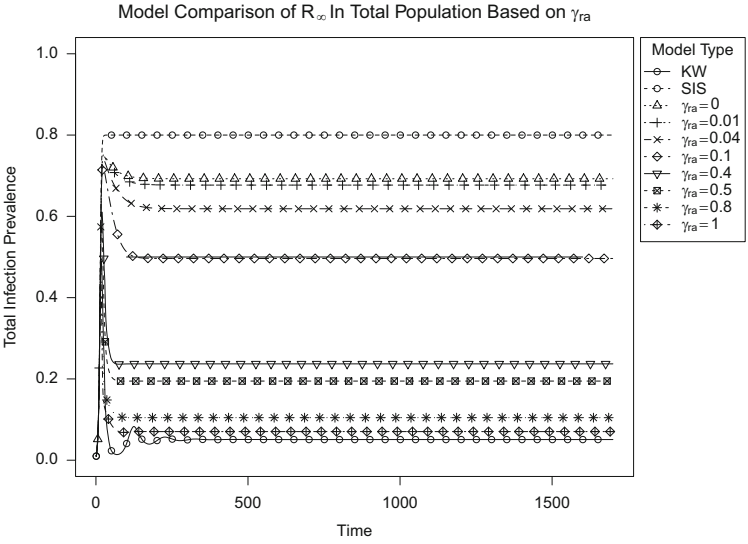

Fig. 8.9 Comparison between SIS, KW, and our model, γ_{ra} variable. Increases in users' ability to become vigilant in response to social pressure is effective at reducing R_∞

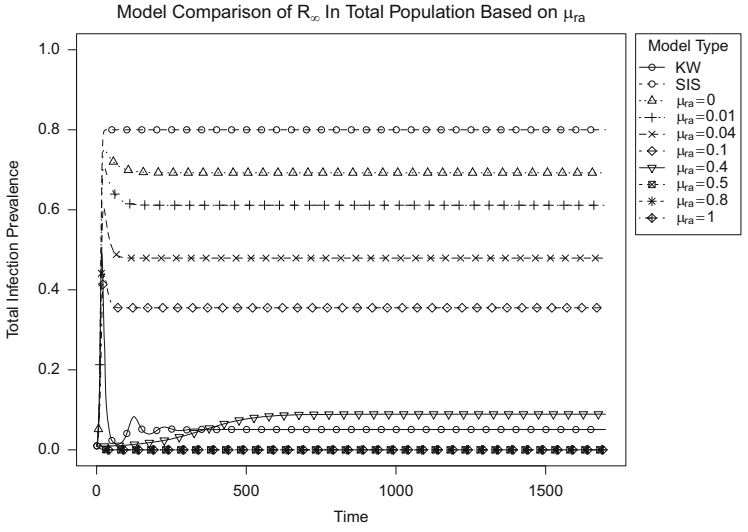

Fig. 8.10 Comparison between SIS, KW, and our model, μ_{ra} variable. Allowing users to recover their systems into the vigilant population is effective at reducing R_∞

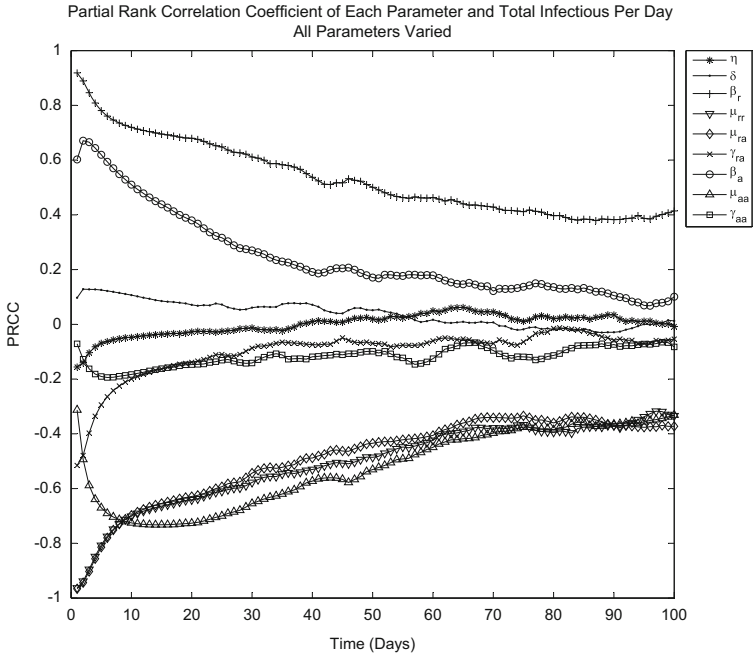

Fig. 8.11 Partial ranked correlation coefficients for all parameters calculated for total infections at time = t

8.4.3 Sensitivity Analysis

We used Latin Hypercube Sampling to examine both the epistemic uncertainty of the model, as well as the sensitivity of output variation to parameter variation [24]. The first step in LHS is to sample the parameter space to create a collection of measured outputs based on those samples. We did this sampling twice: first with all parameters sampled, followed by fixed values for the identified bifurcation parameters. In both cases we sampled the parameter space 1,000 times. Our output of interest was total infection prevalence.

Figure 8.11 shows the changing sensitivity of each parameter as time progresses. In the initial stages of the infection, social response and recovery from risk takers to the risk-averse population is more important than recovery within the risk-averse population. However, it rapidly loses its importance on overall prevalence, while risk-averse recovery increases its importance as time progresses.

All three of the standard recovery parameters (μ_x) are of approximately the same importance in the long-term reduction of prevalence. However, the infection rate in the risk averse group (β_a) loses its sensitivity gradually. The transmissions between susceptible risk takers and susceptible risk-averse (η and δ) are not significant in terms of affecting the global prevalence of a contagion, just as social response within the risk-averse community (γ_{aa}).

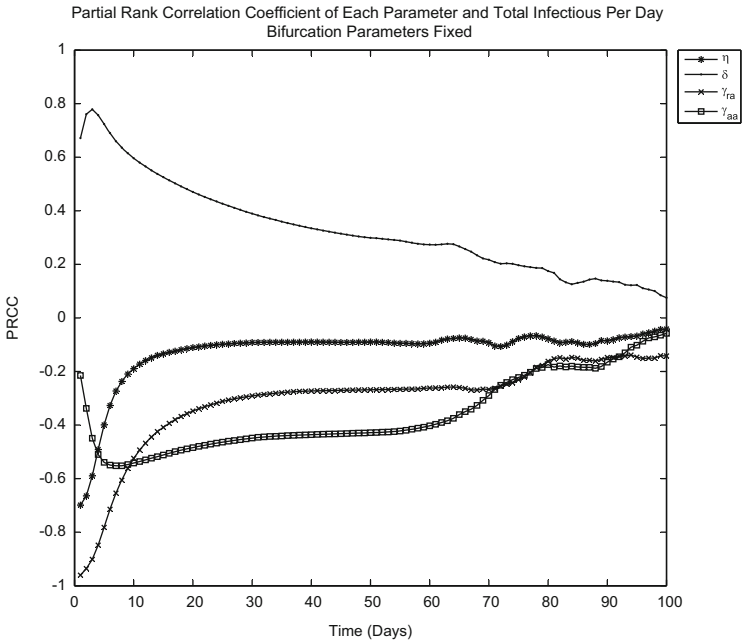

Fig. 8.12 Partial ranked correlation coefficients for non-bifurcation parameters calculated for total infections at time = t

However, when we fix the main bifurcation parameters ($\beta_r = 0.5$, $\beta_a = 0.25$, $\mu_{rr} = 0.1$, $\mu_{ra} = 0.01$, and $\mu_{aa} = 0.02$), we get a better view of the effects of the social parameters. When all parameters are varied, γ_{ra} is significant parameter for reducing prevalence in the initial stages of a contagion, while γ_{aa} is never significant. However, when a contagion exists, we find that both of the social response recovery rates are important at least until the later stages of a contagion, moreso than the transfer from susceptible risk takers to susceptible risk-averse (Fig. 8.12). However, allowing users to recover to the risk-averse population is still effective in the early stages (first 10 days in our simulations) of an outbreak. The effectiveness fades as the number of users able to become risk-averse is overtaken by the much faster infection rates.

8.5 Discussion

In Sect. 8.6 we reify the conclusions of the ten simulations described in this work. In this section we discuss the possible implications of our findings. That extreme changes in β_a have little effect in the equilibrium state of an infection is an encouraging result. The rate of spread of an infection is one variable completely

subject to the control of the attacker. Therefore great efficacy in changes in β would imply that defense could be ultimately futile.

Increasing the roughly equivalent variable, μ_a, is found to be as ineffective as β_a in decreasing the global prevalence of infection. However, there are significant caveats. The outcome assumes that the malware will remain endemic with a roughly constant β and that recovery does not result in immunity to a particular malware component. Yet given the existence of multiple malware attacks, the use of multiple vectors for a single malware variant, the lack of broad immunity upon recovery, and the potential for malware to evolve, these are not unreasonable assumptions.

Individuals choosing the recover due to social pressure (which includes automated pressure, such as Firefox exhortations to upgrade) must be faster than the rate at which the virus is spreading. This is an extremely unlikely case. Yet the social recovery rate, γ_{ra}, is one of the most effective measures in altering the equilibrium when there are two populations (vigilant and otherwise). However, increasing the response rate in the vigilant population has little effect on the global equilibrium. This is a mixed result given that it is arguably easier to alter a response rate in an aware population, but even modest gains in response of the unaware population can significantly reduce the global prevalence.

Transfer rates between the two populations is the most efficacious strategy for reducing long-term equilibrium. This argues that small increases in vigilance can result in significant increases in outcomes. Thus, increased use of healthy behaviors (e.g., contraception use or smoking cessation) can greatly reduce unintended consequences over the population as a whole. Compare this to situations where the entire population must engage in healthy behaviors (e.g., immunization) to result in significant outcomes. This argues for an approach that is closer to risk communication than mandates. Luckily, risk communication is feasible while global mandates are not.

Users must be able to act upon available information, e.g., δ should be quite low. The requires an ease of access to the resources necessary to engage in more secure behavior. Within the public health sector, barriers to treatment and preventive measures have been shown to greatly increase overall costs. For example, Franzini et al. estimated a likely additional cost of $43.6 million in a 1-year period in Texas if adolescents were required to notify parents when they received reproductive health care [16]. This suggests that allowing access to security patches, even in the case of illegal copies, would be effective in lowering system-wide costs, though offering those patches may not be profit-maximizing for a given firm [22].

Moreover, risk communication, when combined with access to treatment resources, has been effective in reducing prevalence in the public health sector. Spain et al. demonstrated the effectiveness of at-risk communication at recruiting at risk groups to utilize reproductive and preventative health care [38]. Several studies demonstrate the effectiveness of Youth Peer Education services at referring at-risk populations to appropriate clinics [12, 23]. When coupled with a voucher system for care, use of clinics increases dramatically [8]. Thus, there are extant systems of response and information that we can take advantage of in regards to encouraging more secure behavior.

The difference between the two populations are rate of recovery (μ), responsive to social pressure (γ), and decreased rate of infection (β). Therefore the findings above, of lack of efficacy of contact rate in the risk-averse population (β_a), social recovery rate (γ_{aa}), and recovery rate (μ_{aa}), are due primarily to the infectious interaction that the risk-averse population exerts as well as the interactions between these variables and the ability to become risk-averse before infection (δ) and the difficulty to remain risk-averse (η). In future work we will extend the model to include this feedback.

This model represents a theoretical model in the same vein as Kephart and White's initial model. However, there are ways to validate the model, but they require better data than we currently have available. A series of cohort studies using different Internet behavior and patch update strategies should be able to discover whether or not risk-averse behavior leads to fewer infections. Watching secondary infections from infected, monitored computers can reveal the effect of the infected population, risk-taking and risk-averse included. As future work, we plan to use a meta-analysis of available data to better estimate parameter distributions for a more accurate picture of malware prevalence in terms of R_∞.

8.6 Conclusions

In this chapter we created and examined the parameters of a two-population SIS epidemiological model in regards to global prevalence of malware. The two populations, vigilant and non-vigilant, interact in many different ways (Fig. 8.2), which affects R_∞, the equilibrium infected population. We examined single-parameter variations within the vigilant population and the system as a whole to identify key components to addressing the spread of malware.

In our first set of simulations, we examined the vigilant population in isolation, seeking to identify the most effective parameter for reducing or removing malware in that population. We found that within the single population it was possible to completely eliminate malware spread by setting $\mu_{aa} + \gamma_{aa} > \beta_a$. We also showed that adjusting the recovery rate μ_{aa} is the most effective way to reduce R_∞ in the vigilant population.

In our second set of simulations, we looked at the entire system and tried to find which parameters were effective at reducing global R_∞, while keeping the infection and recovery rates (β and μ) in the non-vigilant population constant. Here we find that, while we could eliminate the spread of infection within the vigilant population by overcoming the infection rate, adjusting the vigilant parameters had little effect on R_∞ and infections in the non-vigilant population drove the infections. However, when we examined the parameters governing transitions from non-vigilant to vigilant, we discovered several possibilities for infection control.

When we evaluated the transitions between uninfected non-vigilant and uninfected vigilant populations, we found that, while η was effective at making more users vigilant, even a small population of infected non-vigilant users could

negatively impact the vigilant population. Similarly, when we prevented users from returning to the non-vigilant population, we could limit the infection spread to R_{∞_a}. However, this represents an unrealistic expectation of inelasticity (i.e., all users demanding secure behavior, regardless of cost).

Examining the parameters governing the recovery of infected non-vigilant users to uninfected vigilant users, we find that allowing users to clean and repair their systems with updated and secure software is the most effective way to manage the global prevalence of malware infection. Even at low levels, the ability to recover to the risk-averse population (μ_{ra}) greatly reduces the global infection prevalence. Additionally, while not as effective as the risk-averse recovery rate (μ_{ra}), the social response recovery to the risk-averse population (γ_{ra}) is the next most effective parameter. This suggests that coupling social response, along with access to updates for all users, would be an effective measure for reducing the prevalence of global malware.

Acknowledgements We would like to thank Richard Clayton and Tyler Moore for access to their data. We would also like to thank Alessandro Vespignani for his suggestions on model building and analysis. We would further like to thank the Volkswagen Foundation for their support of our research by providing funds to travel to WEIS 2012.

References

1. Anti-Phishing Working Group: Phishing activity trends report 2nd quarter 2010. Tech. Rep. http://www.antiphishing.org/reports/apwg_report_q2_2010.pdf (June 2010)
2. Anti-Phishing Working Group: Phishing activity trends report 2nd half 2010. Tech. Rep. Anti-Phishing Working Group. http://www.antiphishing.org/reports/apwg_report_h2_2010.pdf (Dec 2010)
3. Anti-Phishing Working Group: Phishing activity trends report 1st half 2011. Tech. Rep. Anti-Phishing Working Group. http://www.antiphishing.org/reports/apwg_trends_report_h1_2011.pdf (June 2011)
4. August, T., Tunca, T.I.: Let the pirates patch? An economic analysis of software security patch restrictions. Inf. Syst. Res. **19**(1), 48–70 (2008)
5. Baggaley, R.F., Garnett, G.P., Ferguson, N.M.: Modelling the impact of antiretroviral use in resource-poor settings. PLoS Med. **3**(4), e124 (2006)
6. Bailey, M., Cooke, E., Jahanian, F., Watson, D., Nazario, J.: The Blaster worm: then and now. IEEE Secur. Privacy Mag. **3**(4), 26–31 (2005)
7. Barabasi, A., Albert, R., Jeong, H.: Scale-free characteristics of random networks: the topology of the World-Wide Web. Phys. A Stat. Mech. Appl. **281**(1–4), 69–77 (2000)
8. Bellows, N.M., Bellows, B.W., Warren, C.: Systematic review: the use of vouchers for reproductive health services in developing countries: systematic review. Trop. Med. Int. Health **16**(1), 84–96 (2011)
9. Blower, S., Dowlatabadi, H.: Sensitivity and uncertainty analysis of complex models of disease transmission: an HIV model, as an example. Int. Stat. Rev. **62**(2), 229 (1994)
10. Boily, M.C., Bastos, F.I., Desai, K., Mâsse, B.: Changes in the transmission dynamics of the HIV epidemic after the wide-scale use of antiretroviral therapy could explain increases in sexually transmitted infections: results from mathematical models. Sex. Transm. Dis. **31**(2), 100–113 (2004)

11. Brown, T., Bao, L., Raftery, A.E., Salomon, J.A., Baggaley, R.F., Stover, J., Gerland, P.: Modelling HIV epidemics in the antiretroviral era: the UNAIDS estimation and projection package 2009. Sex. Transm. Infect. **86**(Suppl 2), ii3–ii10 (2010)
12. Burke, H.M., Pedersen, K.F., Williamson, N.E.: An assessment of cost, quality and outcomes for five HIV prevention youth peer education programs in Zambia. Health Edu. Res. **27**(2), 359–369 (2012)
13. Choi, J., Fershtman, C., Gandal, N.: Network Security: Vulnerabilities and Disclosure Policy. Centre for Economic Policy Research, London (2007)
14. Felt, A.P., Finifter, M., Chin, E., Hanna, S., Wagner, D.: A survey of mobile malware in the wild. In: Proceedings of the 1st ACM Workshop on Security and Privacy in Smartphones and Mobile Devices – SPSM'11, Chicago, p. 3. ACM, New York (2011)
15. Ferguson, N., Garnett, G.: More realistic models of sexually transmitted disease transmission dynamics: sexual partnership networks, pair models, and moment closure. Sex. Transm. Dis. **27**(10), 1–10 (2000)
16. Franzini, L., Marks, E., Cromwell, P.F., Risser, J., McGill, L., Markham, C., Selwyn, B., Shapiro, C.: Projected economic costs due to health consequences of teenagers' loss of confidentiality in obtaining reproductive health care services in Texas. Arch. Pediatr. Adolesc. Med. **158**(12), 1140–1146 (2004)
17. Gray, R.T., Beagley, K.W., Timms, P., Wilson, D.P.: Modeling the impact of potential vaccines on epidemics of sexually transmitted chlamydia trachomatis infection. J. Infect. Dis. **199**(11), 1680–1688 (2009)
18. Kephart, J., White, S.: Directed-graph epidemiological models of computer viruses. In: Proceedings: 1991 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, pp. 343–359 (1991)
19. Kephart, J., White, S.: Measuring and modeling computer virus prevalence. In: Proceedings: 1993 IEEE Computer Society Symposium on Research in Security and Privacy, Oakland, pp. 2–15 (1993)
20. Kermack, W.O., McKendrick, A.G.: A contribution to the mathematical theory of epidemics. Proc. R. Soc. Lond. A Contain. Papers Math. Phys. Character (1905–1934) **115**(772), 700–721 (1927)
21. Krebs, B.: Google: Your Computer Appears to Be Infected. http://krebsonsecurity.com/2011/07/google-your-computer-appears-to-be-infected/ (2011)
22. Lahiri, A.: Revisiting the Incentive to Tolerate Illegal Distribution of Software Products. In: 44th Hawaii International Conference on System Sciences, Koloa, Kauai (2011)
23. Liambila, W., Askew, I., Mwangi, J., Ayisi, R., Kibaru, J., Mullick, S.: Feasibility and effectiveness of integrating provider-initiated testing and counselling within family planning services in Kenya. AIDS **23**(Suppl 1), S115–S121 (2009)
24. Marino, S., Hogue, I.B., Ray, C.J., Kirschner, D.E.: A methodology for performing global uncertainty and sensitivity analysis in systems biology. J. Theor. Biol. **254**(1), 178–196 (2008)
25. Meiss, M.R., Menczer, F., Vespignani, A.: Structural analysis of behavioral networks from the internet. J. Phys. A Math. Theor. **41**(22), 224022 (2008)
26. Moore, T., Clayton, R.: Evil searching: compromise and recompromise of internet hosts for phishing. Financial Cryptography and Data Security, Barbados, pp. 256–272 (2009)
27. Moore, D., Shannon, C., Brown, J.: Code-Red: a case study on the spread and victims of an internet worm. In: Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement, Marseille, pp. 273–284. ACM (2002)
28. Moore, D., Paxson, V., Savage, S., Shannon, C., Staniford, S., Weaver, N.: Inside the Slammer worm. IEEE Secur. Priv. Mag. **1**(4), 33–39 (2003)
29. Moores, T.T.: An analysis of the impact of economic wealth and national culture on the rise and fall of software piracy rates. J. Bus. Ethics **81**(1), 39–51 (2007)
30. Newman, M., Forrest, S., Balthrop, J.: Email networks and the spread of computer viruses. Phys. Rev. E **66**(3), 035101-1–035101-4 (2002)
31. Pastor-Satorras, R., Vespignani, A.: Epidemic dynamics and endemic states in complex networks. Phys. Rev. E **63**(6), 1–8 (2001)

32. Pastor-Satorras, R., Vespignani, A.: Epidemic spreading in scale-free networks. Phys. Rev. Lett. **86**(14), 3200–3203 (2001)
33. Perra, N., Balcan, D., Gonçalves, B., Vespignani, A.: Towards a characterization of behavior-disease models. PloS One **6**(8), e23084 (2011)
34. Renton, A.M., Whitaker, L., Riddlesdell, M.: Heterosexual HIV transmission and STD prevalence: predictions of a theoretical model. Sex. Transm. Infect. **74**(5), 339–344 (1998)
35. Science and Technology Committee: Personal internet security. In: 5th Report of Session 2006–2007, vol. I, p. 121. House of Lords (2007)
36. Serazzi, G., Zanero, S.: Computer virus propagation models. In: Performance Tools and Applications to Networked Systems, pp. 26–50. Springer, Berlin (2004)
37. Shiboski, S., Padian, N.S.: Population- and individual-based approaches to the design and analysis of epidemiologic studies of sexually transmitted disease transmission. J. Infect. Dis. **174**(Suppl 2), S188–S200 (1996)
38. Spain, J.E., Peipert, J.F., Madden, T., Allsworth, J.E., Secura, G.M.: The contraceptive CHOICE project: recruiting women at highest risk for unintended pregnancy and sexually transmitted infection. J. Women's Health **19**(12), 2233–2238 (2010)
39. Staniford, S., Paxson, V., Weaver, N.: How to own the internet in your spare time. In: Proceedings of the 11th USENIX Security Symposium, San Francisco, vol. 8, pp. 149–167 (2002)
40. Wang, Y., Wang, C.: Modeling the effects of timing parameters on virus propagation. In: Proceedings of the 2003 ACM Workshop on Rapid Malcode – WORM'03, Washington, DC, p. 61. ACM, New York (2003)
41. Wang, C., Knight, J., Elder, M.: On computer viral infection and the effect of immunization. In: Proceedings 16th Annual Computer Security Applications Conference (ACSAC'00), New Orleans, pp. 246–256. IEEE Computer Society Press (2000)
42. Yook, S.H., Jeong, H., Barabasi, A.L.: Modeling the internet's large-scale topology. Proc. Natl. Acad. Sci. U S A **99**(21), 13382–13386 (2002)
43. Zou, C., Towsley, D.: Routing worm: a fast, selective attack worm based on IP address information. In: Workshop on Principles of Advanced and Distributed Simulation (PADS'05), Monterey, pp. 199–206. IEEE (2005)
44. Zou, C.C., Gong, W., Towsley, D.: Code Red worm propagation modeling and analysis. In: Proceedings of the 9th ACM Conference on Computer and Communications Security – CCS'02, Washington, DC, p. 138. ACM, New York (2002)
45. Zou, C.C., Gong, W., Towsley, D.: Worm propagation modeling and analysis under dynamic quarantine defense. In: Proceedings of the 2003 ACM Workshop on Rapid Malcode – WORM'03, Washington, DC, p. 51. ACM, New York (2003)

Part III
Economics of Privacy

Chapter 9
The Privacy Economics of Voluntary Over-disclosure in Web Forms

Sören Preibusch, Kat Krol, and Alastair R. Beresford

Abstract The Web form is the primary method of collecting personal data from individuals on the Web. Privacy concerns, time spent, and typing effort act as a major deterrent to completing Web forms. Yet consumers regularly provide more data than required. In a field experiment, we recruited 1,500 Web users to complete a form asking for ten items of identity and profile information of varying levels of sensitivity. We manipulated the number of mandatory fields (none vs. two) and the compensation for participation ($0.25 vs. $0.50) to quantify the extent of over-disclosure, the motives behind it, and the resulting costs and privacy invasion. We benchmarked the efficiency of compulsion and incentives in soliciting data against voluntary disclosure alone.

We observed a high prevalence of deliberate and unpaid over-disclosure of data. Participants regularly completed more form fields than required, or provided more details than requested. Through careful experimental design, we verified that participants understood that additional data disclosure was voluntary, and the information provided was considered sensitive. In our experiment, we found that making some fields mandatory jeopardised voluntary disclosure for the remaining optional fields. Conversely, monetary incentives for disclosing those same fields yielded positive spillover by increasing revelation ratios for other optional fields. We discuss the implications for commercial website operators, regulators, privacy-enhancing browser standards, and further experimental research in privacy economics.

S. Preibusch (✉) · A.R. Beresford
Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, UK
e-mail: sdp36@cl.cam.ac.uk; arb33@cam.ac.uk

K. Krol
Department of Computer Science and Security Science Doctoral Research Training Centre (SECReT), University College London, London WC1E 6BT, UK
e-mail: k.krol@cs.ucl.ac.uk

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_9, © Springer-Verlag Berlin Heidelberg 2013

9.1 Forms on the Web

Web users have been typing data into Web forms since they were added to the HTML standard [4] in 1995. Web forms allow interactive search and retrieval requests to servers, and to submit data to the Web server for further processing and storage. The HTML5 working draft proposes richer semantics for Web forms, including typed input fields which will only accept data of a specific kind or format, such as valid telephone numbers or email addresses. The new draft standard also recognises that Web forms may be used "for purposes other than submitting data to a server" [38]. Indeed, the use of form elements to collect data and process it locally inside the Web browser itself—typically using JavaScript—is already common on the Web today. In a few cases, data processed in this way remains within the browser; in most cases however, processed data is uploaded to a remote server at a later point in time.

Consequently, the Web form is the primary mechanism by which companies and governments collect personal data from individuals on the Internet today. Yet, relatively little is known about the behaviour and privacy attitudes of individuals when completing Web forms as part of completing a purchase or other commercial transaction on the Internet. Previous work has shown that consumers do show some reactance to data collection via forms: 25 % of Web users state that they have entered false data into forms [6], and the most frequently entered incorrect data item is their name [5].

Privacy aspects of Web forms have been studied more extensively in survey research. For example, research into *answer behaviour* to sensitive questions has shown that completion rates and data quality do not differ between on-line and paper forms [21]; however, open-ended questions have higher response rates when delivered on-line rather than via paper forms [9]. There remains disagreement on where to place sensitive questions in a questionnaire [36]. Both practice and academia variously advise placement at the beginning [1, 14], in the middle [13], or at the end [27].

Web surveys are typically implemented as forms. By analogy, research into user behaviour and recommendations concerning online questionnaires would also apply to transactional Web forms. However, Web forms and online surveys exhibit a number of systematic differences: first is the motivation for completing them. Whereas soliciting participation in surveys largely relies on social exchange theory, the Web form is often a means to an end (e.g., getting a Web order shipped, setting up an account), which should be motivation enough. Consequently, incentives such as money, sweets, or a lottery—common for completing questionnaires—are rarely found for transactional Web forms. Second, the visual appeal is different: the stand-alone Web form typically spans a single screen page and is as condensed as possible; a survey often continues over multiple pages and features elements such as instructions, and a progress indicator. A questionnaire is the Web page, whereas a standalone Web form is embedded into a Web page. Third, transactional Web forms often feature text input fields which prompt users for data through field labels; a

questionnaire asks questions, often closed. For many psychometric instruments, a battery of items measured on a Likert-scale results in the visual appeal of a matrix of tick boxes. Different activities are required from the user (ticking vs. typing; making a judgement vs. mentally looking up some data). The data items requested are also normally quite different, with the possible exception that surveys ask for contact details for a follow-up or a prize draw.

Completing Web forms is a time-consuming business, and therefore anything which can be done to ease the completion of a form has traditionally been considered sensible. Consequently, user experience practitioners and browser vendors have developed techniques to ease the completion of Web forms. This strand of research focuses on lowering the cognitive and mechanical effort of completing forms: label positioning (above the text field, on the left, below) and formatting (with or without trailing colon), the mechanism for indicating mandatory fields (it is rare [30], but commendable, namely with a red asterisk), and unified text field to reduce tabbing and mouse-keyboard switching [2]. Although some practitioners have strong opinions—for instance on label formatting—it does not seem to matter as long as the user experience is consistent [20].

Autocompletion of Web forms debuted in IE4 in 1997 and was initially called 'Form AutoFill' [23, 24]. The browser uses a combination of cues, including commonly used field names and names of previously completed fields, to match them across websites. The browser can then suggest values for form fields it has seen before, reducing the need to type the same information again. Today, autocompletion is limited to values typed into text fields. For one-time or sensitive entries, such as payment authorisation codes, Web form authors can prevent automatic form filling for an entire form (or parts of it), thereby mandating interactive completion of fields. Web users can configure their browsers not to store and suggest form values and delete individual values from their autocomplete suggestions.

Website authors could also attach semantics to form fields regardless of naming: each field may have a 'VCARD_NAME' attribute to draw information from the 'Profile Assistant', a local repository of identities [25]. For example, if a field is marked as 'vCard.Email', Internet Explorer suggests email addresses previously entered or stored in the Profile Assistant. Whilst the 29 names in the vCard schema cover all practically relevant contact details, the ECML 'Field Names for E-Commerce' defined in RFC 2706 and its successors expands on the idea of semantic annotation by introducing further field names for billing addresses and payment details [11]. Thirteen years later, in 2012, the Chrome browser was criticised for proposing yet another naming scheme to mark up semantically equivalent fields [8].

In the early 2000s, despite built-in browser support for the autocomplete feature, there was also demand for third-party tools, such as FormWhiz [29] and Gator eWallet. Gator eWallet called itself "the smart online companion" [15] and was marketed to "fill in FORMS with no typing". In addition to helping users complete forms, it also transmitted first name, zip code and country to the GAIN Publishing advertising network and served adverts back to the user [15] until it was shut down in July 2006 [16].

The emergence of the autocomplete feature led consumers to question whether their privacy might be violated. For example, PC Magazine in 1999 asked "What's to stop a hacker from stealing your personal data [that] is popping up in the AutoComplete window"? [31, p. 108]. Despite previously entered form data being stored locally in an encrypted file, researchers were able to read out the AutoComplete suggestions in all major browsers [37]. In summary, Web developers were encouraged to use autocomplete because it can "collect demographic data more easily" and faster [22], the common assumption being that the form itself is a nuisance or a "pain" [39, p. 19]. Yet, we are unable to find any thorough study or analysis which demonstrates the extent to which the use of autocomplete encourages data entry on Web forms.

Seemingly obvious assumptions regarding Web form completion have shown surprising results before. In the economics of privacy, it is regularly assumed that monetary or other incentives would encourage Web users to ignore their privacy concerns when filling out Web forms [34]. Yet, recent research has found that consumers show no preference for merchants with less privacy-invasive Web order forms even when all other parameters (such as product and price) are equal [3].

Contribution. In summary, the behaviour of individuals, and their motivations, when providing personal data via a Web form has not been studied rigorously before. Practitioners' literature and blogs abound with design guidelines for online forms on how to ease completion, but the advice is given without reference to any study or data. Evidence from survey research is related but not readily applicable. In this chapter, we deliver what we believe is the first experimental study into Web users' behaviour when providing personal information via a form. We quantify the amount of data provided, the costs of revealing it, and the motives for voluntary over-disclosure of data. Finally, we also benchmark the efficiency of incentivised data collection against voluntary and mandatory data disclosure.

Outline. We briefly revisit motives for voluntary data disclosure (Sect. 9.2) before outlining our research hypotheses (Sect. 9.3) and study methodology and design (Sect. 9.4). We give some descriptive statistics on the sample and their observed form-filling behaviour (Sect. 9.5) before turning to the analysis (Sect. 9.6). Managerial implications and pathways for regulation are discussed before concluding (Sect. 9.7).

9.2 Potential Motives for Over-disclosure of Personal Information

Based on the existing literature and common sense, we briefly review potential explanations for why Web users provide more information on forms than necessary. We focus on the initial act of over-disclosure and not subsequent failure to limit access to the information after it has been disclosed, for instance because of unusable privacy controls.

Over-disclosure by accident. The Web user may reveal more information than requested by accident or out of negligence. The user may ignore the optional status of some of the form fields, perhaps due to the lack of visual cues, deliberately misleading cues, or because she did not read the instructions carefully. This also includes the case of not reading the field labels and typing in more data or more detailed data than is strictly required.

Over-disclosure by proxy. A technical mechanism, such as the autocomplete feature described earlier, or a third-party filling out a form on someone else's behalf, may result in more completed form fields than the data subject intended.

Limit disclosure is costly. In analogy to 'limit pricing', we use the term 'limit disclosure' for disclosing on the edge of mandatoriness. This is difficult: the user may know that some form fields are optional, but is unable to identify them without incurring search cost. For example, if optional fields are not explicitly marked on the page, the user may need to submit the form multiple times with increasing amounts of personal data to determine which subset of the data is truly mandatory before the form is submitted successfully. A risk-averse user may be afraid of losing her entire form submission if she omitted a field. Reading instructions may also be viewed as prohibitively costly in terms of time or cognitive effort, and users may believe it is easier to complete all the fields.

Building social capital. Additional data may be provided in order to "look good" or otherwise stimulate a desired effect. It is a major driver for data disclosure on social networks while seeking mates [12] or job hunting. The analogy to the job market is particularly pertinent in our case, as our experiment is deployed on a crowd-sourcing platform. There could also be a social norm to over-disclose, the violation of which may hurt one's social capital.

Expecting a monetary return. The Web user may expect (to qualify for) a monetary return now or in the future, whether directly in the form of a discounted price, or by receiving promotional offers. To some extent, this explains voluntary data disclosure on online social lending platforms [7].

Expecting a non-monetary return. In the context of online shopping, disclosing (behavioural, preference and profile) information unlocks personalisation, this in turn makes it easier to find products of interest [28]. Further, in the context of online surveys, the respondent may anticipate that answering questions—despite them not being mandatory—will shape public opinion in a manner favourable to her.

Expecting infrastructure improvements. "Voluntary information spillovers" [17] promise economic profits if the information is picked up by companies to innovate products that meet a yet unsatisfied demand; this might be particularly pertinent, if exploiting the information oneself is infeasible or would yield inferior results. The theory of free revealing was originally developed for intellectual property, but we see it could apply to personal information, such as health information, as well.

Acting reciprocally. Reciprocity is a personality trait that facilitates voluntary disclosure of personal information in the context of social exchange. In surveys, social exchange is a strong driver towards participation [18], and incentives significantly increase response rates [19], until saturation is reached [32]. The user communicates gratitude, and returns a favour by revealing non-mandatory data items.

Acting benevolently or altruistically. Data might be provided out of kindness for the person, or organisation behind the Web form. In this scenario, the user fills out optional fields of the form even in the absence of personal benefit. There may also be the desire to help altruistically, and this leads to the assumption that filling out the form will help.

Personality. Some people like talking about themselves. The Web user's personality may be such that she enjoys disclosing information about herself. Filling out forms is subjectively rewarding.

9.3 Research Hypotheses

Our analysis is guided by eight research hypotheses which are designed to tease apart the motivations for over-disclosure as discussed in Sect. 9.2. Our hypotheses also quantify the conditions when over-disclosure will (not) occur. The eight hypotheses are:

H1: Web users provide more personal information than requested by a form, even though they realise there is no prospect of monetary reward for doing so.
H2: The base utility the Web user reaches by submitting the form does not determine the extent of over-disclosure.
H3: Over-disclosure of personal data is not an accident.
H4: Over-disclosure is costly to the user.
H5: Over-disclosure is not seen negatively.
H6: Users have good reasons to over-disclose personal information.
H7: Making some form fields mandatory reduces disclosure for the remaining optional fields.
H8: A reward for some form fields reduces disclosure for the remaining optional fields.

We motivate H2 as follows. For example, if a website pays users $2 for the form, they will not disclose any more optional information on the form than if they were paid $1. Analogously, volunteering information to a Web shop during checkout will be independent of the value of the product purchased. Experiments have also shown that the expected non-monetary benefits (e.g., personalisation) do not determine the extent of over-disclosure [33]. With regard to the exact differences in monetary incentives, we feature two different base rewards in our study ($0.25

vs. $0.50, Sect. 9.4.5); previous survey research found no effect on response rate between those two [19].

9.4 Experiment Methodology

9.4.1 Not a Survey

In privacy economics, surveys are known to yield results with low predictive value for real-world encounters (e.g., [33]). Laboratory and field experiments produce observations with better ecological (external) validity, although there may be trust biases from the 'secure' environment of a university laboratory. The single most important problem with survey-style methodologies is the lack of incentive compatibility when actions or preferences are stated rather than performed or expressed.

Although our design may look like a survey to the casual observer, we stress it is actually a field experiment: instead of asking whether respondents would reveal some personal information, we actually asked for those very data items. Participants did not state a willingness to disclose, but provided personal details at their discretion.

9.4.2 Form Design and Instructions

Figure 9.1 shows a screenshot of the form we asked participants to complete. We did not provide a cover story for data collection nor did we offer any explicit indication of the purpose for data collection. The form was headed "About yourself". The instructions written above the form, detailing which questions were mandatory, differed slightly depending on the treatment administered to the participant (Table 9.2). We deliberately ensured that all the data collection took place on a single page and ensured that it was clear that submitting the form also finished the task for the participant. The form itself did not mention the University of Cambridge, in words or pictures.

All information was collected using text fields. We did not use drop-down lists, radio buttons or tick boxes, even for questions soliciting a yes or no answer. All text fields had the same visual dimensions and we did not perform any input validation. Participants could enter data in any format they wished. For instance the field asking for date of birth did not require the participant to enter data in a specific way. Also partial input of day, month and year was possible. All field labels were phrased as questions, numbered, and were edited by a native English speaker. Due to the location of the participants, American English spelling (e.g., "favorite color") and currency (e.g., "$100") were used throughout. We took care to keep questions short,

Fig. 9.1 Web form used for the experiment; shown are the instructions used in treatment $T_{\underline{50}}$. 'HIT' denotes a task on the crowdsourcing platform we used. This form was embedded in a frame on the crowdsourcing platform, but no other elements were shown. The *right-hand side* gives the proportion of participants who completed each field in $T_{\underline{50}}$ (for all treatments, please see Fig. 9.2). The check questions 5 and 6 were completed by all participants by design

and make sure they were of comparable length. In particular, we made sure that the check questions (5 and 6, Fig. 9.1) did not stand out visually. There was no visual mark for those questions, such as an asterisk, to indicate they were mandatory.

9.4.3 Question Selection

We asked participants 12 questions as shown in Fig. 9.1. There was no randomisation in the order of the questions. Questions covered a variety of phrasings, including Wh-questions (When, Which, What. . .), inversions (Are you, Is it. . .), and with an auxiliary (Do you. . .).

The questions include identity-related and profile information, previously identified as sensitive personal information in another study [33]. We did not ask for data items that could directly identify an individual, such as email address or full name.

Some questions were worded to encourage a yes or no answer (e.g., 4 and 7) or a more elaborate response (e.g., 1 and 9). Every question could be answered with a single word or date. However, we deliberately gave respondents the opportunity to be more talkative: we expected at least some of the participants to elaborate on their yes/no answers. The questions selected fall in multiple, overlapping categories:

- *Identity/family*: Questions 1, 2, 4, and 11 ask for information typically found on an identity card (name, city, date of birth) or relate to the family (siblings and again name, date of birth).
- *Profile*: Questions 3, 7, 8, 9, 10, and 12 ask for profile information in a broad sense, including the user's current spatio-temporal context. Some of these questions require a judgement (3, 10, 12), others are factual (7, 8, 9). Orthogonally, some questions relate to the respondent's personality (namely, 3, 12), whilst others explore the technological context (9).
- *Check questions*: Questions 5 and 6 were the only questions that were always mandatory. These were included to check whether respondents had read and understood the instructions. Depending on the treatment, different answers were correct. Note that for the bonus check question, we intentionally asked about the participant's expectation ("Do you expect. . ." rather than "Is there. . ."): the belief in a payment determines behaviour.
- *Easily verifiable*: Answers to questions 2, 5, 6, and 9 are easily verifiable as they are factually right or wrong (5, 6) or can be checked by referring to meta-data (IP geo-location, HTTP request header).
- *Potentially verifiable*: Some answers could be verified by requesting the participant submit a photo of an ID card (1, 4, 11). This is not feasible however, as such verification is against the terms and conditions of the crowd-sourcing platform we used.
- *Non-verifiable*: Even with offline contact, some data items remain unverifiable (3, 7, 8, 10) by lack of omniscience.
- *Sensitive information*: Health information (10) is considered particularly sensitive (for example, it is listed amongst the "special categories of data" in the EU Data Protection directive); Web users themselves are regularly reluctant to share financial information (8), although there could be reputation gains from disclosing spending. Besides special data, date of birth (11) is one of the data items that consumers are least willing to provide online [3].

9.4.4 Sampling and Deployment

We used Amazon's Mechanical Turk (mTurk), a crowd-sourcing platform, to conduct our field experiment. On mTurk, requesters like us create and publish

tasks, called HITs (originally an acronym for 'Human Intelligence Tasks'). HITs are typically short and pay a few US cents. Workers (participants) choose the tasks they want to work on and submit their work, which is then accepted or rejected by the requester. Requesters can require qualifications for their tasks, such as location, experience or past performance of the worker.

Our choice of deployment had some implications for question selection: To comply with the terms and conditions of the mTurk platform, we did not ask for data items that could directly identify an individual, such as address details, full name, phone number or email address. This is particularly important as workers on the mTurk platform avoid tasks which contravene the terms and conditions. Workers fear they might not get paid for completing these tasks. We received no complaints for the data items requested in the final design.

The mTurk platform allows participants to preview the form before deciding to work on it. Consequently, we expected a few cases of non-response from participants. The ability to abandon the form after previewing it and not entering any data is important: it mirrors Web users' ability to navigate to an alternative Website if they are dissatisfied with the data collection practices of an operator [30].

On mTurk, the experiment was advertised as "Short survey—fast approval" with the description "Five-minute survey with fast approval". We decided the description should not reveal any more detail than the title of the task. By default, the advertisement included our requester name, "University of Cambridge". Amongst the mTurk worker population, "survey" is the term commonly used to describe all tasks that are published by research organisations. We decided that pretending not to be a research organisation would have been deception and harmed the internal validity of our study. Potential trust biases are discussed in Sect. 9.7.

By default, mTurk lists available tasks in the order in which they are advertised. Accordingly, our task moved down the list of available tasks as time progressed and therefore became less prominent. We chose not to re-advertise the task since this would allow a worker to participate in our experiment a second time. Consequently, we cannot use participation frequency as a sensible metric for unit non-response.

Before beginning any data collection, we obtained approval for our study from the Ethics Committee at the University of Cambridge Computer Laboratory.

9.4.5 Treatments

We piloted the form on 40 participants. No changes were necessary to the form itself, but we did learn that our initial estimate of the payment required to encourage participation ($0.65) was overly generous, and we decreased the payments for the main study to $0.50 and $0.25 depending on treatment type. Note that this payment acts as a show-up fee and is unaffected by actual data disclosure.

We varied treatments by task compensation and the amount of mandatory data. The low data requirement means that only the two check questions were mandatory. In the high data requirement, weather and favourite colour were mandatory answers

Table 9.1 Treatments with number of valid observations. 'p' indicates the pilot session

Data requirement		Compensation		
Minimum	Extra	$0.25	$0.50	$0.65
High	—	$T_{\overline{25}}$: 209	$T_{\overline{50}}$: 445	
Low	—	$T_{\underline{25}}$: 202	$T_{\underline{50}}$: 216	$T_{p\underline{65}}$: 38
Low	Bonus for high	$T_{B\underline{25}}$: 181		

Table 9.2 Instructions by treatment

Treatment	Instructions
All	Please provide some information about yourself
$T_{\overline{50}}$, $T_{\overline{25}}$	Questions 3, 5, 6 and 7 are mandatory. All other fields are optional. There is no bonus for this HIT
$T_{\underline{50}}$, $T_{\underline{25}}$, $T_{p\underline{65}}$	Questions 5 and 6 are mandatory. All other fields are optional. There is no bonus for this HIT
$T_{B\underline{25}}$	Questions 5 and 6 are mandatory. All other fields are optional. You will receive a $0.25 bonus when completing fields 3 and 7

in addition. A 2×2 full experimental design was used (Table 9.1). The instructions given on the form were amended to reflect the number of mandatory answers (Table 9.2). In a fifth treatment, $T_{B\underline{25}}$, we only mandated the two check questions, but awarded an extra payment ('bonus') of $0.25 to those participants who voluntarily answered the questions regarding weather and favourite colour. Each treatment was administered to 250 participants, with the exception of $T_{\overline{50}}$, which was administered to 500 participants.

The treatments were deployed as batches on Thursdays and Saturdays in January and February 2012. A two-tailed Mann-Whitney U test on the two batches forming treatment $T_{\overline{50}}$ indicates that the weekday of deployment does not affect response behaviour for sensitive data items (first name, date of birth: $p = 0.94$), or non-sensitive data items (weather, favourite colour: $p = 1.00$). We did not advertise or promote our experiment other than list it as a possible task or HIT on the mTurk platform. Consequently, participants self-selected to take part. We required workers to be based in the United States but placed no further restrictions on participation. Repeated participation was prevented.

9.4.6 Follow-up Questionnaire

We actively followed up with participants after they had submitted the form. At least 1 day later, they received an invitation to complete a feedback questionnaire for an additional payment of $0.65. Response rate on the follow-up was 74 %.

We reminded the participant of the original form they completed with a small screenshot, and then asked a series of 12 questions regarding their motives for

participating, time spent, enjoyment, and willingness to participate in a similar study; finally, we asked them whether they had revealed any personal or sensitive information, and, if so, which data items they considered as such.

Depending on their original submission, we also asked the participants for their motives for (not) telling us their date of birth. We asked for expected data use, and whether they had any objections against us sharing their data with an online shop. Reciprocity as a personality trait was measured with six-item battery of pre-established reliability [35, Question 126].

9.4.7 Data Processing and Coding

All responses were manually coded by a single skilled analyst prior to analysis. We excluded all participants without correct answers to both check questions. The proportion of correct answers was 72 % for $T_{B\underline{25}}$ and varied between 81 and 89 % for the other, non-incentivised treatments. Whilst a Kruskal-Wallis test indicates these differences across all treatments are significant ($p < 0.0001$), pairwise two-tailed Mann-Whitney U tests over all treatment combinations do not find systematic variations between the treatments. If it is true that the most diligent workers choose to undertake tasks more quickly than average workers, then this might explain the decline in submission ratios, since the tasks were scheduled sequentially (with $T_{B\underline{25}}$ administered last).

We checked the respondents' answers for plausibility. Because of the limitations associated with the mTurk platform, we could not fully verify their submissions. The percentage of obvious fake answers was very low, probably because response to most fields on the form was optional.

Taking the example of first name for illustration purposes, we have 1,110 correct submissions across all treatments except for $T_{B\underline{25}}$, whose 181 correct submissions we considered separately. Amongst all respondents, 824 (74 %) provided their full first name, 1 % initials only, and 25 % did not provide an answer. A single one respondent submitted a first name which is very likely to be fake ("Gaius Julius").

For analysis purposes, we group $T_{\underline{50}}$ and $T_{\overline{\underline{50}}}$ as the high-paying (hereafter denoted collectively as T_{50}) and $T_{\overline{\underline{25}}}$ and $T_{\underline{25}}$ as the low-paying treatments (denoted T_{25}) respectively. $T_{\underline{50}}$ and $T_{\underline{25}}$ are grouped as the treatments with low data requirements (denoted $T_{\underline{\star}}$) and $T_{\overline{\underline{50}}}$ and $T_{\overline{\underline{25}}}$ as those with the high data requirements (denoted $T_{\overline{\star}}$).

9.5 Descriptive Statistics

Figure 9.2 shows the proportion of respondents who provided each data item, broken down by treatment. Across our sample, full date of birth was the data item omitted most often. Of all submissions, 57 % included full details, that is day, month and

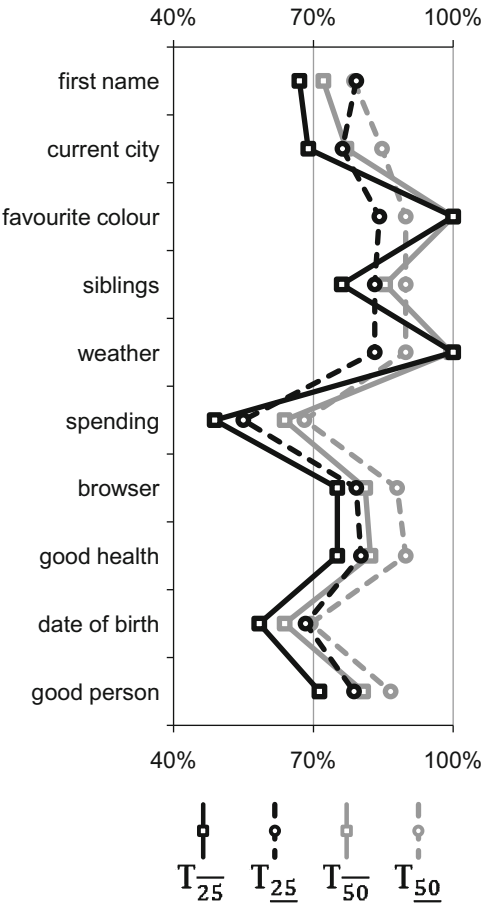

Fig. 9.2 Proportion of respondents who provided each data item, broken down by treatment. The data items are ordered by their order of appearance in the form. *Solid lines* represent treatments with a high data requirement ($T_{\overline{\star}}$) for which favourite colour and current weather were mandatory. *Black lines* correspond to treatments with a low base reward (T_{25})

year; 68 % provided parts of their date of birth. Details about the date of birth were submitted significantly less often than the second- and third-most-often omitted data items, which were spending and first name (two-tailed paired t-test: $p = 0.005$ and $p < 0.0001$ respectively). We will thus consider date of birth *the* sensitive item. Spending is a profile-related data item; the low response rate for this question may also originate in respondents' inability to recall their last major purchase. Answers concerning weather and favourite colour were included most frequently.

Neither in the form, nor in the follow-up did we ask for demographics. All the same, age and gender may be inferred for those who gave us their date of birth and first name. These details may be fabricated, just like answers to direct demographic questions. The reported year of birth ranged from 1941 to 1996 with a median of 1981, corresponding to an age of about 30 years at the time of the experiment, after outlier correction. The cohort of the 1980s accounts for 41 % alone. When dividing the entire participant population at the median age, there is a significant ($p < 0001$,

Table 9.3 Prevalence of the eight most common browsers used by our participants (IE: Internet Explorer). 50 of the 67 participants using a Firefox older than version 6 were users of version 3.6.

			Count by version lag				
Browser	total	Current version	−0	−1	−2	−3	older
Firefox	422	9	242	27	11	6	67
Chrome	365	16	321	2	3	1	7
IE	221	9	80	105	35	1	0
Safari	68	5	63	4	1	0	0
iPad	11						
Opera	9						
Android	5						
Chromium	2						

two-tailed t-test) trend that older participants answered fewer questions. However, age only explains 2 % of the variance in completion rates and the effect size is minute. On average, the older half provided 0.18 items less than the younger half of participants.

We guess participants' gender by matching their first names against the list of common first names from the 1990 US census. For ambiguous names, the dominant gender is chosen. According to this inference, 34 % of participants were male and 33 % female. For the remainder, the first name was not given or could not be found in the census name files. There is no significant difference in completion rates between the males and females ($p = 0.22$, two-tailed t-test).

Of all respondents, 99 % had JavaScript enabled in their browser; 2 % were participating through a mobile device. Of all respondents, 66 % (78 %) were running the most (or second-most) recent version of their browser, as far as we could tell from the HTTP request headers (Table 9.3). Our sample was using more recent browsers than the general online population [26]. The four most prevalent browsers, Firefox, Chrome, Internet Explorer and Safari, accounted for 97 % of all observed browsers.

As part of the form, we asked participants for the browser they were using. Nineteen per cent left this field blank. Amongst those who provided an answer, 96 % correctly named their browser identified from the HTTP request headers. Less than 2 % provided an incorrect answer, such as "Google?", "Windows 7" or "Word". The remaining 2 % indicated a browser that did not match the HTTP headers; from mTurk Web forums, we know that some workers use several browsers simultaneously.

Such high awareness of browser type, and the use of such modern browsers, suggests that our sample may have a higher-than-usual level of computer literacy.

9.6 Analysis

9.6.1 Multivariate Analysis into Disclosure Behaviour

In addition to analysing systematic associations between the participants psychometrics, attitudes and their disclosing behaviour (Sect. 9.6.2 and following), we performed multivariate ordinal logistic regressions into the number of data items disclosed and the disclosure of date of birth in particular. The −2-Log-Likelihood model fitting criterion exhibited a very good significance in both cases ($p < 0.0001$, Chi-square test). Treatment parameters (base reward T_{50} vs. T_{25}, data requirement $T_{\underline{\star}}$ vs. $T_{\overline{\star}}$, presence of a bonus), response to the check questions, the browser used by the respondent, enjoyment, motives for participating, and perceiving data as sensitive or personal were used as categorical factors, plus reciprocity (negative and positive) as a metric covariate. We report the relevant factors here; the full parameter estimates are available from the authors.

If the correct completion of the check questions are taken as an indicator of having read and understood the instructions, then date of birth is disclosed significantly more often when the instructions are not understood ($p < 0.0001$). Disclosure decreases when the data provided is perceived as personal ($p = 0.003$). Enjoying the form significantly increases disclosure ($p = 0.002$). Neither reciprocity, nor any of the different motivations for participating and submitting the form are systematically associated with disclosing behaviour for date of birth.

The total number of data items provided above and beyond the check questions is also not systematically influenced by reciprocity as a personality trait. Amongst all coded motivations, only enjoyment increases disclosure weakly significantly ($p = 0.09$). Again, not passing the check questions results in more data items provided ($p = 0.02$). Differences between the treatments are without systematic influence. As an aside, users of Internet Explorer are more likely to fill in more data fields ($p = 0.004$). Anecdotally, using an old version of a browser instead of a current one is associated with fewer data items being provided (not significant). Higher computer literacy decreases privacy concerns [10], which could facilitate over-disclosure.

Results from regression analysis suggest that differences in payment, as the independent variable manipulated across treatments, does not impact on disclosing behaviour. There is however strong evidence for over-disclosure by accident due to not reading the instructions.

9.6.2 Hypothesis 1

Web users provide more personal information than requested by a form, even though they realise there is no prospect of monetary reward for doing so.

In all but two treatments in our experiment, questions 5 and 6 were mandatory and all other questions were optional. In $T_{\overline{\star}}$, questions 3, 5, 6 and 7 were mandatory and all other questions were optional. A data item is revealed significantly less often when it is optional instead of mandatory (Fisher's exact test on favourite colour and sunny weather in treatments $T_{\underline{\star}}$ vs. $T_{\overline{\star}}$: $p < 0.0001$). Across all treatments, optional questions were answered by a significant proportion of the participants (Fisher's exact test on full date of birth in $T_{\overline{25}}$ as the limit case of lowest disclosure: $p < 0.0001$).

Hypothesis 1 is therefore *supported*. We also found strong and significant evidence for overly detailed disclosure, as discussed below (Sect. 9.6.5).

9.6.3 Hypothesis 2

The base utility the Web user reaches by submitting the form does not determine the extent of over-disclosure.

For analysis, we combine all treatments with low base reward, T_{25}, and all with high base reward, T_{50}, respectively. For low-sensitivity data items, the disclosure ratio[1] increases significantly with the base reward (weather: $p = 0.001$, favourite colour: $p = 0.001$; G-test); it also increases in the base reward for medium-sensitivity data items (good person: $p = 0.003$). For date of birth and first name, however, the high-sensitivity data items, there is no significant association of the reward level and the disclosure behaviour.

Hypothesis 2 is therefore *partially supported* for high-sensitivity data items, but otherwise *rejected*. The results do not differ by high or low reciprocity as a participant's personality trait. Note that in this experiment, with the exception of the bonus treatment $T_{B\underline{25}}$, the base reward was independent of participants' actual disclosure.

9.6.4 Hypothesis 3

Over-disclosure of personal data is not an accident.

Two check questions were built into the form, the answers to which revealed whether or not the participants had read and understood the instructions. Of all participants, 93 % correctly identified that none of the answers (except the check questions themselves) were mandatory. We observe that over-disclosure is

[1] Revelation ratio or *disclosure ratio* is the proportion of times that a given input field on a form was completed versus the total number of times this form was submitted.

significantly more prevalent amongst those who did not read the instructions (date of birth: $p < 0.0001$, 67 % vs. 87 %; good person: $p = 0.0001$, 81 % vs. 90 %; Fisher's exact test). Nevertheless, the majority of participants who understood the instructions over-disclosed (Sect. 9.6.2).

The participants knew they had disclosed personal information. In the follow-up questionnaire, 62 % of all participants indicated their submission contained personal data, and 8 % felt this personal data was sensitive. In an open-ended questions without any prompts, respondents to the follow-up named data items considered sensitive. Date of birth was listed as sensitive by 2 % of all respondents, or 43 % of those who gave some data item in their free-text responses.

There is a significant positive association between completing a field in the form and indicating, in the follow-up questionnaire, that the participant felt they had revealed 'personal' information (date of birth: $p < 0.0001$, good person: $p < 0.0001$, weather: $p = 0.003$, favourite colour: $p = 0.001$; G-test). Only the provision of date of birth made participants describe their submission as containing 'sensitive' information ($p < 0.05$).

We conclude that information was, first, provided knowingly voluntarily, and, second, perceived as personal. Hypothesis 3 is therefore *supported*.

9.6.5 Hypothesis 4

Over-disclosure is costly to the user.

The user has to spend more effort completing optional fields. This effort includes both the time taken and the physical typing activity. We observe that participants who complete all instead of none of the optional fields take significantly longer ($p < 0.0001$, t-test). A regression analysis reveals that participants spent around 57 s reading the form plus additional 3.5 s per field completed ($p < 0.0001$, t-test on the regression coefficients; outlier detection based on inter-quartile range). Completion times are depicted graphically in Fig. 9.3. Interestingly, most participants largely over-estimate the time spent on the form. In the follow-up questionnaire, 86 % of respondents had an estimate of the time spent that exceeded the actual time. For 13 % of the participants, their estimate was more than 10 times larger than the actual time.

Participants who over-disclose also have to type more. The minimum number of characters typed in total increases linearly in the number of fields completed at a rate of 4.0 characters per field ($p < 0.0001$, t-test on the regression coefficients; outlier detection based on inter-quartile range; 98 % variance explained). However, the median number of characters typed in total increases quadratically (!) in the number of fields completed (99 % variance explained).

As a special case of over-disclosure, we consider overly verbose answers to simple questions, taking the example of weather and spending. For this analysis, we distinguish between three levels of disclosure, again, only considering correct

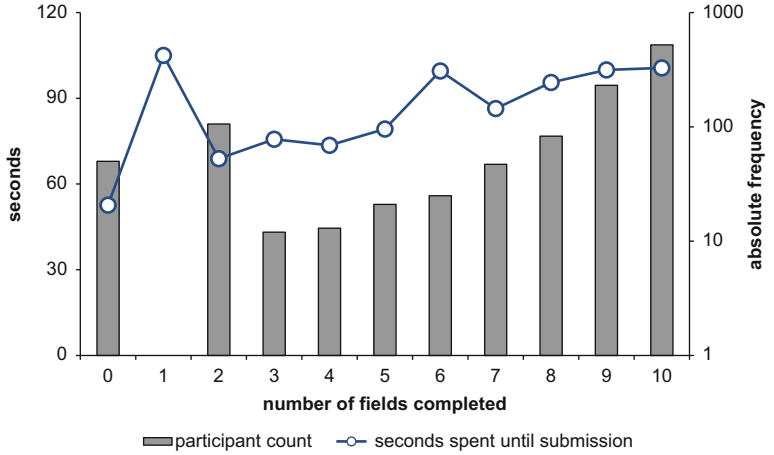

Fig. 9.3 Time spent on completing the form and number of participants, by number of fields completed on *top* of the check questions

submissions: field left blank, simple answer provided, simple answer given with details that were not asked for. Examples for responses of the latter type include: "No. It's currently cloudy and rainy" or "no, its cloudy and snowing" when asked for sunny weather, and "last week on textbooks" or "4 days ago getting groceries" when asked when they last spent over $100. 6 % of all participants answering the weather questions provided details that were not asked for; 14 % of those indicating the time of their $100+ purchase also indicated the purpose of spending. In both cases, the prevalence of overly detailed answers is significant ($p < 0.0001$, Fisher's exact test).

From the consistent evidence, we conclude: Hypothesis 4 is *supported*.

9.6.6 Hypothesis 5

Over-disclosure is not seen negatively.

We were concerned that participants might have over-disclosed personal information involuntarily. Norms and perceived expectations could compel users to over-disclose. The evidence from the follow-up questionnaire indicates otherwise: From the follow-up questionnaire, we know that 98 % of all participants enjoyed completing the form. Acknowledging that taking the follow-up may have introduced a sampling bias, we observe that this share corresponds to 74 % of all of the original participants. Of all respondents to the follow-up, 99 % (or 74 % of the original sample) said they wanted more of these form-filling tasks in the future.

Given the high prevalence of enjoyment, more fine-grained analysis is limited by the low number of those who did not enjoy the task. We combine treatments by their data requirement and take the example of date of birth. Regardless of the data requirements, there is no significant association between enjoying the form and disclosing ($T_{\underline{\star}}$: $p = 0.22$, $T_{\overline{\star}}$: $p = 0.57$, Fisher's exact test). Also, enjoyment as a motivation to participate is not systematically associated with over-disclosure ($T_{\underline{\star}}$: $p = 0.17$, $T_{\overline{\star}}$: $p = 0.90$, Fisher's exact test). The non-significant trend shows that enjoyment was more prevalent amongst those who provided their date of birth.

Hypothesis 5 is therefore *supported*.

9.6.7 Hypothesis 6

Users have good reasons to over-disclose personal information.

Using an open-ended question, we asked participants in the follow-up questionnaire why they had completed the form in the first place. The free-text answers were coded into up to three reasons for each respondent. Money had motivated 54 % to participate. This comes as no surprise since we recruited our participants on a crowd-sourcing platform. The form looked easy to 30 %, so they completed it.

Of all respondents, 15 % indicated they had participated out of joy; 25 % because it was interesting. Original responses include: "I enjoy filling out surveys", "I enjoy doing surveys as a way to destress [sic]" or "It looked interesting, fun and easy to do". This is opposed to the received wisdom that form-filling is a nuisance. To have their opinion heard or to help research was named by 3 and 8 % of the follow-up respondents respectively. Exemplary answers include: "I think it's really cool to be part of a statistic analysis, to contribute my thoughts and experiences to a collective body of information", "my information goes towards creating a change in something", "the opportunity to present an underrepresented demographic (conservatives, mothers) in surveys" or "I like taking surveys to get my opinions heard". This is in line with the motives for over-disclosure identified in Sect. 9.2. Those being motivated by helping research named both the researcher and research per se, such as: "I enjoy helping researchers", "Any help I can be for research, I am glad to do" or "I appreciate helping (even if only a little) with research".

These motives work towards completing specific fields in the form (e.g., for opinion shaping) and completing more fields on the form (e.g., when wanting to help research). They can be understood as antecedents for over-disclosure. There is a high proportion of participants motivated by the base reward or the form itself, combined with the low proportion of those who used the form to express opinions (2 %), or were motivated by trust in the university (4 %). This is promising as it hints that our results might have more general validity. A G-test for each treatment individually and for all treatments together indicates that over-disclosing date of birth, being a good person and favourite colour, does not depend on whether participants were motivated by money or not.

Respondents who provided more data items (nine or ten versus eight and below completed fields) also enjoyed the form more, although the direction of a causal relationship, if any, remains to be determined (approaching significance: $p = 0.06$, G-test).

We also saw evidence that participants were motivated by other reasons named under Sect. 9.2, to disclose more personal data than required. We report anecdotally some of the reasons given for disclosing date of birth. Respondents abided by social or self-imposed norms to submit a full form: "I feel a certain obligation to completely fill out surveys", "A completionist [sic] instinct", "because it was asked", "Completeness", "i felt that i should complete all aspets [sic]", "I like to fully comply with requests". The last example in particular indicates that the socially desirable behaviour for an optional field may be to complete it rather than skip it. The desire to be helpful (in an altruistic or reciprocal sense) was also frequently reported.

Entering the data also happened habitually ("Force of habit", "Habbit [sic]", "habit", "I probably automatically put it down without thinking"). Many respondents reported they did not know why they had provided their date of birth, which indicates the lack of a conscious decision (or deliberation)—a strong sign of a habit. We even observed cases of extroversion ("I have a unique birthday, it being on christmas [sic], so i just wanted to share").

Some participants anticipated future payoffs and intended to improve their social capital on the platform ("to boost my mturk hit approval rate", "Even though it was optional, I thought that if I did not disclose the information, you would be unable to classify me for future HITs and I would miss out on the opportunities").

From the combined evidence, we conclude: Hypothesis 6 is *supported*.

9.6.8 Hypothesis 7

Making some form fields mandatory reduces disclosure for the remaining optional fields.

We compare data disclosure in treatments $T_{\underline{\star}}$ and $T_{\overline{\star}}$ to see if the mandatory revelation level makes a difference for the disclosure of the remaining, optional fields. In $T_{\overline{\star}}$, weather and favourite colour were mandatory; they are two data items revealed voluntarily most often in $T_{\underline{\star}}$.

Making two low-sensitivity fields mandatory decreases the revelation ratio for the high-sensitivity item date of birth ($p < 0.02$, G-test) as well as for the medium-sensitivity item of being a good person ($p < 0.04$, G-test).

We now only consider participants who provided answers to questions 3 and 7, regardless of whether they were in a treatment where these questions were mandatory or optional. Amongst this cohort, disclosure behaviour for the remaining fields depended on whether questions 3 and 7 were marked as mandatory or optional. The average number of fields completed when questions 3 and 7 were marked as mandatory is reduced by about 1.3 fields in $T_{\overline{\star}}$ compared to $T_{\underline{\star}}$ ($p < 0.0001$,

two-tailed t-test). The data therefore suggests that as the number of mandatory fields in a form is increased, the total number of completed fields reduces. Also negatively affected is the revelation ratio for date of birth: fewer participants are willing to disclose it in $T_{\overline{\star}}$ than in $T_{\underline{\star}}$ ($p < 0.0001$, G-test).

Hypothesis 7 is therefore *supported*.

9.6.9 Hypothesis 8

A reward for some form fields reduces disclosure for the remaining optional fields.

We benchmark incentivised disclosure against voluntary and mandatory disclosure by comparing $T_{B\underline{25}}$ with treatments $T_{\underline{25}}$ and $T_{\overline{50}}$. $T_{\underline{25}}$ is the limit case for $T_{B\underline{25}}$ respondents who do not accept the incentive; $T_{\overline{50}}$ is the limit case for those who choose to collect the incentive through extra disclosure.

Incentives yield a similar disclosure ratio as mandatoriness ($T_{B\underline{25}}$ vs. $T_{\overline{\star}}$: $p = 0.55$, Fisher's exact test). They improve disclosure for fields that are optional ($T_{B\underline{25}}$ vs. $T_{\underline{\star}}$: $p < 0.0001$, Fisher's exact test).

To our surprise, we do not see evidence for crowding-out of incentives, whereby a monetary incentive would replace the intrinsic motivation and yield an overall lower inclination to cooperate. On the contrary, there is strong evidence for crowding-in. When comparing $T_{B\underline{25}}$ and $T_{\overline{50}}$, we find that incentives for disclosing low-sensitivity data also increase disclosure for the remaining, optional, medium- and high-sensitivity fields on the same form (good person: $p = 0.002$, date of birth: $p < 0.001$; Fisher's exact test).

Hypothesis 8 is therefore *rejected* and we find significant evidence for the opposite relationship.

9.7 Summary and Discussion

Forms are ubiquitous on the Web. They are the primary mechanism used to collect personal information relating to one's identity or profile. They are the metaphor for the explicit invasion of privacy.

The received wisdom is that completing Web forms is a nuisance. User experience practitioners have argued for various styles to make the form filling exercise more comfortable. However, their design recommendations do not appear to have been backed by empirical evidence. At the same time, browser vendors and add-on programmers have eased the mechanics of form filling, in particular through the autocomplete feature.

In privacy economics, we are interested in two aspects of filling in forms: the time spent (including mechanical effort) and the invasion of privacy. The traditional assumption is that Web users complete as few fields as possible on a Web form to

Table 9.4 Summary of findings from the field experiment: overview of supported and rejected hypotheses and operationalisations used. The worst significance level is reported if the same operationalisation applied to several data items of different sensitivity (sensit.); no significance levels are reported for operationalisations based on pure occurrence

Hyp.	Support	Operationalisations	Signif./prop.
H1	Supp.	Optional revealed less often than mandatory	$p < 0.0001$
		Optional revealed more often than necessary	$p < 0.0001$
H2	Partial	Discl. ratio increases in base reward (low sensit.)	$p = 0.001$
		Discl. ratio increases in base reward (medium sensit.)	$p = 0.003$
		Discl. ratio increases in base reward (high sensit.)	n.s.
H3	Supp.	Not reading instructions and over-disclosure	$p = 0.002$
		Over-disclosed subjectively personal information	$p = 0.003$
H4	Supp.	Over-disclosure is time-consuming	$p < 0.0001$
		Overly detailed disclosure is prevalent	$p < 0.0001$
H5	Supp.	Enjoying the form	n/a: 98 %
		Wanting more of such forms	n/a: 99 %
		Over-disclosure and enjoying less	$p = 0.19$
		Over-disclosure and less motivated by joy	$p = 0.32$
H6	Supp.	Motivated by joy	n/a: 15 %
		Motivated by interest	n/a: 25 %
		Motivated by ease	n/a: 30 %
		Motivated by opinion shaping opportunity	n/a: 2 %
		Motivated by contributing to science	n/a: 8 %
		Motivated by monetary prospects	n/a: 54 %
		Motivated by trust in university	n/a: 4 %
H7	Supp.	Medium sensit. revelation ratio	$p < 0.04$
		High sensit. revelation ratio	$p < 0.02$
		Average number of fields completed	$p < 0.0001$
		High sensit. revelation ratio given compliance	$p < 0.0001$
H8	Rej./opp.	Incentives increases disclosure for trigger fields	$p < 0.0001$
		Incentives increases disclosure for remaining fields	$p = 0.002$

reduce their privacy exposure and save on typing. We challenge this assumption with the first field experiment into the prevalence and extent of voluntary over-disclosure on Web forms. The empirical evidence (Table 9.4) gives a consistent picture.

Firstly, over-disclosure occurs commonly and this is no accident. Across all levels of sensitivity, Web users provide data items for which they know disclosure is optional and not rewarded. In doing so, they reveal information subjectively considered as personal and they incur significant costs in terms of typing effort and time spent on the form.

Secondly, Web users have good reasons for disclosing personal data despite the negative side-effects. These motives include well-being by abiding to social norms and one's personality, reciprocity, shaping public opinion, and also the build-up of social capital. A base reward, independent of and not systematically associated with disclosure, remains the strongest driver for submitting the form at all.

Thirdly, for website operators, optional fields deliver a good data return, even for sensitive data items, which may explain why we still find them on the Web. Operators should be cautious, however, that increasing revelation ratios by making fields mandatory can backfire, because it jeopardises voluntary disclosure for the remaining fields on the form. A better approach is incentives for voluntarily provided optional data. Rewards for extra disclosure have the added benefit of crowding-in, stimulating further disclosure on the form beyond the incentive.

9.7.1 Recommendations and Managerial Implications

The implications for industry are quite profound. The single most important message is to mandate fewer fields. More mandatory fields mean less voluntary data disclosure whatever the sensitivity of a data item. Optional fields yield a good data return—in particular when users are unaware of their optionality. If indistinguishable from mandatory fields, voluntary fields are rather filled in than skipped.

Next, websites should capitalise on Web users' motives for voluntary disclosure. It can be helpful to frame data collection as a social exchange rather than an economic exchange. Our results (Hypothesis 2) also suggest that the privacy-friendly opt-in to personalisation features is viable. The increased base reward from a personalised service can stimulate further disclosure on low and medium sensitivity data items. We also recommend the use of free text input fields even for yes/no answers: they allow fine-tuned hiding and over-disclosure alike, thereby appealing to both privacy concerned and unconcerned users. Free text fields give opportunity to talk so that customer service can learn issues and needs.

We add to the policy debate with the following ideas for regulation and for browser behaviour as de facto standards. In line with current data protection legislation, website operators should make the purpose of data collection explicit. Otherwise, Web users come up with their own good reasons for providing personal information, typically resulting in over-disclosure. Regulators assessing the privacy invasion of a Web form should be aware that an optional field is often as privacy-invasive as a mandatory field.

Browsers can help users to limit their flow of personal information. With HTML5, there is now an attribute to distinguish optional fields from mandatory fields. The browser could blur optional fields, delay or disable autocomplete for optional fields, and warn the user if a form submission contains optional fields.

In the meantime, educating Web users to identify form fields as optional is crucial so they can spot opportunities for data hiding. One of our participants reported in the follow-up: "I will be much more careful in the future about giving out my personal information. Thank you for this very important lesson that I have learned."

To academics across disciplines using Web forms, for instance in a survey or exit-questionnaires, we also recommend re-assessing the privacy invasion of optional fields. Further, privacy was salient in our setup: two third of our respondents were

aware they had submitted personal information. We also take the opportunity to reiterate good practice in field experimentation: test thoroughly, pilot, monitor and be open to receive feedback from your participants.

We also caution researchers in privacy economics to prepare their control treatments carefully. The voluntary over-disclosure of personal information warrants further research into privacy-friendliness as a desirable property! Web users' preference to reveal personal data could be so strong that they prefer privacy-invasive alternatives over privacy-friendly alternatives—a serious threat to the validity of control treatments. Given the ambiguous role of optional fields, we also recommend experimenters to consider making all fields mandatory.

9.7.2 Limitations and Future Work

We are aware of the limitations of our findings. They are of three kinds.

Firstly, our results may exhibit a *trust bias* that originates from our university status rather than as an unknown commercial entity; having said this, only a minority of respondents (4 %) named trust in the university as a driver for participation. Our results may not generalise to other transactional Web forms found in electronic commerce or online social networking. We knowingly incurred this limitation in external validity for the sake of internal validity. We also emphasise that well-known retailers or social networks may benefit from similar trust biases resulting from brand effects. Indeed, a trust bias in favour of our experiment may have originated in good ratings on relevant mTurk forums rather than in our status as a university.

Secondly, owing to the deployment specifics of our field experiment, we were limited in our ability to perform *data verification*. We checked users' responses regarding their Web browsers and found truthful reporting for the overwhelming majority of participants. We planned to verify participants' answers for their current city with the location returned by geo-IP. However, in the end we did not do so, as we were unable to obtain detailed knowledge of nested geographical areas. We inspected all other data items for syntax correctness and plausibility. We note that a commercial website is similarly handicapped in its ability to test the accuracy of users' personal information, but the motives for voluntary over-disclosure work against lying. Also, misreporting only partially affects the economics of over-disclosure: the typing effort for an answer is independent of its truthfulness. Still, as future work, we are currently considering mechanisms to enforce truthful reporting or at least assess the prevalence of misreporting.

We performed analyses only on participants who had successfully passed both check questions that tested for understanding the instructions and the optionality of the data items explained therein. We acknowledge that this results in underestimating the extent of over-disclosure: the 14 % of participants who did not answer the check questions correctly, and supposedly ignored the voluntariness of disclosure, were strongly significantly more likely to over-disclose.

Thirdly, we acknowledge a potential *sampling bias*. By deploying on mTurk, it is possible that we only recruited form-lovers: the skills and the mindset of mTurk workers may be such, and they are trained to complete forms quickly. When optimising for speed, uniformity and thereby over-disclosure may be more desirable than time-consuming, selective disclosure. However, we carefully checked compliance with instructions and removed participants from our analysis who had not passed the check questions. The mTurk platform does not provide access to participant statistics, such as the number of previously completed tasks for each worker, which could have been moderating variables. We considered requesting those details in the follow-up questionnaire, however, in the end, we decided other questions were more important given our budget constraints and the need for a high response rate (and therefore short follow-up questionnaire).

Workers on the mTurk platform may deliberately over-disclose to increase their chance of future working opportunities. In our case, the participant might have believed the Web form was the first one in a series. In the exit questionnaire, we did ask participants for their motivation and found only limited evidence for signalling behaviour. We further notice that the quest for future tasks cannot explain over-disclosure at the level of detail, and that most mTurk workers are trained for compliance rather than volunteering personal data. Reputation building on mTurk works mainly for and against the requester whose tasks are chosen (or ignored) by the worker population.

Our participants' submissions indicated privacy concerns. We also consider our sample more representative of the Western online population than a convenience sample of computer science or psychology undergraduates. All in all, we are therefore confident that our findings generalise beyond the sample at hand. In particular, the following findings should hold beyond the mTurk environment: the relative magnitudes of revelation ratios; the moderating factors (or their lack of influence) for base reward, personality and signals such as the browser used; the effects from incentives and mandatory fields; and our estimates of typing effort and time spent.

Exploring the effects of aforementioned limitations are one possible strand for future research. Other promising avenues for future work include: How is over-disclosure affected by the number of fields on the form? Will over-disclosure persist when the penalties are increased beyond time, effort and sensitivity? As a priming/salience effect, does the explicit mention of data collection purposes impact disclosure ratios? Is form filling affected by the number of pages over which form fields are spread out? How do past experiences with forms, including misuse of data entered into them, affect disclosing behaviour in the long run? Privacy economics meet usability research in the ultimate conundrum for privacy advocates: why is it so easy to collect Web users' personal information?

Acknowledgements The authors wish to thank Google for providing financial support for this work through a Focused Research Award. Sören Preibusch and Kat Krol were supported by a Volkswagen Foundation travel grant to present this work at WEIS 2012. Kat Krol is supported by a scholarship from EPSRC (grant number EP/G037264/1).

References

1. Acquisti, A., John, L.K., Loewenstein, G.: The impact of relative standards on the propensity to disclose. J. Mark. Res. **49**(2), 160–174 (2012)
2. Anthony, T.: Why users fill out forms faster with unified text fields. http://uxmovement.com/forms/why-users-fill-out-forms-faster-with-unified-text-fields/ (2011)
3. Beresford, A., Preibusch, S., Kübler, D.: Unwillingness to pay for privacy: a field experiment. IZA Discussion Papers 5017, Institute for the Study of Labor (IZA) (2010). http://ftp.iza.org/dp5017.pdf
4. Berners-Lee, T., Connolly, D.: Hypertext Markup Language – 2.0. http://tools.ietf.org/html/rfc1866 (1995)
5. BITKOM: 12 Millionen Deutsche machen Falschangaben im Web. http://www.bitkom.org/62107_62102.aspx (2010)
6. BITKOM: Jedes vierte Mitglied flunkert in sozialen Netzwerken. http://www.bitkom.org/de/presse/70864_67989.aspx (2011)
7. Böhme, R., Pötzsch, S.: Privacy in online social lending. In: Proceedings of AAAI Spring Symposium on Intelligent Information Privacy Management, Stanford (2010)
8. Braun, H.: Chrome hilft beim Formular-Ausfüllen. http://heise.de/-1422502 (2012)
9. Denscombe, M.: Item non-response rates: a comparison of online and paper questionnaires. Int. J. Soc. Res. Methodol. **12**(4), 281–291 (2009)
10. Dinev, T., Hart, P.: Internet privacy, social awareness, and internet technical literacy–an exploratory investigation. In: Proceedings of the 17th Bled eCommerce Conference, Bled (2004)
11. Eastlake, D., Goldstein, T.: ECML v1: field names for E-commerce. http://tools.ietf.org/html/rfc2706 (1999)
12. Ellison, N.B., Steinfield, C., Lampe, C.: The benefits of Facebook "friends:" social capital and college students' use of online social network sites. J. Comput. Mediat. Commun. **12**(4), 1143–1168 (2007)
13. European Commission: EUROPA – EuropeAid – evaluation – guidelines: how is a questionnaire developed? http://ec.europa.eu/europeaid/evaluation/methodology/egeval/tools/too_qst_how_qst_en.htm (2005)
14. Frick, A., Bächtinger, M.T., Reips, U.D.: Financial incentives, personal information and drop-out rate in online studies. In: Current Internet Science. Trends, Techniques, Results, Nürnerg. Online Press, Zürich (1999)
15. GAIN Publishing: Gator.com – Home. http://web.archive.org/web/20031031020937/http://www.gator.com/home2.html (2003, 2012)
16. GAIN Publishing: Gator.com – Home (Important information about GAIN software). http://web.archive.org/web/20060630073649/http://www.gator.com/home2.html (2006, 2012)
17. Harhoff, D., Henkel, J., von Hippel, E.: Profiting from voluntary information spillovers: how users benefit by freely revealing their innovations. Res. Policy **32**(10), 1753–1769 (2003)
18. Helgeson, J.G., Voss, K.E., Terpening, W.D.: Determinants of mail-survey response: survey design factors and respondent factors. Psychol. Mark. **19**(3), 303–328 (2002)
19. James, J.M., Bolstein, R.: The effect of monetary incentives and follow-up mailings on the response rate and response quality in mail surveys. Public Opin. Q. **54**(3), 346–361 (1990)
20. Jarrett, C., Gaffney, G.: Forms That Work: Designing Web Forms for Usability. Morgan Kaufmann, San Francisco (2008)
21. McCabe, S.E., Boyd, C.J., Young, A., Crawford, S., Pope, D.: Mode effects for collecting alcohol and tobacco data among 3rd and 4th grade students: a randomized pilot study of web-form versus paper-form surveys. Addict. Behav. **30**(4), 663–671 (2005)
22. Microsoft: Collect demographic data more easily with Internet Explorer 5. http://msdn.microsoft.com/en-us/library/bb250414.aspx (1999)
23. Microsoft: How to use the autocomplete feature in Internet Explorer 4. http://support.microsoft.com/kb/171230 (2007)

24. Microsoft: How to use the autocomplete feature in Internet Explorer 5 and 6. http://support.microsoft.com/kb/217148 (2007)
25. Microsoft: Using autocomplete in HTML forms. http://msdn.microsoft.com/en-us/library/ms533032.aspx (2008)
26. Net Applications.com: Desktop browser version market share (Jan 2012). http://marketshare.hitslink.com/browser-market-share.aspx?qprid=2&qpcustomd=0&qptimeframe=M&qpsp=156 (2012)
27. NHS Wirral: Questionnaire design tips. http://www.wirral.nhs.uk/document_uploads/Governance/QuestionnaireDesignTips.pdf (2010)
28. Personalization Consortium: Personalization & privacy survey. http://personalization.org/SurveyResults.pdf (2000, 2005). Via Internet Archive
29. Philippot, P., Canter, S.: Fill in web forms automatically. PC Mag. **18**(16), 205 (1999)
30. Preibusch, S., Bonneau, J.: The privacy landscape: product differentiation on data collection. In: The Tenth Workshop on the Economics of Information Security (WEIS), Fairfax (2011)
31. Rubenking, N.J.: Autocomplete for web forms–is it safe? PC Mag. **19**(3), 105–108 (2000)
32. Schupp, J., Kroh, M.: Incentives and response rates – experience from the SOEP-innovation-sample 2009. In: 4th Conference of the European Survey Research Association, Lausanne (2011)
33. 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, Tampa, pp. 38–47. ACM, New York (2001)
34. Taylor, D., Davis, D., Jillapalli, R.: Privacy concern and online personalization: the moderating effects of information control and compensation. Electron. Commer. Res. **9**, 203–223 (2009)
35. TNS Infratest Sozialforschung: Living in Germany: survey 2005 on the social situation of households (individual question form). Technical report, SOEP, DIW Berlin (2005)
36. Tourangeau, R., Yan, T.: Sensitive questions in surveys. Psychol. Bull. **133**(5), 859–883 (2007)
37. van Eitzen, C.: Auto-complete: browsers disclose private data – update. http://h-online.com/-1043122 (2010)
38. W3C, Hickson, I. (ed.): HTML5. A vocabulary and associated APIs for HTML and XHTML, section 4.10 forms. http://www.w3.org/TR/html5/forms.html. W3C Working Draft 25 May 2011
39. Wroblewski, L.: Web Form Design: Filling in the Blanks. Rosenfeld Media, Brooklyn (2008)

Chapter 10
Choice Architecture and Smartphone Privacy: There's a Price for That

Serge Egelman, Adrienne Porter Felt, and David Wagner

Abstract Under certain circumstances, consumers are willing to pay a premium for privacy. We explore how choice architecture affects smartphone users' stated willingness to install applications that request varying permissions. We performed two experiments to gauge smartphone users' stated willingness to pay premiums to limit their personal information exposure when installing applications. When participants were comparison shopping between multiple applications that performed similar functionality, a quarter of our sample indicated a willingness to pay a $1.50 premium for the application that requested the fewest permissions—though only when viewing the requested permissions of each application side-by-side. In a second experiment, we more closely simulated the user experience by asking them to valuate a single application that featured multiple sets of permissions based on five between-subjects conditions. In this scenario, the requested permissions had a much smaller impact. Our results suggest that many smartphone users are concerned with their privacy and are willing to pay premiums for applications that are less likely to request access to personal information, but that the current choice architectures do not support this. We propose improvements for smartphone application markets that could result in decreased satisficing and increased rational behavior.

S. Egelman (✉) · D. Wagner
University of California, Berkeley, CA, USA
e-mail: egelman@cs.berkeley.edu; daw@cs.berkeley.edu

A.P. Felt
Google Inc., Mountain View, CA, USA
e-mail: felt@google.com

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_10, © Springer-Verlag Berlin Heidelberg 2013

10.1 Introduction

> Architecture starts when you carefully put two bricks together. There it begins.
>
> –Ludwig Mies van der Rohe

Nearly 90 % of U.S. adults own cellular phones [38], and over 40 % of these are smartphones [31]. Smartphones pose a challenging information security problem: users need to regulate how applications access their private information. Smartphones often store sensitive personal data, such as contacts, financial information, location information (e.g., GPS), and sensor data (e.g., cameras, microphones, and accelerometers). Smartphones need to simultaneously protect this data and support the installation of a variety of third-party applications.

Google's Android addresses this problem with user-granted *permissions*. Permissions govern an application's ability to make use of either personal data or sensor hardware. For example, an application can only read the user's list of contacts if it has the `READ_CONTACTS` permission. When a user installs an application from the Android Market, the central application repository, he or she is shown a warning screen that displays the set of permissions that the respective application requires. In order to complete the installation, the user must consent to granting all of the requested permissions to the application. Currently, this notice-and-consent process is all-or-nothing; the user cannot selectively grant or decline a subset of the permissions (i.e., the user must decline installation to deny the requested permissions).

We evaluate how the Android Market *choice architecture*[1] influences users' abilities and desires to protect their privacy, as evidenced by their stated willingness to pay premiums for applications that request fewer permissions. To explore this topic, we performed two online experiments: one to examine the extent to which users will consider permissions when comparison shopping, and another to examine the role of permissions when users are valuating a specific application. We designed our experiments to study the two primary shopping behaviors supported by the Android Market: function-specific searches and application-specific searches. While we performed our experiments using the Android platform, we believe our results are generalizable to other smartphone platforms.

During a *function-specific search*, users seek applications to perform specific tasks. When performing function-specific searches, users do not have a particular application in mind and are therefore willing to consider several different applications to fulfill the desired function (e.g., choosing one flashlight application amongst many). During an *application-specific search*, users seek a particular application that is known to them, such as through word of mouth, "popular" application lists, or advertisements. When performing application-specific searches, users are unlikely

[1] The term "choice architecture" refers to the way in which options are presented to people, as these design decisions can have a profound impact on decision-making [34].

to compare alternatives. For example, the decision of whether or not to install Angry Birds is an example of an application-specific search.

In our first experiment, users selected one application from a set of similar applications that requested different permissions. This experiment tested whether participants were willing to pay a privacy premium for an application that requested fewer permissions than the cheaper alternatives, when those alternatives were presented side-by-side. We found that 25 % of our participants stated a willingness to pay a $1.50 premium on a $0.49 application in order to grant the fewest permissions.

In our second experiment, we focused on application-specific searches. We told participants that we were a software company seeking beta testers for a new application. Participants submitted bids for the amount of compensation required to regularly use our application. These bids were proxies for participants' willingness to install the application. We constructed several between-subjects conditions by varying the permissions that participants saw. We also asked participants whether they would rather use a $0.99 version of the application or a free version supported by behavioral advertising. We made it clear that the advertisements would be targeted based on data collected as a result of the requested permissions.

Unlike our first experiment, wherein privacy-conscious participants opted to pay the highest premium for the fewest permissions, we observed that participants were satisficing under the more realistic conditions of the second experiment. We observed that only the request for a user's list of contacts had a significant effect on their bids; requests for location data or access to the user's photos had no observable difference over the control condition. Additionally, around 80 % of participants expressed a willingness to receive advertisements, regardless of the permissions used for the targeting, if it would save them $0.99. Our contributions are as follows:

- Prior work has focused on smartphone users' preferences for sharing location data. We found that users are less concerned about location data than other types of data commonly accessed by smartphone applications; participants were significantly more concerned over the use of address book data.
- We contribute to the literature on willingness to pay for privacy by examining decisions holistically: we measure privacy behaviors as part of a larger value proposition. We show that 25 % of participants in our first experiment were willing to pay the highest premium in order to grant permission to the least amount of personal data, when the options were presented side-by-side for easy comparison. However, current smartphone application markets make such comparisons very difficult. Our second experiment better approximated these current choice architectures. We found that when users are considering a particular application, they satisfice by downplaying their privacy concerns in favor of other considerations until those concerns reach a threshold.
- Our results lead to two suggestions. First, privacy-conscious users may be willing to spend more money for an application if the choice architecture supported comparison shopping for privacy. Second, users may be less likely to satisfice if

the decision to install a particular application were decoupled from the decision to grant it a set of permissions. Specifically, our results indicate that users may be better served by presenting permission requests when the data is actually needed, rather than requiring all permissions to be granted at install-time.

10.2 Background

In this section we provide an overview of application permissions on the Android platform, previous research on smartphone privacy that has focused on location sharing, and previous research on willingness to pay for privacy.

10.2.1 Android Permissions

Users find applications in the Android Market by searching by name, keyword, or by browsing lists of popular applications. When they select an application from a search result or list, they arrive at the application's description page. The description page includes developer information, screenshots, user reviews, the price, and a "Download" button. When they press "Download," the Market displays a warning screen that lists the application's permissions (Fig. 10.1).

Permissions govern access to privacy- and security-relevant actions on the phone, such as reading contacts or call history, connecting to the Internet, or sending text messages. Developers specify which permissions their applications require, and the user must agree to grant all of the application's requested permissions in order to install it. If the user does not consent to all of the permissions, the only option is to cancel the installation.

Smartphone applications often request access to private information for advertising, analytics, and other secondary uses [4, 16, 17]. These practices have generated public outrage; for example, consumer complaints forced Apple to promise to restrict how iOS applications can access contacts [32]. Free applications often request more permissions than paid applications because they are subsidized by sales of user data to advertising networks [10, 30]. Consequently, users who compare free and paid applications can choose to withhold personal information by paying a premium. This motivates our exploration into whether users are willing to pay to protect the personal information that is on their smartphones.

Researchers have built several tools to help users control how applications use their private information. Appfence [24] helps users protect their private data from exfiltration; it substitutes shadow data in place of private data and blocks network communication that contains user data. Kirin [15] operates under the assumption that users do not understand permissions and provides security rules to automatically accept or reject sets of permissions. Apex [29] lets users selectively

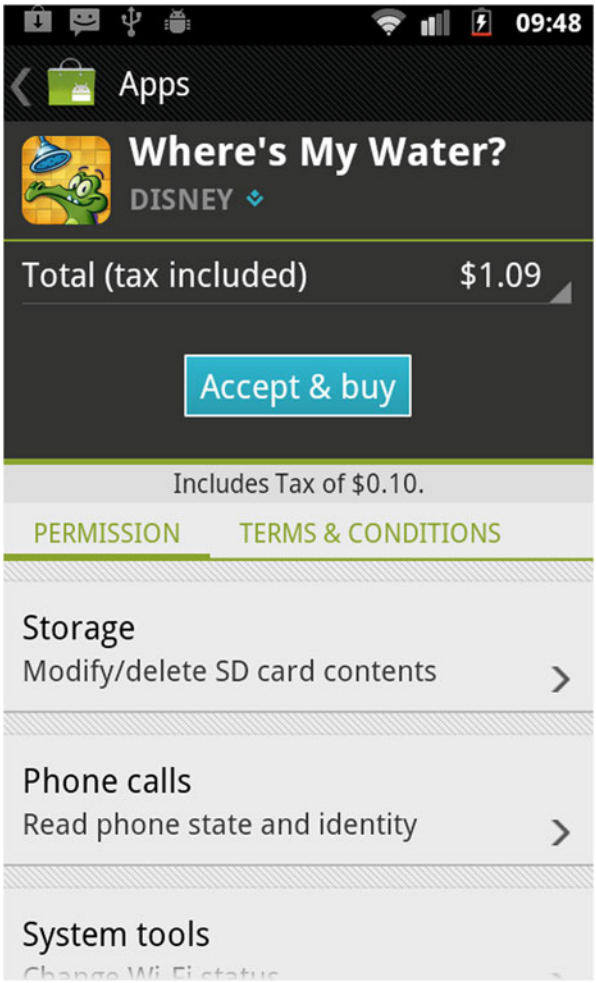

Fig. 10.1 Example of a permission request screen in the Android Market

grant or reject applications' permission requests. Rather than building tools that may never be adopted, we focus on evaluating users' willingness to share personal information.

We previously evaluated whether Android users pay attention to, comprehend, or act on permissions [20]. However, that study was limited to observing attention and comprehension rates, rather than the factors that go into installation decisions. We extend that work by empirically testing the influence of permissions on application selections. We also designed this study's experiments to be robust to the low attention and comprehension rates that we previously reported. In our experiments, we showed participants screenshots so that they did not ignore the permissions

by rapidly clicking through them, and in our second experiment we modified the permission warnings to avoid the comprehension problems that our previous study described. As such, our current results represent the decisions of fully-informed users.

10.2.2 Location Privacy

Previous research on smartphone privacy has focused on users' willingness to share location data with their social contacts [7,26]. Three independent studies [11,27,37] found that the identity of the data recipient was the most important factor that influenced users' sharing decisions, whereas Barkhuus [6] found that the user's current location matters more.

We explore users' willingness to share smartphone data beyond just location information. We examined several data types, such as contacts, audio, and photos. Also, we did not prime participants to think about sharing with social contacts or advertisers. As in current smartphone application markets, study participants determined on their own how they thought the requested permissions would be used by applications. Different users might be concerned about social contacts, employers, advertisers, law enforcement, governments, insurance companies, etc.

10.2.3 Willingness to Pay for Privacy

Users do not always act in accordance with their professed privacy concerns [2,33]. People are sometimes willing to trade privacy for convenience, functionality, or financial gain, even when the gains are very small [23]. Good et al. asked people to install applications after viewing privacy statements; regardless of privacy, they found that people will install applications if they think the utility is high enough [22]. Other studies have attempted to quantify the price that researchers and corporations would need to pay to buy users' location information [12, 13]. Thus, privacy decisions are not always consistent, and vary based on how the choices are framed. Two studies reported that users valuate their privacy differently when asked to pay to protect it rather than when asked to accept payment for disclosure (though in both cases, most placed a small financial value on their privacy) [3,23].

Acquisti hypothesized that privacy-concerned people are willing to trade privacy for small gains because they are not economically rational agents with respect to privacy [1]. He attributed users' actions to three factors that reduce economic rationality: incomplete information (i.e., unawareness of the risks associated with disclosure), bounded rationality (i.e., an inability to calculate all of the payoffs associated with privacy preservation), and psychological distortions (e.g., hyperbolic discounting, self-control problems, and immediate gratification). A survey

of 119 people supported this hypothesis: many respondents greatly overestimated or underestimated the likelihood and magnitude of privacy abuses, were unable to remember all of the parties involved in standard financial transactions, and were less likely to use privacy-enhancing technologies if the perceived risk was in the distant future [2].

A corollary to Acquisti's hypothesis is that people will act more like economically rational agents if they operate within a system that mitigates the effects of incomplete information, bounded rationality, and psychological distortions. Several studies have explored this. Gideon et al. [21] and Tsai et al. [35] asked users to purchase items using Privacy Finder, a search engine that included privacy ratings in search results. The search results provided users with additional information and made the tradeoffs easier to compute. Both studies found that participants were willing to pay premiums to purchase privacy-sensitive goods from merchants with better privacy policies when privacy ratings were displayed. Good et al. [22] similarly found that privacy and security are important factors when choosing between two applications with similar features, but not when considering an application on its own. This indicates that the timing and placement of privacy information is crucial. Egelman et al. [14] validated this by testing the timing and placement of privacy indicators when shopping online and found that even non-privacy-conscious shoppers will pay more for privacy when indicators are presented before visiting websites rather than after the user has already selected a website to visit.

Like past research on the economics of privacy, we aim to measure users' willingness to trade privacy for financial gain. However, we focus specifically on Android applications and smartphone data. We performed two experiments: the first showed participants side-by-side privacy information similar to that of Privacy Finder, and the second showed participants an individual application similar to the choice architecture of the Android Market and other smartphone application repositories. In the second experiment, we asked users to bid in a reverse auction for a chance to participate in a beta test of a particular smartphone application. We used these bids as proxies for willingness to pay and modeled our reverse auction on past studies that have used reverse auctions to gauge users' willingness to share information [12, 13, 25].

10.3 Privacy and Comparison Shopping

In October 2011, we deployed a survey to test whether smartphone users would be willing to pay more for an application if it requested fewer permissions than less-expensive alternatives. We asked participants to view screenshots from the Android Market of four fictitious applications. We counterbalanced the names, descriptions, and imagery, while controlling for price and requested permissions.

10.3.1 Methodology

For our survey, we created screenshots of four fictitious news aggregation applications as they might appear in the Android Market, but with one important difference: current smartphone application markets only allow application permissions to be viewed serially; we showed the four applications side-by-side to aid participants in contrasting their differences. We asked participants to choose which of the four they would be most willing to purchase. The purpose of this experiment was to examine whether participants would be willing to pay a privacy premium, when that option was apparent to them. The amount of personal information collected by each application was signaled by a permission request screen. Each fictitious application featured one of four possible prices: $0.49, $0.99, $1.49, and $1.99. These prices corresponded to four sets of requested permissions:

1. **$1.99**—`INTERNET`
2. **$1.49**—`INTERNET` and `ACCESS_FINE_LOCATION`
3. **$0.99**—`INTERNET` and `RECORD_AUDIO`
4. **$0.49**—`INTERNET`, `RECORD_AUDIO`, and `ACCESS_FINE_LOCATION`

We chose to focus on these three permissions for the following reasons:

- Access to the Internet (`INTERNET`) is the most frequently requested permission [18]. It also would be needed by a news reader to serve its intended purpose.
- GPS location data (`ACCESS_FINE_LOCATION`) has been heretofore the focus of most smartphone privacy research (see Sect. 10.2).
- Other research has found that the ability to record audio (`RECORD_AUDIO`) may be one of the most concerning permissions to users [19, 20].

We paired permissions to prices such that the least privacy-invasive application, which only requested `INTERNET`, had the highest price. The most privacy-invasive application, which requested all three permissions, had the lowest price. Participants with privacy concerns would need to pay a premium of $1.50 over the base price of $0.49 for the least privacy-invasive application. Since a previous study suggested that users are less concerned with location privacy than with an application's ability to make audio recordings [19], we set the price of the application with the `INTERNET` and `RECORD_AUDIO` permissions to be the second least expensive. Finally, the application with the `INTERNET` and `ACCESS_FINE_LOCATION` permissions cost $1.49, the penultimate price.

If participants had viewed four identical applications that only differed based on price and permissions, the purpose of the study would be obvious, which might have an impact on participants' responses. To minimize the potential for the Hawthorne effect, we created four applications with different names, manufacturers, screenshots, descriptions, and icons (Fig. 10.2). We counterbalanced these features such that each of the four price and privacy combinations was equally likely to be

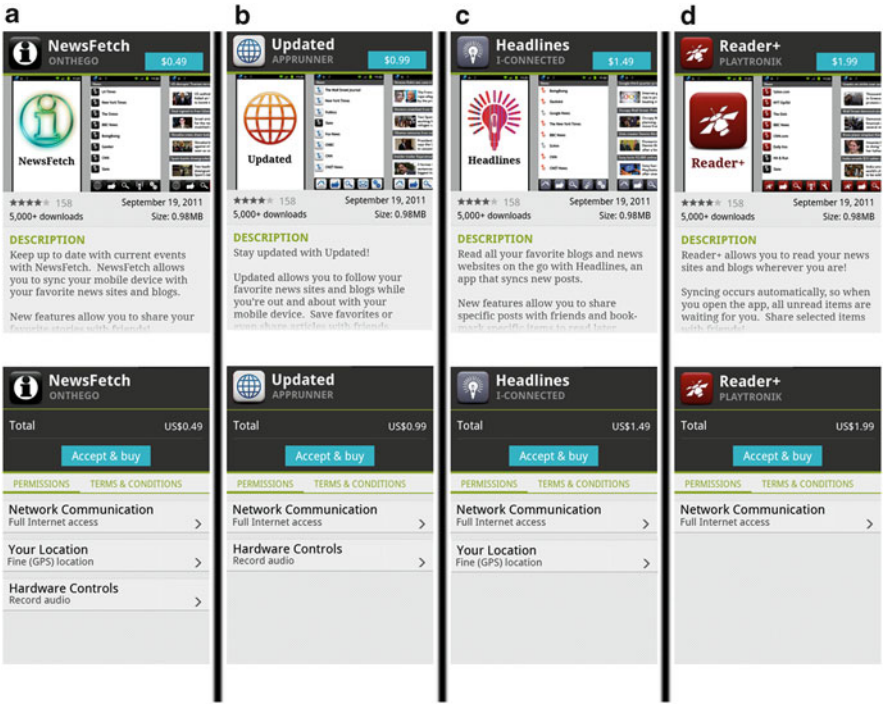

Fig. 10.2 We asked participants to choose the application they would be most willing to purchase. Privacy and price were inversely proportional, such that the most expensive application requested the fewest permissions. The other distinguishing features were counterbalanced

assigned to each application. A second concern was that users would anchor on the first price. To compensate for anchoring effects, we also counterbalanced the order in which the price and privacy conditions were presented; the conditions were either ordered from the lowest to the highest price, or from the highest to the lowest.

We were initially concerned that participants' responses would differ based on their general interest level in applications of this type. In particular, participants who did not have any interest in news readers might not put much thought into their selections. To test whether this occurred, we used the first page of the survey to randomly display one of the four applications and asked participants to indicate their willingness to purchase it using a five-point Likert scale (i.e., "extremely unlikely," "unlikely," "indifferent," "likely," or "extremely likely"). In order to equally distribute anchoring effects, we also randomized the price that was shown alongside the application. (For this preliminary question, we showed the application's description page but not the requested permissions, since those were irrelevant to this question.) Using Pearson's correlation, we observed no statistically significant correlations between participants' stated willingness to

Table 10.1 The four price/privacy variants and the number of participants who chose each variant

Price	Permissions requested	Total	
$1.99	INTERNET	120	(24.84 %)
$1.49	INTERNET ACCESS_FINE_LOCATION	74	(15.32 %)
$0.99	INTERNET RECORD_AUDIO	77	(15.94 %)
$0.49	INTERNET ACCESS_FINE_LOCATION RECORD_AUDIO	212	(43.89 %)

install an application and their selection in the subsequent experiment. As such, we concluded that we did not need to remove any participants due to lack of interest.

We recruited Mechanical Turk users who were U.S. residents over the age of 18. We did not limit participation to smartphone users, but we did ask about smartphone usage. We received a total of 483 valid survey responses, after screening out ten incomplete and questionable responses.[2]

10.3.2 Results

We considered three hypotheses:

H_0 = *Each price/privacy variant will be chosen with equal probability*
H_1 = *Cost-sensitive participants will choose the least-expensive option*
H_2 = *Privacy-sensitive participants will choose the high-privacy option*

We tested the null hypothesis to see if participants chose at random because they did not have any financial stake and found that this was not the case. A chi-square test showed that we can reject H_0 ($\chi_3 = 102.9$, $p < 0.0005$). A plurality (43.9 % of 483) selected the cheapest variant, whereas the second most popular variant was the most expensive one (24.8 % of 483), which afforded the most privacy (Table 10.1). We applied the Bonferroni correction to account for multiple tests ($\alpha = 0.01$) and found significant differences between the high-privacy variant and each of the others ($0.49: $\chi_1 = 25.49$, $p < 0.0005$; $0.99: $\chi_1 = 9.39$, $p < 0.002$; $1.49: $\chi_1 = 10.91$, $p < 0.001$). Thus, while more participants chose the cheapest variant than any others, more participants chose the high-privacy variant than the other two.

[2]We identified invalid results based on two factors. First, we included several questions that required free text responses, such as, "why or why not would you purchase this application." Using these questions, we deleted surveys that contained nonsensical responses. Second, in addition to asking participants to select the application that they were most willing to purchase, we also asked them to select the application that they were least willing to purchase. We removed participants who gave the same answer to both questions.

Table 10.2 Participants reported the factor that most influenced their decision to select a particular price/privacy variant. Each column lists the top three factors listed for each variant

	$1.99		$1.49		$0.99		$0.49	
1.	Permissions	(46 %)	Description	(41 %)	Cost	(33 %)	Cost	(62 %)
2.	Description	(23 %)	Permissions	(18 %)	Description	(15 %)	Description	(15 %)
3.	Icon	(9 %)	Cost	(16 %)	Icon	(14 %)	Rating/reviews	(6 %)
n	120		74		77		212	

10.3.2.1 Influencing Factors

Participants used a five-point Likert scale to rate the influence of the following factors:

- Number of Downloads
- Icon
- Size of App
- Permissions Requested
- Description
- Name of App
- Rating/Reviews
- Familiarity with App
- Cost
- Manufacturer of App

After applying the Bonferroni correction ($\alpha = 0.005$), we observed significant Pearson correlations between the extent to which participants reported being influenced by price and the price of the variant they selected ($r = -0.39$, $p < 0.0005$). We also observed a significant correlation between the degree to which participants reported being influenced by permissions and the price of the variant they selected ($r = 0.33$, $p < 0.0005$). These correlations support H_1 and H_2: participants who chose lower-price variants were more concerned about price, whereas participants who chose higher-price variants were more concerned about the permissions.

Despite counterbalancing most features to make each variant unique, participants who chose pricier variants were more likely to say that they were influenced by the icon ($r = 0.13$, $p < 0.004$). This was not because one icon was more appealing; each of the four icons appeared next to each variant with equal probability. Instead, we believe that the attention participants gave to each of these factors can be modeled as a zero-sum game: participants who were less concerned with price were more willing to base their selections on other factors (e.g., privacy and icons).

Finally, we asked participants to select the primary factor that influenced their selections. The majority of those who chose the $0.49 low-privacy variant claimed that cost was the primary factor (62.3 % of 212), whereas a plurality of the participants who chose the $1.99 high-privacy variant claimed that the permissions were the primary factor (45.8 % of 120). Table 10.2 lists the top three primary factors.

10.3.2.2 Permission Necessity

Participants may have believed that the permission requests reflected differences in application functionality not listed in the descriptions (e.g., they may have viewed the less-expensive alternatives as more functional rather than more privacy-invasive). To test this, we provided participants with a list of permissions and a picture of one of the applications' description screens,[3] and asked them to select all of the permissions that they believed would be required for the application to function as described. The description read:

> Read all your favorite blogs and news websites on the go with Headlines, an app that syncs new posts. New features allow you to share specific posts with friends and bookmark specific items to read later.

The permissions from which they could choose were as follows:

- Modify and/or delete your accounts
- Determine your physical location
- Read incoming or outgoing text messages (SMS)
- Read incoming or outgoing email messages
- Access the Internet
- Read your list of contacts
- Read your web browsing history
- Determine your phone number
- Determine which other apps are running
- Record audio
- Record video
- Send email messages
- Send text messages (SMS)
- Prevent device from sleeping
- Modify your storage card contents
- None of the above

We performed Phi correlations, corrected for multiple testing ($\alpha = 0.0125$), between whether participants thought a particular permission was required and whether they previously selected an application that requested that permission. We observed no significant correlations with regard to `RECORD_AUDIO`, since very few participants believed that the application needed this ability to function (7.9 % of 483). However, we did observe a statistically significant correlation with regard to `ACCESS_FINE_LOCATION` ($\phi = 0.21$, $p < 0.0005$). This indicates that some participants may have chosen the $0.49 and $1.49 variants because they believed that the `ACCESS_FINE_LOCATION` permission signaled desirable location-based

[3] We did not show the permission request screen. To negate priming, all participants viewed the $1.99 version, which was associated with only the `INTERNET` permission in the previous tasks.

features. Of all 483 participants, 91.3 % correctly understood that Internet access would be required for all four applications to function as desired.

We conclude that nearly all participants understood that the `RECORD_AUDIO` permission was unnecessary (92.1 % of 483). However, 59.3 % of these 445 participants were unwilling to pay a premium to deny the application this extraneous data. Although more people thought that the `ACCESS_FINE_LOCATION` permission was required, 51.9 % of the 316 participants who thought that it was unnecessary were unwilling to pay a $0.50 premium to deny the application this data.

10.3.2.3 Demographics

We collected participants' ages, genders, types of phones, and general privacy sensitivities. Our sample was 52.6 % female with an average age of 31.58 ($\sigma = 10.25$). Upon performing a Mann-Whitney U test, we found no observable differences between gender and the price selected. Based on our sample, there was no evidence that age or gender were correlated with willingness to pay for privacy.

A total of 372 respondents (77.0 % of 483) reported owning a smartphone, of which 42.7 % were Android-based. One potential confound is that existing Android users may understand the significance of permissions better than others. While Android users were significantly more likely to report that permissions influenced their decisions ($U = 19,848.5$, $p < 0.0005$), we observed no statistically significant differences with regard to which price/privacy variant they ultimately selected. Thus, while Android users may have been more familiar with the UI or the word "permissions," they were no more likely to factor them into their decisions.

To gauge general privacy sensitivity, we asked participants to rate three statements using a five-point Likert scale (from "I strongly disagree" to "I strongly agree") so that we could categorize them along the Westin privacy index [36]:

1. Consumers have lost all control over how personal information is collected and used by companies.
2. Most businesses handle the personal information they collect about consumers in a proper and confidential way.
3. Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

We classified participants according to Westin's metric [36]. Those who agreed with the first statement and disagreed with the second and third statements were classified as Privacy Fundamentalists (26.1 % of 483). Those who disagreed with the first statement while agreeing with the second and third statements were classified as Privacy Unconcerned (5.6 % of 483), while the remaining participants were classified as Privacy Pragmatists (68.3 % of 483). While 30.2 % of Privacy Fundamentalists were willing to purchase the high-privacy variant, compared to 23.3 and 18.5 % of Privacy Pragmatists and Privacy Unconcerned, respectively, these differences were not statistically significant. Thus, we did not find a correlation between the Westin privacy index and participants' behaviors.

10.4 Privacy in Context

The side-by-side comparison scenario that we tested in our first experiment is an idealized choice architecture that is not representative of any current smartphone application market. We performed a second experiment to examine how permission requests impacted behavior during application-specific searching: we asked participants to provide a bid for the amount of compensation that they would need to install a given application and to recommend a price at which we should sell our application in the Android Market. We varied the permissions that the application requested so that these bids were proxies for willingness to pay for privacy.

10.4.1 Methodology

The goal of our second experiment was to quantify the effect of permission requests on participants' valuations of a single application. We posed as a company named "AirZoom" and recruited participants from Amazon's Mechanical Turk who were 18 years old, based in the U.S., and current Android users to participate in a "private beta test" of an application. We performed a reverse Vickrey auction, modeled after Danezis et al.'s study on willingness to pay for location privacy [13]. In their study, participants bid on the amount they would need to be compensated in order to be tracked. However, in practice, a request for information is almost exclusively part of a larger value proposition (e.g., desirable location-based features). We wanted to examine the permission requests within the context of a larger value proposition: the amount of compensation participants would demand to install a given application. We asked participants how much they would need to be compensated to install and use a fictitious application for a month, as well as to suggest a reasonable price for us to charge for this application in the Android Market. We constructed several between-subjects treatments that differed only based on the permissions requested.

We displayed screenshots of our application, shown in Fig. 10.3, and asked participants to provide a bid for how much they would need to be compensated to participate. We also asked them to provide a suggested price for us to charge in the Android Market. We used language similar to Danezis et al.'s study [13]:

> We are recruiting current Android users to participate in a private beta test of this app. If you are selected to participate, you must purchase this app from the Android Market for $0.99 and install it onto your smartphone. You will be expected to use the app for at least one hour per week over the course of a month.
>
> Each person who is selected to participate in the beta will receive monetary compensation. We are running an auction to select those who will take part. We invite you to submit a bid for the amount of money you require to take part in the beta. Successful bidders will be those who bid the lowest amounts, and each will be paid the amount of compensation demanded by the lowest unsuccessful bidder. (We have yet to decide how many participants we will require.)

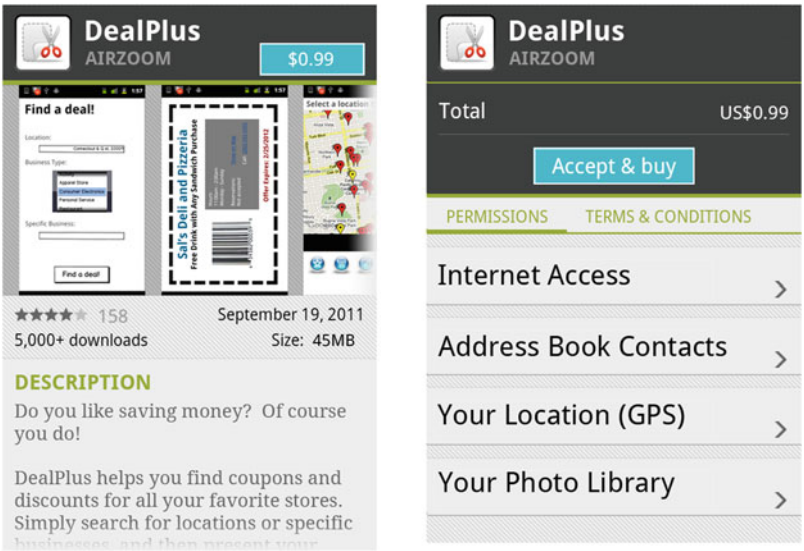

Fig. 10.3 Screenshot of our fictitious application that participants were asked to valuate. The permissions requested were randomly assigned from a set of five conditions

The fictitious application screenshots contained a request for up to four possible permissions: Internet access (`INTERNET`), location (`ACCESS_FINE_LOCATION`), address book contacts (`READ_CONTACTS`), and photos from the photo library (`PHOTOS`).[4] We chose the latter two because they were unrelated to the application's core functionality, as it was described to participants. We also chose them because our previous research indicated that they would be likely to raise concerns [20]. As before, we included a request for the `ACCESS_FINE_LOCATION` permission because location data has been the subject of extensive privacy research, and we wanted to compare participants' concerns about location to their concerns about other permissions. We also included the `INTERNET` permission because it would have been required for the application to function as described.

Each user saw one of five sets of permission requests: the `INTERNET` permission paired with one of the three other permissions, all four permissions together, or the `INTERNET` permission alone. We selected these sets of permissions to isolate the effects of individual permissions and observe the synergistic effect of the combination. (We did not test every combination to limit the number of experimental conditions.) We conducted a pilot experiment on a separate sample of 320 participants to determine whether participants differentiated between the `INTERNET` permission and the absence of any permission requests. We did not

[4]This Android permission does not actually exist; no permission is needed to access stored photos.

observe any significant differences and therefore used the INTERNET permission by itself as our control.

Android users often misunderstand permission requests because they only read the permission category rather than the full description [20]. To counter this, we removed the category text and increased the font size of each permission descriptions. We also altered the text of the permissions to remove other ambiguities that we previously observed [20]. Since we expect that participants better understood our modified permission warnings, our results are likely an upper bound.

We told participants that if selected, they would need to purchase the application from the Android Market for $0.99. Consequently, they needed to account for this in their bids. We added this caveat to minimize cheating and to limit variance by anchoring participants to an initial price. In our pilot, we included a condition that featured no price, which resulted in extremely divergent suggested prices.[5] Therefore, we decided to intentionally anchor participants to a default price of $0.99. We also hoped to make participants feel more invested in the application, and therefore more willing to pay attention to its details.

Our survey was short to maximize participation. We asked a total of seven questions on four pages. The first two questions solicited participants' bid amounts and price suggestions. On the second page of the survey, we asked participants to sort several factors that may have influenced their perceptions of the application ("description," "icon," "manufacturer," "permissions requested," "name of application," "cost," and "size of application") from "most influential" to "least influential."

The purpose of the survey was to examine how participants would change their valuations of the application based on whether it was collecting information for secondary purposes. The bids were our primary experimental measure, but we also asked an explicit question about advertising. We told participants:

> We're also considering making a free version of the app. It will have targeted advertisements that will be relevant to you, based on how you use the app and your mobile device. For instance, ads may be selected based on:
>
> - How you use the app (e.g., the specific deals you view)
> - Your address book contacts (e.g., the ads viewed by friends who also use the app)
> - Your location (e.g., the businesses you visit while carrying your smartphone)
> - Your photo library (e.g., activities depicted in your photos)

Each participant saw a subset of the bullets, based on his or her assigned experimental condition. The first bullet was present for every participant because all participants saw the INTERNET permission. The second bullet was displayed for participants who saw the READ_CONTACTS permission, the third bullet was

[5] When we made the price "free," skewness and kurtosis were 8.36 and 71.03, respectively ($n = 159$). Whereas when we set the price to "$0.99," skewness and kurtosis were 1.72 and 5.74 ($n = 163$). This anchoring effect was statistically significant: $U = 10078.5$, $p < 0.0005$, $\mu_{free} = \$2.94$ ($\sigma = 11.09$), $\mu_{\$0.99} = \$1.11$ ($\sigma = 0.57$).

displayed for participants who saw the ACCESS_FINE_LOCATION permission, and the fourth bullet was displayed for participants who saw the PHOTOS permission. Participants who saw all four permission requests also saw all four bullets. We then asked participants to select the version they would be *more* likely to install: the "$0.99 version with no advertisements" or the "free version with targeted advertisements."

We collected survey responses during February 2012. After screening out 26 responses due to obvious cheating, non-Android users, and incomplete responses, we were left with 368 responses. These responses corresponded to 139 females (37.8 % of 368) and 227 males (61.7 %), while two respondents omitted their genders. We observed no statistically significant differences with regard to gender, nor age ($\mu = 29.3$, $\sigma = 8.57$), and therefore did not further analyze demographic factors.

10.4.2 Results

We considered five hypotheses:

H_{0a} = Bids will not change based on the permissions
H_{1a} = Bids will positively correlate with permission requests
H_{0b} = Suggested prices will not change based on the permissions
H_{1b} = Suggested prices will negatively correlate with number of permissions
H_{0c} = Popularity of the ad-supported version will not change with permissions

We noticed several clear outliers for the open-ended bids and suggested prices (e.g., one participant bid $10,000). We compensated for these outliers by excluding every data point above the 95th percentile.[6] Thus, we analyzed 353 responses.

10.4.2.1 Bids as a Proxy for Privacy Concerns

We performed a linear regression between participants' bids and which of the three permissions they were shown (i.e., PHOTOS, ACCESS_FINE_LOCATION, or READ_CONTACTS). Our results were statistically significant, though not to the degree we had expected: READ_CONTACTS was the only permission that had a statistically significant impact on participants' bid amounts (see Table 10.3). The 139 participants who were exposed to this permission demanded significantly more compensation than the remaining 214 ($\mu_0 = \$15.11$, $\mu_1 = \$23.32$, $U = 12900.0$, $p < 0.034$). In summary, we reject H_{0a}, and H_{1a} is supported for contact data.

[6]This corresponded to bids over $100 and suggested prices over $2.99. Prior to removing outliers, the skewness and kurtosis for the bids were 18.65 and 353.15, respectively. After removing outliers, they became 2.15 and 4.10. Regarding the suggested prices, the original skewness and kurtosis were 5.87 and 50.27, but were reduced to 0.63 and 1.79, after removing outliers.

Table 10.3 Regression of participants' bids as a function of which permissions they were shown

	β	t	Sig.
(Constant)		7.435	$p < 0.001$
READ_CONTACTS	0.168	3.104	$p < 0.002$
ACCESS_FINE_LOCATION	0.011	0.215	$p < 0.830$
PHOTOS	−0.027	−0.494	$p < 0.622$
$F_3 = 3.294$, $p < 0.021$			

Table 10.4 The rows depict the five between-subjects permission requests, the average bid amounts, the standard deviations, and the number of participants assigned to each condition

Permission	μ	(σ)	n
INTERNET	$13.03	(13.85)	69
INTERNET ACCESS_FINE_LOCATION	$16.84	(25.14)	81
INTERNET READ_CONTACTS	$24.82	(29.70)	71
INTERNET PHOTOS	$15.17	(17.17)	64
INTERNET ACCESS_FINE_LOCATION READ_CONTACTS PHOTOS	$21.77	(30.32)	68

We did not observe a significant correlation between the suggested prices and participants' bids. This suggests that the anchoring effect of including a price in the screenshot overshadowed any effects from the permission requests; participants reported an average suggested price of $0.98 ($\sigma = 0.56$), which was similar to the $0.99 anchor. Thus, we conclude that the suggested prices were an inadequate proxy for participants' willingness to pay (i.e., H_{0b} cannot be rejected and H_{1b} cannot be accepted). Thus, our focus remains on participants' bids for compensation.

We performed a Pearson correlation between the number of permissions each condition requested and participants' bids and found this to be statistically significant ($r = 0.10$, $p < 0.031$, one-tailed); participants requested more compensation as the application requested more permissions. Table 10.4 lists participants' average bids.

10.4.2.2 Influential Factors

Participants ranked seven factors that influenced their bids:

1. Cost ($\mu = 2.39, \sigma = 1.54$)
2. Description ($\mu = 2.64, \sigma = 1.58$)

3. Name ($\mu = 3.36, \sigma = 1.73$)
4. Icon ($\mu = 4.41, \sigma = 1.85$)
5. Permissions Requested ($\mu = 4.64, \sigma = 1.81$)
6. Size of Application ($\mu = 4.83, \sigma = 1.76$)
7. Manufacturer ($\mu = 5.37, \sigma = 1.59$)

We observed that participants in the control condition ranked the requested permissions as the 5th largest factor that influenced their bids, whereas the other participants ranked it as 4th. While a statistically significant difference ($U = 6227.0$, $p < 0.015$), permissions were not a primary decision factor.

10.4.2.3 Willingness to See Targeted Ads

We asked participants whether they would prefer a free, advertising-supported version of the application instead of the $0.99 version that they previously saw. We indicated that the advertisements would be targeted using data from the granted permissions (e.g., we told participants who saw the `ACCESS_FINE_LOCATION` permission that the advertisements would be location-based). Overall, we found that 22.3 % of our 368 participants indicated a preference for paying $0.99 to avoid advertisements. However, chi-square tests with regard to the specific permissions to which participants were exposed yielded no statistically significant results: we fail to reject H_{0c}. Thus, an aversion to advertising does not appear to be based on what data is collected to support targeted advertising. The corollary to this is that 77.7 % of our participants would prefer advertisements if it meant saving $0.99, regardless of the personal data that is collected and used to target those advertisements.

10.5 Implications

The choice architecture of the smartphone application marketplace can have a profound impact on users' privacy decisions: users weigh privacy more heavily when they can easily compare applications' permission requests. In this section, we explore the various factors that compete for users' attention during application selection and installation. We conclude with suggestions for future work to help improve the Android Market and other smartphone application repositories to better support users' privacy concerns.

10.5.1 Decision Factors

In our first experiment (Sect. 10.3), we observed that there were three factors that influenced participants' application choices: the amount of consideration given

Table 10.5 Regression of participants' price selections based on three factors: whether they believed the ACCESS_FINE_LOCATION permission was appropriate, their perceived importance of permission requests in general, and their perceived importance of price

	β	t	Sig.
(Constant)		16.488	$p < 0.0005$
ACCESS_FINE_LOCATION	−0.140	−3.581	$p < 0.0005$
Permissions	0.315	8.071	$p < 0.0005$
Cost	−0.367	−9.373	$p < 0.0005$
$F_3 = 59.843$, $p < 0.0005$			

to permissions in general, the amount of attention paid to application cost, and whether they believed the ACCESS_FINE_LOCATION permission was relevant to the application's functionality We performed a linear regression with these factors, which we observed to be highly statistically significant (Table 10.5).

The coefficients in Table 10.5 indicate that when participants were able to compare similar applications side-by-side, cost was the primary factor behind their decisions, but the requested permissions were a near second. As a result of these factors, we observed that when the choice architecture allowed them to compare multiple applications of similar functionality, a quarter of participants indicated a willingness to pay a 300 % premium for an application that collected the least data. This effect was pronounced when it was clear that the data was extraneous to an application's core functionality. In this choice architecture, in which participants were able to directly compare permissions between applications, participants who selected the high-privacy variant indicated that the permissions were the primary factor behind their decisions.

The choice architecture that participants encountered in our second experiment (Sect. 10.4) did not allow them to view the permissions of multiple applications. Instead, we displayed varying permission requests and asked participants to indicate their willingness to install the application. We observed that their stated willingness was only correlated with the request for one particular permission, READ_CONTACTS. Participants were not observably less willing to install the application when it requested access to a user's photo library (PHOTOS) or the location reported by onboard GPS hardware (ACCESS_FINE_LOCATION). More importantly, even though only one particular permission triggered privacy concerns, its presence did not force privacy concerns to the forefront of the decision process. Privacy-concerned participants indicated that permissions ranked a meager fourth in terms of the factors they considered, as compared to fifth for participants who were not exposed to extraneous permission requests. Our results suggest that because this choice architecture does not allow participants to contextualize the appropriateness of applications' permission requests, they give less weight to their privacy concerns.

We hypothesize that participants' devaluation of privacy in the second experiment was a result of bounded rationality. In the first experiment, the side-by-side

display of varying permission requests made it obvious to participants that they can directly choose between multiple applications that afford varying levels of privacy, as well as the fact that some applications may not actually *need* some of the requested permissions. Therefore, for the privacy-concerned, they had the option to simply avoid applications that they believed posed a conflict to their privacy preferences. Our data indicate that 25 % of our participants exercised this option. In the second experiment, participants were not exposed to multiple sets of permission requests and therefore even if they believed some permissions may be extraneous to an application's functionality, they may have felt they had no choice but to grant them, because a choice was not apparent to them—though they could have reflected this apprehension in their bids. It was only when the permission requests crossed a particular "privacy threshold" that participants demanded additional financial incentives (i.e., when the `READ_CONTACTS` permission was requested). Thus, they satisficed by accepting extraneous permissions that did not rise to this threshold.

10.5.1.1 Location, Location, Location?

Previous smartphone privacy research has focused on location disclosure. However, our results suggest that the value proposition offered by sharing location data with applications is generally seen as a net positive: participants in our first experiment were more likely to see location requests as a signal for desirable location-aware features than an inappropriate intrusion upon their personal privacy. In our second experiment, participants did not significantly alter their bids in the presence of the `ACCESS_FINE_LOCATION` permission, nor did it prompt an observable change in participants' decisions in the context of behavioral advertising. This speaks to the acceptance of ubiquitous location-aware applications in the marketplace.

We observed that the `RECORD_AUDIO` permission signaled privacy concerns across our first experiment's participants. Examining the Likert data used to report concern levels, we found that the mean concern level over the ability to record audio was significantly higher than concerns over determining location (`RECORD_AUDIO`: $\mu = 4.91, \sigma = 2.23$; `ACCESS_FINE_LOCATION`: $\mu = 4.62$, $\sigma = 2.05$; $Z = -2.69$, $p < 0.007$).

Previous research has suggested that user attention is a finite resource [8, 9]. Therefore, prompting users to approve requests for access to data that does not concern them is likely to increase habituation, which could result in a failure to notice more-concerning requests. Our results suggest that a more effective choice architecture needs to account for the relative levels of concern for the varying permissions.

10.5.1.2 Users and Behavioral Advertising

McDonald and Cranor reported that only 20 % of survey respondents would be interested in targeted advertising and that 11 % would pay to avoid advertisements

altogether [28]. However, in our second experiment, we observed that 77.7 % of our 368 participants (95 % CI: 73.1–81.9 %) stated that they were unwilling to pay $0.99 for an application in order to avoid targeted advertising. This divergence may be due to their study being performed 3 years prior to ours; privacy attitudes may have changed during the interim as users become accustomed to location-based services. Another possibility is that users are more accepting of behavioral advertising when it is part of a much larger value proposition. In their survey, participants were simply asked whether they would prefer it, whereas we specified an exact cost ($0.99).

We were surprised that participants' support for behavioral advertising did not change as a function of the data used to target the advertisements. This may indicate that users take an all-or-nothing approach to behavioral advertising: if forced to look at advertisements, they may want those advertisements to be as relevant as possible. We did not gather data to explicitly test this.

10.5.2 Open Problems and Future Work

Our findings suggest that it may be possible to improve the Android choice architecture to better address users' privacy and security concerns. Our study also suffered from several limitations, which future studies should address.

10.5.2.1 Privacy Annotations

Our results suggest that a choice architecture that allows users to make side-by-side comparisons between similar applications encourages privacy-preserving behavior. Currently, the Android Market does not support this. We believe that such an architecture is in the best interest of all stakeholders. Many users desire greater privacy features and are willing to pay premiums for them. Such premiums are in the interest of the platform owner, as they increase total marketplace revenue.

We expect to empirically test this in the near future in the laboratory and field by modifying participants' smartphones to support similar search result annotations. Given the limited screen real estate, and that many of the current permissions are unconcerning to users, users are likely to comparison shop based on only a small subset of permissions. We are currently designing icons to represent the permissions that users find most concerning. We expect that if these icons appear alongside search results, users will be more likely to install applications that are aligned with their privacy and security preferences, even if this means paying a premium.

10.5.2.2 Runtime Permissioning

In our second experiment, we observed that participants' privacy-preserving behaviors were much more nuanced than in our first experiment: when participants

viewed a single application, their willingness to install it only changed significantly when it requested one particular permission, `READ_CONTACTS`, which previous participants ranked as one of the most concerning. Other extraneous permission requests did not concern participants enough for them to significantly alter their valuations of the application; similar extraneous requests in the first experiment resulted in privacy-concerned participants choosing alternate applications with fewer permission requests. This result suggests that smartphone users are likely to install desirable applications regardless of whether or not they request extraneous permissions, even when these requests conflict with their stated privacy preferences—the desire to install the applications outweighs users' privacy concerns. This may be due in part to hyperbolic discounting, where the immediate desire for the application results in devaluation of future privacy concerns. Regardless of the exact cause, we believe this result points to another limitation of the choice architecture: when participants are evaluating an application's entire value proposition, privacy concerns are but one aspect; many other factors may overshadow even a privacy-conscious user's apprehension to disclose personal data.

The current choice architecture under-values privacy because it frames the choice as one between accepting the privacy risks or not installing the application—without providing alternatives. These shortcomings could be addressed by decoupling the decision of whether or not to install an application from the decision of whether or not to grant it a particular permission. Studies are needed to validate this hypothesis. We expect to perform a field study using a modified version of the Android OS to examine whether or not users make the same decisions regarding whether or not to grant applications permissions when those permissions are requested at runtime, when the data is actually needed by the application. Such a choice architecture will need to consider many factors: for example, how often to prompt the user for particular types of data, why the data is being requested, and how to phrase the requests.

10.5.2.3 Limitations

Both of our studies were based on users' stated preferences, rather than observing their actions in the Market. That is, unlike real interactions with the Market, where users are paying actual money and disclosing their actual personal data, our study did not expose them to these costs. At the same time, in our second experiment, we led users to believe that they would be paying for and installing an actual application, and they had no reason to disbelieve us. In fact, our statistically significant results suggest that they weighed these costs in their decisions.

Furthermore, our first experiment displayed four applications side-by-side, which allowed participants to directly compare the full set of permissions requested between all four applications. Due to screen size limitations on most mobile devices, this scenario is a best case for the ability to do side-by-side comparisons. Therefore our survey results likely represent upper bounds.

In both of our experiments, we exposed participants to varying permission requests. However, we only collected data on their hypothetical behaviors for five

different permissions. These permissions represent just 4 % of the 124 permissions available in the most recent release of Android [5]. It is likely that there are many permissions that users find even more concerning than the ones we examined in this study, as well as many more that users universally find unconcerning. An improved choice architecture would need to account for the full spectrum of permissions. Likewise, because of this and the aforementioned limitations stemming from the realism of the tasks, we cannot make generalizations about how much users may be willing to pay to avoid granting particular permissions.

Finally, another limitation of our study was that the results from our first experiment cannot be quantitatively compared with the results from our second experiment, since they used different metrics and were performed over different periods of time (not to mention that they involved completely different methodologies). Instead, we qualitatively compare the results of the two studies to show how changes to the choice architecture can have profound impacts on users' decisions. In future work, we expect to directly address these limitations by conducting laboratory and field experiments wherein participants face real financial and privacy risks.

Acknowledgements The authors would like to thank Jaeyeon Jung and Stuart Schechter for their feedback. This work was supported by Intel, through the ISTC for Secure Computing. The co-author Adrienne Porter Felt was affiliated with the University of California, Berkeley, at the time of this work.

References

1. Acquisti, A.: Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the ACM Electronic Commerce Conference (EC '04), New York. ACM, New York (2004)
2. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Secur. Priv. **3**(1), 26–33 (2005). http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1392696
3. Acquisti, A., John, L., Loewenstein, G.: What is privacy worth? In: Twenty First Workshop on Information Systems and Economics (WISE), Phoenix (2009)
4. Agele, M., Kruegel, C., Kirda, E., Vigna, G.: Pios: detecting privacy leaks in iOS applications. In: Proceedings of the Network and Distributed System Security Symposium (NDSS), San Diego (2011)
5. Android Developers: Manifest.permission. http://developer.android.com/reference/android/Manifest.permission.html. Accessed 28 Dec 2011
6. Barkhuus, L.: Privacy in location-based services, concern vs. coolness. In: Workshop on Location System Privacy and Control at MobileHCI '04, Glasgow (2004)
7. Barkhuus, L., Dey, A.: Location-based services for mobile telephony: a study of users' privacy concerns. In: INTERACT'03, Zurich, pp. 702–712 (2003)
8. Beautement, A., Sasse, M.A., Wonham, M.: The compliance budget: managing security behaviour in organisations. In: Proceedings of the 2008 Workshop on New Security Paradigms, NSPW '08, Lake Tahoe, pp. 47–58. ACM, New York (2008)
9. Böhme, R., Grossklags, J.: The security cost of cheap user interaction. In: Proceedings of the 2011 New Security Paradigms Workshop (NSPW), Marin County. ACM, New York (2011)
10. Chia, P.H., Yamamoto, Y., Asokan, N.: Is this App safe? A large scale study on application permissions and risk signals. In: World Wide Web Conference, Lyon (2012)

11. Consolvo, S., Smith, I.E., Matthews, T., LaMarca, A., Tabert, J., Powledge, P.: Location disclosure to social relations: why, when, & what people want to share. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI '05, Portland. ACM, New York (2005)
12. Cvrcek, D., Kumpost, M., Matyas, V., Danezis, G.: A study on the value of location privacy. In: Proceedings of the 2006 Workshop on Privacy in an Electronic Society (WPES'06), Alexandria (2006)
13. Danezis, G., Lewis, S., Anderson, R.: How much is location privacy worth? In: Proceedings of the Workshop on the Economics of Information Security (WEIS 2005), Cambridge (2005)
14. Egelman, S., Tsai, J., Cranor, L.F., Acquisti, A.: Timing is everything? The effects of timing and placement of online privacy indicators. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI '09, Boston. ACM, New York (2009)
15. Enck, W., Ongtang, M., McDaniel, P.: On lightweight mobile phone application certification. In: Proceedings of the 16th ACM Conference on Computer and Communications Security (CCS), Chicago. ACM, New York (2009)
16. Enck, W., Gilbert, P., Chun, B.G., Cox, L.P., Jung, J., McDaniel, P., Sheth, A.N.: Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones. In: Proceedings of the 9th USENIX Conference on Operating Systems Design and Implementation, OSDI'10, Vancouver. USENIX Association, Berkeley (2010)
17. Enck, W., Octeau, D., McDaniel, P., Chaudhuri, S.: A study of Android application security. In: Proceedings of the 20th USENIX Security Conference USENIX Association, Berkeley (2011)
18. Felt, A.P., Greenwood, K., Wagner, D.: The effectiveness of application permissions. In: Proceedings of the 2nd USENIX Conference on Web Application Development, WebApps'11, Portland, pp. 7–7. USENIX Association, Berkeley (2011)
19. Felt, A.P., Egelman, S., Wagner, D.: I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In: Proceedings of the ACM CCS Workshop on Security and Privacy in Smartphones and Mobile Devices (SPSM), Raleigh (2012)
20. Felt, A.P., Ha, E., Egelman, S., Haney, A., Chin, E., Wagner, D.: Android permissions: user attention, comprehension, and behavior. In: Proceedings of the 2012 Symposium on Usable Privacy and Security (SOUPS), Washington, DC (2012)
21. Gideon, J., Egelman, S., Cranor, L., Acquisti, A.: Power strips, prophylactics, and privacy, oh my! In: Proceedings of the 2006 Symposium on Usable Privacy and Security, Pittsburgh (2006)
22. Good, N., Dhamija, R., Grossklags, J., Aronovitz, S., Thaw, D., Mulligan, D., Konstan, J.: Stopping spyware at the gate: a user study of privacy, notice and spyware. In: Proceedings of the Symposium on Usable Privacy and Security (SOUPS 2005), Pittsburgh (2005)
23. Grossklags, J., Acquisti, A.: When 25 cents is too much: an experiment on willingness-to-sell and willingness-to-protect personal information. In: Proceedings (online) of the Sixth Workshop on Economics of Information Security (WEIS), Pittsburgh (2007)
24. Hornyack, P., Han, S., Jung, J., Schechter, S., Wetherall, D.: These aren't the droids you're looking for: retrofitting Android to protect data from imperious applications. In: Proceedings of the 18th ACM Conference on Computer and Communications Security (CCS), Chicago. ACM, New York (2011)
25. Huberman, B., Adar, E., Fine, L.: Valuating privacy. IEEE Secur. Priv. **3**(5), 22–25 (2005)
26. Iachello, G., Smith, I., Consolvo, S., Chen, M., Abowd, G.D.: Developing privacy guidelines for social location disclosure applications and services. In: Proceedings of the 2005 Symposium on Usable Privacy and Security, SOUPS '05, Pittsburgh, pp. 65–76. ACM, New York (2005)
27. Lederer, S., Mankoff, J., Dey, A.K.: Who wants to know what when? Privacy preference determinants in ubiquitous computing. In: CHI '03 Extended Abstracts on Human Factors in Computing Systems, CHI EA '03, Ft. Lauderdale, pp. 724–725. ACM, New York (2003)
28. McDonald, A.M., Cranor, L.F.: Beliefs and behaviors: internet users' understanding of behavioral advertising. In: 38th Research Conference on Communication, Information and Internet Policy (Telecommunications Policy Research Conference), Arlington (2010)

29. Nauman, M., Khan, S., Zhang, X.: Apex: extending Android permission model and enforcement with user-defined runtime constraints. In: Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS), Beijing. ACM, New York (2010)
30. Pearce, P., Felt, A.P., Nunez, G., Wagner, D.: Addroid: privilege separation for applications and advertisers in Android. In: Proceedings of the ACM Symposium on Information, Computer and Communications Security (ASIACCS), Seoul. ACM, New York (2012)
31. Purcell, K.: Half of adult cell phone owners have apps on their phones. Pew Internet & American Life Project. http://pewinternet.org/Reports/2011/Apps-update.aspx (2011)
32. Simonite, T.: Apple ignored warning on address-book access. Technology Review (MIT). http://www.technologyreview.com/communications/39746/ (2012)
33. Spiekermann, S., Grossklags, J., Berendt, B.: E-privacy in 2nd generation E-commerce: privacy preferences versus actual behavior. In: Proceedings of EC'01: Third ACM Conference on Electronic Commerce, Tampa, pp. 38–47 (2001)
34. Thaler, R., Sunstein, C.: Nudge: Improving Decisions About Health, Wealth, and Happiness. Yale University Press, New Haven/London (2008)
35. Tsai, J., Egelman, S., Cranor, L., Acquisti, A.: The effect of online privacy information on purchasing behavior: an experimental study. In: Proceedings of the 2007 Workshop on the Economics of Information Security (WEIS'07), Pittsburgh (2007)
36. Westin, A.F.: E-Commerce & Privacy: What Net Users Want. Privacy & American Business, Hackensack (1998)
37. Wiese, J., Kelley, P.G., Cranor, L.F., Dabbish, L., Hong, J.I., Zimmerman, J.: Are you close with me? Are you nearby? Investigating social groups, closeness, and willingness to share. In: Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp '11, Beijing, pp. 197–206. ACM, New York (2011)
38. Zickuhr, K.: Generations and their gadgets. http://pewinternet.org/Reports/2011/Generations-and-gadgets/Report/Cell-phones.aspx (2011). Accessed 2 Oct 2012

Chapter 11
Would You Sell Your Mother's Data? Personal Data Disclosure in a Simulated Credit Card Application

Miguel Malheiros, Sacha Brostoff, Charlene Jennett, and M. Angela Sasse

Abstract To assess the risk of a loan applicant defaulting, lenders feed applicants' data into credit scoring algorithms. They are always looking to improve the effectiveness of their predictions, which means improving the algorithms and/or collecting different data. Research on financial behavior found that elements of a person's family history and social ties can be good predictors of financial responsibility and control. Our study investigated how loan applicants applying for a credit card would respond to questions such as "*Did any of your loved ones die while you were growing up?*" 48 participants were asked to complete a new type of credit card application form containing such requests as part of a "Consumer Acceptance Test" of a credit card with lower interest rates, but only available to "financially responsible customers." This was a double-blind study—the experimenters processing participants were told exactly the same. We found that: (1) more sensitive items are disclosed less often—e.g., friends' names and contact had only a 69 % answer rate; (2) privacy fundamentalists are 5.6 times less likely to disclose data; and (3) providing a justification for a question has no effect on its answer rate. Discrepancies between acceptability and disclosure were observed—e.g., 43 % provided names and contact of friends, having said they found the question unacceptable. We conclude that collecting data items not traditionally seen as relevant could be made acceptable if lenders can credibly establish relevance, and assure applicants they will be assessed fairly. More research needs to be done on how to best communicate these qualities.

M. Malheiros (✉) · S. Brostoff · C. Jennett · M.A. Sasse
University College London, Gower Street, London WC1E 6BT, UK
e-mail: m.malheiros@cs.ucl.ac.uk; s.brostoff@cs.ucl.ac.uk; c.jennett@cs.ucl.ac.uk

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_11,

11.1 Introduction

To lend money responsibly, as well as protect their own business, lenders assess the risk of applicants not repaying their loans. For the assessment process, lenders collect personal data items directly from applicants, and from organizations such as credit reference agencies, and feed the data collected into their credit scoring algorithms. The lenders will reject loan requests from applicants who fall above a certain risk threshold. The goal is to ensure that the lending business remains profitable, but it also prevents applicants who would not be able to afford the loan from getting into financial hardship.

Lenders are continuously looking to improve the accuracy of their risk assessments, either by improving the algorithms used, or by collecting new types of data. Based on the literature on credit scoring and interviews with experts in personal finance and credit risk, we identified factors that seem to be associated with financial behavior, but are not widely used (and if they are used, the general public is not aware of it). These include a person's relationship with parents while growing up [22, 42], social links [13], and bill payment history [5, 43] among others. Such data is clearly sensitive, but using it in this way is no different from how health data is used by insurance companies, and psychometric and drug tests data by some companies to assess job applicants. But such data could also be beneficial for some loan applicants: new types of data with predictive value could help those with "thin" credit histories, who currently find themselves excluded from many financial services because they cannot prove their creditworthiness.

We first review the literature on credit scoring, and present results from interviews with experts in personal finance and credit risk; we then discuss factors known to influence privacy perceptions of individuals. We then present a study in which participants were asked to complete a credit card application in which they had to disclose data commonly requested in this process, and some alternative data. The results from the experiment and the post-experiment questionnaire show that providing justifications for questions has no effect on disclosure rates. Surprisingly, participants did disclose some data they rated as "*unacceptable for lenders to request*", but were less likely to disclose such information about people other than themselves. We conclude that lenders should avoid collecting indices of social capital for the time being, and should keep in mind the potential mismatch between the perceived relevance of a data request and its actual relevance in an empirically based credit scoring algorithm.

11.2 Background

11.2.1 Credit Scoring

Credit can be a force for good: it can be an investment—for example, buying a car might enable someone to obtain a job which they otherwise might not be able to

get to, or it can help to manage unexpected expenses, such as emergency repairs. However, individuals obtaining loans they cannot repay has serious consequences on their lives, as well as the lenders' balance sheets: in the UK, for instance, 331 people are declared insolvent or bankrupt every day [11].

To minimize the number of loan defaults and maximize profit (not giving a loan to applicants who could repay it equals lost profit), lenders assess the likelihood that an applicant will repay a loan. This process is known as credit scoring, first used in the 1940s, when it relied on human judgement: credit analysts read an application form and made a decision based on the 5 C's [46]: "the *character* of the person (do you know the person or their family?); the *capital* (how much is being asked for?); the *collateral* (what is the applicant willing to put up from their own resources?); the *capacity* (what is their repaying ability. How much free income do they have?); the *condition* (what are the conditions in the market?)".

Today, credit scoring is based on automatic statistical algorithms which are fed data from the applicant's application form, data related to past dealings with the lender, and their credit report—obtained from a credit bureau (see, for example, [44]). The risk of an applicant defaulting is inferred from the performance of borrowers whose data profile is similar [10, 25]. Credit scoring algorithms are faster, more consistent and less prejudiced than human decision makers, and there is evidence that these algorithms are better predictors of which applicants would be "good" or "bad" customers [46].

But credit scoring algorithms are not perfect. Mistakes occur in the classification of some applicants (good risks classified as bad risks and vice-versa), because of limitations in the building of the algorithms themselves (data used to develop predictive models sometimes has poor quality and is based only on samples of accepted borrowers [15]), interactions between variables that become outdated (people's behavior changes over time), and because some factors that are the cause of bankruptcy are difficult to predict—e.g., divorce, health problems or unemployment (Expert 1, 2010, Discussion on consumer finance statistics, personal communication with university professor with background in consumer finance statistics research) [25]. To improve the accuracy of their credit scoring, lenders can improve the way their algorithms are built—by adjusting how variables are transformed—or by collecting more data. The latter is seen as a more promising approach because the statistical methods underlying credit scoring are well understood, and no imminent breakthroughs in improving their performance are expected (Expert 1, 2010, Discussion on consumer finance statistics, personal communication with university professor with background in consumer finance statistics research).

11.2.2 Alternative Indicators

Based on our review of literature on personal credit (Brostoff et al., Privacy value networks project: financial services case study. Technical report, unpublished), and interviews with experts on credit risk and financial behavior, we identified several

types of data that are potential indicators of financial behavior, but which are not currently requested in loan applications. These types of data include: bill payments (other than utilities), tax payments, employer recommendations, health condition, stability in life, and social relationships.

Utility payments, for example, are considered to be a measure of willingness to pay debts. There have been initiatives in the US for applicants with no traditional credit history to use their history of utility payments as a measure of their willingness to pay, and these data have been incorporated into credit report products offered by mainstream credit reference agencies—for example the "PRBC credit report with FICO expansion score" from Fair Isaac. Some utility payments are now part of the UK credit bureau data, but it is not clear whether applicants realize this, or how they would perceive an explicit request for this data. Data such as TV license payments are not yet collected, and it is not clear to what extent applicants consider them to be utilities (as opposed to less socially acceptable categories of expenditure), and how this personal classification might be reflected in perceptions of requests for the data.

The same applies to accommodation-related payments. Rent [43] and Council Tax payments indicate that the applicant makes regular payments and demonstrates responsible behavior. A larger number of insurance claims might also indicate that they are a riskier person (resulting in higher insurance premiums); too many may indicate a propensity for fraud.

Sometimes employers vouch for new employees, so that they can get bank accounts (Expert 2, 2010, Discussion on financial behavior of immigrant populations in London, personal communication with an academic specializing in migration and immigration). A recommendation from the employer could therefore function as signal for creditworthiness.

Health condition may also be linked to ability to repay. Body-mass index (BMI), for example, has been linked to some aspects of self-control [27], which can be perceived as being related to ability to pay. Also, some lenders purchase insurance to recover the loan in the case of the borrower's death, and these policies require declarations of health and pre-existing conditions. Health checks can reveal lifestyle choices that correlate with responsibility and self-control, and ability to pay back loans. Moreover, mental illness, disability, and physical illness are large risk factors for borrowers not paying back debts (Expert 3, 2009, Discussion on peer-to-peer lending, personal communication with executive from a peer-to-peer lending company).

Stability in applicants' lives is a key predictor for creditworthiness. One way to assess stability is by asking whether the applicant lives with a partner or spouse (Expert 4, 2009, Discussion on risk management in lending, personal communication with a risk management consultant for a financial services authority). Kirchler et al. [29] suggest that relationship dynamics can have an impact on credit decisions, with mutual social influence of the partners potentially changing their behavior.

Stability and attitudes to money are also corrected with experiences while growing up. Analysis of case studies of over-spenders found that these often have

a family background where money was used as a method of control, where the relationships with fathers were problematic, distant and mediated by money [42], and where the patient had experienced major and unresolved loss [22].

In a study that examined the performance of listings in a peer-to-peer lending service (*Prosper*[1]), the structural component "degree centrality" of the applicant's social network was related to their probability of being granted a loan: applicants who had more friends and were more central in their social networks were more likely to receive loans. Lin et al. [33] found that the number and type of friends an applicant had was related to how likely they were to receive a loan—with likelihood increasing with the number of friends who were lenders on *Prosper*. Friend lists may therefore be used as a way of estimating social capital—if an applicant has friends who are rich, powerful and trustworthy, then s/he is seen as trustworthy and less risky to lend to. It is also seen to assist fraud prevention because such connections facilitate tracing of a defaulting borrower who has changed address. Similarly, the names, addresses and phone numbers of people that know the applicant well could be obtained. This is already done by some sub-prime lenders [26]. Although Lin et al. [33] did not study it, it is plausible that some measure of message flow between an applicant and their social network is an indication of the strength of ties between that applicant and their network, and so could be used as an index of social capital, and therefore trustworthiness to receive loans.

11.2.3 Privacy Factors

Even though the data items discussed above could potentially improve the assessment, their use by lenders raises the number of questions. The key one—which we address in this study—is whether requesting them would raise privacy concerns. Past research has identified three criteria that are like to impact applicants' privacy perceptions: sensitivity, transparency, and privacy values.

11.2.3.1 Sensitivity

Adams and Sasse [2] investigated privacy perceptions from a user-centric perspective, and found that users' assessment of privacy risks depends on three main factors: (1) information receiver; (2) information usage; and (3) information sensitivity. The first factor refers to how much the user trusts the person or people who will have access to their data. The second factor addresses the way users think receivers use their data in the present, and are going to use it in the future. When individuals perceive that they have some degree of control over future usage of their personal data, they react in a more positive manner to its collection [12].

[1] www.prosper.com

The third factor consists of the users' perceptions of the data being disclosed and how others (e.g., the receivers) will interpret it. Believing that data portrays individuals in a fair and accurate manner is an important acceptance factor—from a privacy perspective—of technologies and processes that collect personal data [12, 34]. Metzger [36] investigated the effect of sensitivity on disclosure and found individuals were more likely to withhold items they found more sensitive.

We believe that the different data sensitivities of the various items requested will have an impact on disclosure rates. Consequently, we propose that:

H1: The proportion of participants disclosing each data item will be correlated with the sensitivity of the data items.

11.2.3.2 Transparency

Relevance or legitimacy of the data request in the context of the interaction has also been identified as an important privacy factor [12, 19]. Annacker et al. [3] identify legitimacy of a data request as a significant driver for privacy costs, i.e., the lower the perceived legitimacy of the data request the more privacy individuals felt they were giving away. Drawing from the concept of "contextual integrity" (see [38]) O'Hara and Shadbolt [39] describe examples in which there is a negative reaction to a type of data request in one context, but not another: e.g., collecting data about one's marital status may be appropriate during a date, but is inappropriate in the context of a job interview. In a previous study, Jennett et al. [24] suggested that transparency of purpose of data requests, in the context of credit applications, could make individuals feel more comfortable with answering questions. Thus, in the current study we advance the following hypothesis:

H2: Participants will disclose more data when a reason for the data request is given, compared to when no reason is given.

11.2.3.3 Privacy Values

Individual differences may also contribute to different privacy perceptions of specific data requests: some individuals are more sensitive to privacy issues than others. There have been several attempts to develop ways to measure privacy concern (see [8] for a review). One of the most widely used privacy scales is Westin Privacy segmentation [48], which requires participants to rate three statements on a 4-level scale. Based on their answers participants are assigned to one of three groups: (1) privacy fundamentalists, who have strong feelings about privacy and are very defensive of their personal data; (2) privacy unconcerned, who do not have many concerns about privacy or disclosing personal data; and (3) privacy pragmatists, the majority of people, who are willing to disclose personal data when they see a legitimate use for it and see the benefits of doing so [45]. In our study,

we expect participants categorized as privacy fundamentalists to be more protective of their personal data, therefore, our third hypothesis states that:

H3: Privacy fundamentalists will disclose less data than privacy unconcerned or privacy pragmatists.

11.2.4 Privacy Attitudes vs. Privacy Behavior

Privacy research has identified a discrepancy between stated privacy attitudes and concern and actual disclosure behavior (see [1] and [6]). Most privacy research has relied on data collection techniques such as questionnaires and interviews to capture privacy perceptions and attitudes. In the past two decades, several surveys have identified privacy as a serious concern for consumers and citizens in general [9, 28, 32, 40]; yet there are many documented examples of individuals surrendering their personal data for seemingly small rewards [23, 30]. Thus, it is important to observe how people act in situations where they are confronted with real trade-offs involving their personal data, rather than just ask them about hypothetical scenarios. Our study explores the difference between the stated acceptability of some questions, and the actual disclosure behavior of the same participants on those questions.

Actual privacy behavior is guided by cost-benefit considerations. When organizations providing a service request personal data from individuals, these assess the potential economic or social benefits that will result from the exchange, and weigh them against the costs of providing the data [37, 41]. If the benefits are perceived to outweigh the costs, individuals will agree to the exchange; if they do not, they will withhold or falsify data to reduce the privacy costs, while still obtaining the benefits of the exchange [20, 36].

Some studies have investigated disclosure behavior when economic rewards (such as money, future convenience or time savings) are offered in exchange for personal data [14, 17, 18, 21]. Results indicate that there is a point—albeit variable from context to context—at which individuals will trade their data for material benefits. When individuals apply for credit there is also a potential economic reward that can be obtained through the disclosure of personal data. However, to our knowledge, no empirical research has been conducted on privacy perceptions and decision-making in the context of credit application forms. It is not clear whether privacy decision-making when individuals apply for credit follows the same rules as in other contexts. The research described here tries to address this gap in the literature by simulating an application process for a credit card that requires different types of personal data to be disclosed.

In the following section we describe our experimental design. We start by describing a preliminary survey aimed at collecting sensitivity ratings for several data items which are not currently used in risk assessment but which experts believe may be associated with financial behavior. We then describe our main experimental study which investigated participants' disclosure behavior in the context of a simulated credit card application.

11.3 Preliminary Survey

11.3.1 Demographics

A UK nationally representative sample of 285 participants answered the survey. One hundred eighty-one (63.5 %) were female and 104 male (36.5 %). Forty-five (15.8 %) were between 18 and 24 years old, 36 (12.6 %) between 25 and 39 years old, 100 (35.1 %) between 40 and 59, 104 were 60 years old or over.

11.3.2 Survey

We generated 53 hypothetical questions which are thought to have relevance for assessing creditworthiness, but which are not normally collected in loan application processes. These include "internet payment history", "any insurance claims", "list of friends from your social networking sites". For each item, participants were asked to rate on a 5-point scale to what extent they were comfortable with giving a lender this data.

After an initial principal components analysis (PCA) with Cattell's scree plot method, we identified 5 main factors that the 53 questions varied on. These 5 factors accounted for 57 % of the total variance. The varimax rotation provided a far more interpretable solution than the direct oblimin rotation. Therefore the varimax rotation was interpreted. The five factors produced were seen to have common themes in the items they contained and as such were given the tags: (1) personal/sensitive, (2) bills, (3) attitudes, (4) social network, and (5) partners and children. We selected 14 items for use in the experimental study (see Table 11.2 below) that were representative of these factors—but that could also be changed into a question that could be "responded to" by a participant.

11.4 Experimental Study

11.4.1 Demographics

There were 48 participants in the study. Ages ranged from 19 to 31 years, average age 20 years old (s = 1.97). Thirty five (72.9 %) participants were female and 13 (27.1 %) were male. Thirty six (75 %) participants were UCL psychology students; 8 (16.7 %) were students in other degrees at UCL; 2 (4.2 %) were students at other universities; and 1 (2.1 %) was not a student.

11.4.2 Procedure

Participants were told that they would be helping to test "*the acceptability of the application process for a new Super Credit Card that beats all other cards on the market. Because the deal is so good it can only be offered to people who are very reliable at repaying. The bank (we cannot reveal which one because of commercial sensitivity) thinks it has discovered a better way of assessing financial responsibility, but it requires more and also different information than is used in the standard credit reference reports.*"

The application process consisted of an online application form with 24 questions. Participants were asked to complete and submit the form. They could submit once they had answered at least 20 out of the 24 questions, and were paid £5 (approx. $8) regardless of whether they were able to submit the form or not. Participants were told that no actual credit card would be awarded, but that the person who was found to be most creditworthy would receive a £50 (approx. $80) prize. One factor that could potentially affect the way personal data disclosure decisions are made in the context of credit applications is the large value of the credit service being offered compared to the privacy cost of disclosing sensitive data. Thus, this reward was meant to create a real trade-off between disclosing personal data and obtaining an economic benefit similar to what happens in real life credit applications.

To disincentivize submission of false data, participants were told that "*the card can only be offered to people that are completely honest during the application procedure, if you lie on a single item you are not eligible. [. . .] all application data is being sent to a credit reference agency for validation. . . [using a] . . . sophisticated combination of cross-comparisons between data in the application form, the individual's current credit record, and also comparison to the Agency's most advance customer profiling system.*" Again, the goal was to simulate as realistically as possible a real application process for obtaining credit.

After filling in the application form, participants answered Westin's privacy segmentation questions, and were interviewed about the acceptability of the form's questions and whether they had engaged in any privacy protection behaviors (such as lying). Participants were told that this questionnaire and interview were not part of the evaluation of the bank's application form, but instead part of the research group's own investigation into the acceptability of the data requests. They were further reassured that the experimenters would not share the interview data with the bank.

To prevent bias, the study was conducted "double-blind". The experimenters who processed the participants—three psychology students—were told the same story as participants. The experimenters were told that the research group was conducting a consumer acceptance trial of the new application process for the bank, and also wanted to determine if people would be inclined to lie on those forms.

The study design was submitted to the university's ethics approval process, and received approval before the study commenced. After the study participants had

Table 11.1 List of *Basic* items

	Items
1.	Full name
2.	Gender
3.	Date of birth
4.	Current home address
5.	Mobile phone number
6.	Home phone number
7.	Nationality
8.	Employment status
9.	Have you had a credit card before?
10.	What is the name of your bank?

been processed, experimenters and participants were informed (face-to-face and by email respectively) that the bank did not really exist. The £50 reward was given to a participant selected at random out of those who did submit.

11.4.2.1 Application Form

The application form in the current study began with 10 *Basic* questionnaire items that are present on existing credit application forms (see Table 11.1). These were included to make participants believe that the data was really going to be checked against credit reference agency data, and be used by the bank to identify them. We also assumed that—given how the study was advertised—participants would expect that they had to provide these items—giving a baseline to compare the more sensitive items with.

These were followed by the 14 *Novel* items (see Table 11.2). Responses were either textual data, or required the participant to tick a box to state that he/she consented to his/her data being looked up by the bank.

A progress bar was set up so that participants had to give a certain amount of data before they could submit their application to the bank. We deliberately set the threshold high (20/24) to increase the likelihood of participants providing some *Novel* data items: even if participants gave all *Basic* data (10 items), they needed to provide 10 of the 14 *Novel* items. If participants tried to submit their answers before the progress bar reached 100 %, they received an error message (see Fig. 11.1). If participants chose "not applicable" (N/A), this did not contribute towards the tally, since in a real credit application, an applicant would have to submit alternative data items if s/he was unable to answer a question.

This was part of the deception: when a participant clicked "submit", their data was not sent anywhere, but deleted instead, i.e., no record of the content of participants' responses was kept. Instead, experimenters' kept notes on which questions participants answered in the form. Experimenters did record audio of the post-scenario interviews for later analysis.

Table 11.2 List of *Novel* items

	Items
1.	Did any of your loved ones die while you were growing up? Please give their relation to you (e.g., mother, brother, friend, etc.)
2.	Do you suffer from any medical conditions? Please list...
3.	Did you live with both your mother and father while you were growing up?
4.	Could you list the names and either phone numbers or email addresses of three of your closest friends?
5.	Do you give us permission to contact your local council to get a copy of your council tax payment history?
6.	Do you give us permission to obtain a copy of your TV licence payment history?
7.	Do you give us permission to obtain a copy of your gas or electricity payment history?
8.	Please provide the name and address (or other contact details) of a previous employer so that we can request a copy of the last recommendation from him/her about you...
9.	What is the job of your partner/spouse? Please describe...
10.	What are the names of three people that you are friends with on a social networking site (Facebook, Twitter) whose profiles you would be happy to share with us? Please list...
11.	What are the names of three people that you are friends with on a professional networking site (LinkedIn, Orkut) whose profiles you would be happy to share with us? Please list...
12.	Will you allow us to measure the typical number and length of messages between you and your friends on social networking sites?
13.	What is the length of the longest relationship you have had with a partner/spouse? (years/months/weeks)
14.	May we obtain a copy of your insurance claims (e.g., car, house)?

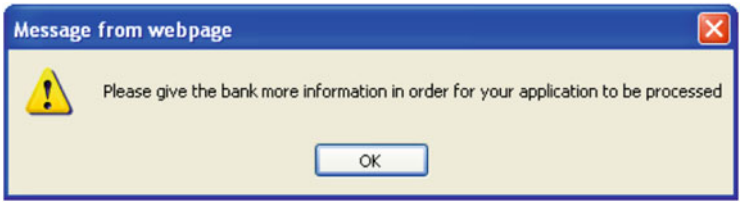

Fig. 11.1 Insufficient information error message

11.4.2.2 Different Versions of the Application Form

Past research suggests that individuals are more comfortable with disclosing personal data when they understand and agree with the purpose of its collection and usage (e.g., [12, 19], or [2]). To test this, we set up two versions of the form:

- *Explanation* condition: Participants were given a brief explanation of why each item was needed by the bank (small text that was presented below the item)
- *No Explanation* condition: Participants were not given an explanation of why each item was needed by the bank.

For example, for half of the participants the question "*Did any of your loved ones die while you were growing up?*" was accompanied by the following explanation: "*We need this information to help judge how your early experiences might shape your behavior as an adult—early loss has been related to later financial behavior.*"

For each of these conditions, we created a *Normal Order* version and a *Reverse Order* version to control for item order. In both versions the 10 *Basic* items were always presented first. In the normal order the *Novel* items were presented as above in Table 11.2—and in the reverse order the *Novel* items were presented in reverse.

11.4.2.3 Privacy Values Questionnaire and Follow-up Interview

As noted in Sect. 11.2, there is evidence that some people are more privacy-sensitive than others. Thus, as well as controlling for *age* and *gender*, we also collected level of privacy concern as assessed by the Westin privacy segmentation [48]. In the Westin scale participants are asked to rate three statements on a 4-point Likert type scale, where 1 = strongly disagree and 4 = strongly agree. The three statements are:

- Consumers have lost all control over how personal information is collected and used by companies.
- Most businesses handle the personal information they collect about consumers in a proper and confidential way.
- Existing laws and organizational practices provide a reasonable level of protection for consumer privacy today.

Privacy fundamentalists are respondents who agreed (strongly or somewhat) with the first statement and disagreed (strongly or somewhat) with the second and third statements. *Privacy unconcerned* are respondents who disagreed with the first statement and agreed with the second and third statements. All other respondents are categorized as *privacy pragmatists*.

A short interview followed, where participants were invited to discuss the acceptability of each of the 24 questions in the application form. If they had decided not to submit the form they were asked about their reasons. They were also asked about whether they had lied on or exaggerated any of their answers.

11.5 Results

11.5.1 Submission and Answer Rates

Twenty eight (58.3 %) participants submitted the application form, which means they answered at least 20 questions out of the 24 asked. All participants answered at

Table 11.3 Answer rates

Item	N	Answer	No answer	N/A	%answer	%answer (excl. N/A)
Grew up with both mother and father	48	48	0	0	100	100
Current home address	48	48	0	0	100	100
Employment status	48	48	0	0	100	100
Gender	48	48	0	0	100	100
Mobile phone number	48	48	0	0	100	100
Nationality	48	48	0	0	100	100
Full name	48	48	0	0	100	100
Date of birth	48	47	1	0	97.9	97.9
Ever had a credit card	48	47	1	0	97.9	97.9
Loved ones passed away while growing up	48	45	3	0	93.8	93.8
Name of your bank	48	45	1	2	93.8	97.8
Copy of TV licence payment history	48	28	1	19	58.3	96.6
Medical conditions	48	45	3	0	93.8	93.8
Copy of gas/electricity payment history	48	38	3	7	79.2	92.7
Home phone number	48	24	2	22	50.0	92.3
Length of longest relationship	48	34	3	11	70.8	91.9
Copy of council tax payment history	48	24	3	21	50.0	88.9
Previous employer contact details	48	26	4	18	54.2	86.7
Social networking profiles of three friends	48	37	6	5	77.1	86.0
Copy of insurance claims	48	23	4	21	47.9	85.2
Job of partner/spouse	48	17	3	28	35.4	85.0
Number and length of mobile text messages	48	33	13	2	68.8	71.7
Name and phone number/email of three friends	48	33	15	0	68.8	68.8
Professional networking profiles of three friends	48	4	5	39	8.3	44.4

least one question, however, even participants who did not submit the form, and the answer rate across all participants for each question is shown below in Table 11.3.

Of the 10 *Basic* information items, 6 were answered by all participants for whom they were applicable (100 %), 3 were not answered by one participant each for whom they were applicable (98 %), 1 was not answered by two participants for whom it was applicable (92 %), giving an average answer rate of 99 %.

The answer rate for *Novel* items was lower, averaging 85 % and ranging from 100 to 44 % of participants answering data items that applied to them. Only one of the *Novel* data items was answered by all respondents—"*Grew up with both mother and father*".

11.5.2 Testing the Hypotheses

Hypothesis 1 predicted that the answer rate for each data item request would be correlated with the sensitivity of the item as measured in the preliminary survey (see Sect. 11.3). This hypothesis was supported: the percentage of participants who answered an item (excluding N/A answers) was significantly correlated with the sensitivity of that item (measured on a 5-point comfort scale) $\rho = 0.624$, $p < 0.01$.

Hypothesis 2 stated that participants would be more willing to disclose personal data in the version of the form where a justification was given for each question. The data did not support this hypothesis. There was no association between the presence of explanations for the questions and whether participants submitted the form or not: $\chi^2(1) = 0.34$, which is below the critical value of 3.84 ($p = 0.05$). There was also no association between the presence of explanations and the number of questions participants answered: t value was not significant ($p = 0.05$). Finally, there was no association between the presence of explanations and whether participants had answered a particular question: Pearson's Chi Square or Fisher's Exact Tests were conducted for each item and none were significant ($p = 0.05$).

Hypothesis 3 stated that participants categorized as *privacy fundamentalists* according to Westin's privacy scale would be less willing to disclose data. This hypothesis was supported by the data, but only when *privacy unconcerned* and *privacy pragmatists* were blocked. When comparing the behavior of privacy fundamentalists against that of privacy pragmatists and unconcerned separately no statistically significant relationship was found. We believed that we did not have enough participants for the test to have enough power to detect the difference. In order to increase the power of the test, we attempted to sharpen the differences in predicted behavior between the segmentations by contrasting fundamentalists with the other two Westin segmentation groups, using the statistical technique of blocking. There was a significant association between whether participants were *privacy fundamentalists* and whether they submitted the form $\chi^2(1) = 4.39$, $p < 0.05$. Based on the odds ratio, the odds of a person submitting the form were 5.6 times higher if they were *non-fundamentalists*.

11.5.3 Acceptability of Data Requests

We transcribed the recordings of the post-session interviews and analyzed participants' comments[2] using thematic analysis [7]. We identified several factors that influence the acceptability of a data request (see Fig. 11.2). These factors help clarify why some data requests are considered acceptable while others are not.

[2]Participants are identified by the letter P and a number when quoted to maintain their anonymity.

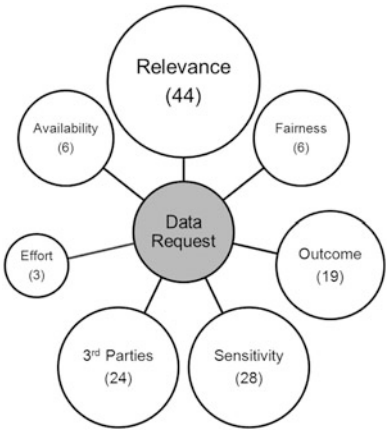

Fig. 11.2 Acceptability themes and frequencies of participants that mentioned each one

The acceptability of a data request is related to its perceived **relevance**. A relevant data request is one where the data item is perceived to be related to financial behavior, personality of the applicant, or probability of debt repayment. Relevant data requests were perceived more positively than irrelevant ones:

> I don't think it's acceptable, it's got nothing to do with my credit status. (P6)

> Yeah it's good, because the bank needs to know how much income you've got. (P13)

Some participants questioned the **fairness** of using certain items to assess an applicant. Fairness perceptions are associated with relevance perceptions; however, while perceived relevance seems to be related to how acceptable it is to use an item to draw conclusions from a statistical point of view, perceived fairness is related to how acceptable it is to use the item from an ethical perspective.

Perceptions of the consequences of disclosing a data item had an influence on acceptability as well. When participants thought that a data disclosure would result in a positive or neutral **outcome**, they saw it as more acceptable. On the other hand, participants perceived data disclosures they thought could harm them in future as less acceptable:

> I did reply, I answered, but only because I don't suffer from a medical condition. Probably if I did I might have reacted differently. (P17)

> I did disclose it on the answers because again I had nothing to hide, it would all go in my favor. (P29)

> I know that because I have medical conditions it could be used to discriminate against me. (P40)

The **sensitivity** of a data request has an influence on how acceptable it is perceived to be. When participants considered a request too personal, sensitive, or invasive, they perceived it as less acceptable.

> I found that very intrusive. I don't think that's acceptable. (P48)

Requests for data related to **third parties**, such as colleagues or friends of the participants, were perceived as less acceptable:

> Sharing other people's details is always something I find like quite hard to do. (P48)

Participants said they feared their friends might be hassled by the bank, that disclosing their data would be a privacy invasion, that it was not their data to give, and that their friends had not consented for their data to be disclosed:

> I wouldn't really want them to impose on my friends' personal space without them giving consent to that. (P25)

The **effort** required to answer a data request may also impact how it is perceived with request that are involve more work being seen more negatively:

> It would be difficult to get hold of the information, so again I was less inclined to provide it. (P30)

> Depending on how long the form is, I wouldn't mind doing it. (P36)

Asking for data that was already **publicly available** from other sources was perceived by some participants as more acceptable. In fact, a couple of participants even disclosed items they thought were unacceptable because they believed the data was already public:

> Yes I thought this was acceptable, insofar that social networking sites are sort of publicly accessible, and so giving the details of people with whom I have connections on these sort of sites is a reasonable thing to ask. (P23)

11.5.4 Acceptability and Disclosure

As expected, the acceptability ratings of items correlated significantly with their previously measured sensitivity ratings, $\rho = 0.607$, $p < 0.01$. However, the association between participants finding an item acceptable and disclosing it was only significant for three questions: *insurance claims* $\chi^2(2) = 10.44$, $p < 0.05$, *council tax* $\chi^2(2) = 10.10$, $p < 0.05$, and *emails and phone numbers of friends* $\chi^2(2) = 8.42$, $p < 0.05$.

For some items there was no association between acceptability and disclosure rate because every participant (or almost every participant) found the item acceptable and disclosed it. There were several items which a large proportion of participants found unacceptable, but still disclosed (see Table 11.4).

Participants who answered data requests they considered unacceptable were asked why they did. Fourteen participants said that, **on reflection, they did not mind disclosing** the data:

> I did, though I felt I shouldn't... they don't need to know that [...] although I did answer the question, because then I thought it might not be that bad. (P17)

Table 11.4 Acceptability vs. disclosure

Item	N[a]	Unacceptable & disclosed	% unacceptable & disclosed	% unacceptable & disclosed (excl. N/A)
Loved ones passed away while growing up	46	26	56.5	56.5
Social networking profiles of three friends	47	25	53.2	61.0
Name and phone number/email of three friends	47	20	42.6	42.6
Number and length of mobile text messages	46	19	41.3	43.2
Length of longest relationship	47	18	38.3	50.0
Grew up with both mother and father	44	18	40.9	40.9
Medical conditions	46	11	23.9	23.9
Professional networking profiles of three friends	45	3	6.7	33.3
Job of partner/spouse	46	3	6.5	15.8
Copy of insurance claims	41	2	4.9	7.1
Previous employer contact details	46	2	4.3	6.7
Copy of TV license payment history	45	2	4.4	7.1
Copy of gas/electricity payment history	45	1	2.2	2.8
Copy of council tax payment history	46	1	2.2	3.8

[a] Participants who, in the interview, did not answer clearly whether they found an item acceptable or not were deleted pairwise

Ten participants answered that even though they considered a question **generally unacceptable**, they personally had no problem with answering it:

> Again I did disclose it, but I don't think the general public would be happy [. . .] because I see myself as quite an open person, so I would be happy. (P28)

Five participants said they did disclose because they **wanted to complete the form**:

> I did disclose some things mainly just to complete the questionnaire. But it didn't seem a great question. (P27)

Other reasons for answering unacceptable data requests included:

1. Answering is **not harmful** to me (four participants);
2. The data is **publicly available** anyway (two participants);
3. The **bank will not actually look** at the data (two participants);
4. **Was not thinking** about it when I answered (one participant);
5. I felt safe answering because I was **part of a study** (one participant).

11.5.5 Privacy Protection Behaviors

All participants were asked during the post-experiment interview if they had lied or exaggerated on some items when completing the form: 22.9 % of participants said that they had. Examples include saying that the bank could check on their electricity bills when they actually do not pay any, and writing friends' initials instead of their names. One reason mentioned to do this was to increase the amount of data disclosed to the minimum required to be able to submit the form. Another reason given was to protect the privacy of friends.

11.6 Discussion

Our study investigated the role of sensitivity, transparency, and privacy values in decision-making about disclosure in the context of a simulated credit card application form. We also wanted to explore the interaction between stated acceptability of a data request, and disclosure behavior regarding the same data request.

11.6.1 Sensitivity, Transparency, and Privacy Values

Hypothesis 1 stated that the number of participants sharing each data item would be inversely correlated with the sensitivity of the data items. In fact, the answer rates for each question showed a significant negative correlation with the sensitivity rating of the question (as measured in a previous study), thus supporting the hypothesis. Past research found that more sensitive items were more likely to be withheld [36]. The importance of our finding is that it can be used to estimate a priori how an application or registration form will fare, before actually deploying it. Knowing how sensitive certain data items are perceived to be in general makes it possible to predict the likelihood of applicants withholding such items, and weigh the impact of missing data on the lender's business processes to determine whether it is actually worth requesting it.

H3 stated that *privacy fundamentalists* would disclose less data than *privacy unconcerned* or *privacy pragmatists*. As expected, participants who were categorized as *privacy fundamentalists* on Westin's scale were significantly less likely to submit the form than non-fundamentalists. Privacy fundamentalists are generally more concerned about the risks of their personal data falling into the wrong hands and of the harmful effects that disclosing personal data can have on their lives [49]. This would explain their reluctance in submitting their personal data to an unknown party for an uncertain reward, i.e., the reward would have to be larger to offset the perceived cost of answering and submitting the form.

H2 predicted that participants would disclose more data when a reason for the data request was given than when no reason was given. However, even though previous studies identified lack of transparency and legitimacy as promoters of negative reactions [3,24], in our study the presence of explanations for the questions being asked had no significant effect on participant behavior. Thus, this hypothesis was not supported. One possible explanation is that participants did not notice the explanations positioned below each question. Another possibility is that they saw the explanations, but did not read them. Past research on privacy policies found that people rarely read them, or other terms online, because of the time cost, which has been estimated as an average of 10 min per policy [35]—so our participants may not have wanted to spend time reading the explanation. If participants read the explanations they may not have understood, or believed them—we did not ask our participants about this. In future studies, user behavior, such as mouse and eye movements, should be tracked to check whether participants are noticing and reading the explanations.

11.6.2 Disclosure and Acceptability of Novel Items

Overall, the disclosure rates for the *Novel* items (excluding N/A answers) can be considered high: 85 % or more for all but three items. Items related to family history had surprisingly high disclosure rates (100 and 93.8 % respectively for "*Grew up with both mother and father*" and "*Loved ones passed away while growing up*"), as did "*Medical conditions*" and "*Length of longest relationship*". These are all items generally considered to be very sensitive.

One possible explanation for the high disclosure rates is that no relationship was found between acceptability of a question and its disclosure rate. Even though acceptability and sensitivity ratings were significantly correlated, the acceptability and disclosure rates for individual questions were not with many participants both rating questions as unacceptable and answering them. For example, 56.5 % of participants considered the question about "*Loved ones passed away while growing up*" unacceptable, but still answered it. The only exceptions were "*Copy of insurance claims*", "*Copy of council tax payment history*", and "*Name and phone number/email of 3 friends*"

The thematic analysis provides some insights into why this happens. Several participants said that—even though they found a particular question generally unacceptable—they personally did not mind answering it. This suggests that the assessment of the acceptability of a data request precedes the actual individual cost-benefit evaluation of the disclosure. Participants may believe that it is wrong for a lender to ask for particular data items, but feel that in their personal case it is beneficial (or not costly) to answer. This is further supported by some participants saying they answered "unacceptable" questions so they could submit the application form. They weighed the effort already invested plus the benefit of entering the prize draw against the costs of disclosure, and decided for disclosure. This suggests that

when individuals answer surveys about privacy, they may be answering according to the perceived abstract acceptability of certain data practices which may differ from their personal cost-benefit assessment in a real situation. This would help explain why a difference between privacy attitudes and behaviors has been observed in the literature [1, 6].

Items relating to social networks had among the lowest disclosure rates. These items included:

- Number and length of mobile text messages;
- Name and phone number/email of three friends;
- Professional networking profiles of three friends;
- Social networking profiles of three friends.

All of these can be taken as indexes of participants' social capital. We have already noted that social capital is related to trustworthiness [33]. However, these data items are about individuals other than the participant and with whom the participant is friends. The thematic analysis revealed that participants were not comfortable revealing data about their friends without their permission. Similarly, "*Partner's job*" was among the least disclosed items.

Items related to bill payment history, such as utility, TV license, and council tax payment history had high disclosure rates, and a low proportion of participants found them unacceptable. This gives support to the current trend for lenders to use some types of bill payment history as indicators of creditworthiness, especially when applicants have "thin" credit histories, to make credit scoring more accurate.

Several factors identified in the thematic analysis confirm previous results. In a previous study on applicants' perceptions of loan application forms [24], participants similarly raised issues with: perceived lack of relevance of data requests; level of detail needed to reply to some requests; potential negative outcome of a disclosure; and perceived unfairness of application process. Relevance of a data request, sensitivity of data, and disclosure outcome are all also identified by Culnan [12] when reviewing factors which impact perceptions of secondary use of information. Culnan [12] argues that individuals are less likely to perceive that their privacy was invaded when the collected personal data is considered to be relevant for the interaction taking place and will be used to draw reliable conclusions about them. Sensitivity of data is generally considered to be related to privacy perceptions (see [2] for a privacy model in multimedia communications, or [36] for findings in e-commerce). In a study focused on privacy perceptions in serious games, Malheiros et al. [34] also identified perceived outcome of sharing data as an important factor. The emergence of factors in our thematic analysis which have been identified in studies focused on different types of contexts suggest that the process through which individuals assess data requests may be context-independent, which does not mean the assessments themselves are.

11.6.3 Privacy Protection Behaviors

Twenty-three percentage of participants admitted to falsifying some of the data they submitted as a way to obtain the benefits of submitting the form (and the chance to get a £50 prize) while minimizing the data actually disclosed. Metzger's [36] study found an almost identical correlation between item sensitivity and disclosure (0.61) as this study (0.62), but a higher proportion of participants that falsified (40 % of participants that falsified at least one data item). Metzger's participants were asked about falsification in a self-administered questionnaire, whereas ours were asked face-to-face by the experimenter. Survey work to estimate the prevalence of socially undesirable behavior (for example sexual infidelity in marriage) has found that more people admit to these behaviors to self-administered questionnaires than to experimenters. The difference can be large—six times as many admitted infidelity when asked by a form than by interview [50]. We hypothesize that social desirability effects due to the presence of experimenters may have led to under reporting of falsification in our study, and encourage other researchers to address this source of bias more effectively when designing their studies, by employing methods that are more resistant to this bias, for example: self-report questionnaires, or random response techniques (such as participants flipping a coin to answer truthfully or answer yes [4]) that make it impossible to tell if each individual respondent's answer is truthful, but allow an accurate assessment of the true proportion in the sample as a whole.

No data was collected in this study on the rate of falsification per item (we made sure participants' data was not saved to comply with ethics guidelines), but if a relationship could be found between sensitivity of an item and its falsification rate (as in [36]), then the data quality impact for lenders of asking for certain items could be bounded.

11.6.4 Limitations

Our participants were university students with an average age of 20. We acknowledge that this limits the generalizability of our results, and plan to repeat the study with a larger, more representative sample. We would, however, argue that the findings of our study have face validity when considered in the context of previous results. Westin's Privacy segmentation has been repeatedly given across many different samples in different years. A consistent finding is that approximately 25 % of respondents fall into the *privacy fundamentalist* category [31]. Our participants had a smaller proportion, with 16.7 % being *privacy fundamentalists*. This agrees with Tsarenko and Tojib's [47] finding that young people were more pragmatic in their privacy concerns *viz* financial institutions than other segments of the population. We argue that by being more pragmatic and unconcerned than the general population, the disinclination shown by our participants for disclosing

certain data items can be expected in the general population, and that our results would form an upper bound for disclosure of these items in the general population. Also, the preliminary survey was conducted with a larger (N = 285), nationally representative sample, and the sensitivity ratings correlated significantly with the experimental study's disclosure rates; thus, contributing to the external validity of our findings.

11.7 Conclusions

From a methodological point of view, this study breaks with common practice by deceiving participants into thinking the data they submitted was actually going to be used to assess financial reliability. A monetary reward for the most creditworthy participant was also offered to nudge participants into submitting their form and answering questions in a truthful manner. Furthermore, experimenters were under the same deception as the participants, to minimize bias. Since privacy decision-making and disclosure behavior are highly contextual it is important to capture and observe them in conditions as realistic as possible.

A goal of this study was to discover which novel data items could potentially be used as alternatives for evidence of credit worthiness for applicants who do not have conventional credit histories, and so could not otherwise participate in and receive the benefits of low cost credit. Among the most sensitive of the novel data items studied in this research (as measured by sensitivity score and disclosure rates) were those relating to people other than the participant. Although the results need to be validated with a wider socio-demographic (where we estimate individuals to be less pragmatic), we consider this study to be a warning that use of indices of social capital as signs of creditworthiness may currently not be acceptable. The explicit collection of items associated with bill payment history, on the other hand, seem to be less sensitive. Items such as TV license and council tax payment history, which are not currently collected, could be used for credit scoring in situations where applicants have "thin" credit histories.

In the context of applying for credit, we found a direct relationship between an item's sensitivity, and its likelihood of being disclosed, and that this relationship might be employed in a cost/benefit analysis during the design phase for credit application procedures. However, care must be taken in the choice of language when assessing sensitivity using survey methods: we found that similar language can tap quite distinct constructs that relate very differently to observed behavior. We found no relationship between items' "acceptability" and their disclosure; many people disclosed information whilst reporting that the antecedent information requests were unacceptable. We hypothesize that there are two separate but related tests employed by credit applicants for assessing information requests—one for testing the requests' general acceptability (that has little impact on disclosure behaviors), and one with respect to the individual's costs and benefits (with much greater impact on disclosure).

A growing body of privacy research is starting to look at privacy decision-making as outcome-oriented: individuals assess the costs and benefits of trading their personal data for some kind of reward. Our research provides some insights into the factors that guide this decision-making process.

The impact of perceived relevance and fairness in particular should be of note to any organization that collects personal data and uses it for profiling purposes. Empirical score-carding [16], for example, may find a relationship between a data item and likelihood of default which, while statistically sound, may not be understood by applicants. In fact, these relationships are usually kept hidden from applicants to prevent gaming of the application process, which makes it more difficult for applicants to perceive the relevance of certain data requests. Furthermore, even if the collection of certain types of data is seen as statistically relevant, applicants may still consider the practice unfair or unethical.

We detected no effect of request transparency on disclosure—participants were just as likely to disclose data whether or not an explanation was given for the request. This suggests that, in contexts where there is a low perceived relevance of data requests, organizations should explore new ways to assure individuals that their data collection and data use practices are actually relevant and fair.

We also found that 23 % of our participants admitted to falsifying, exaggerating or omitting information when completing our simulated application form. We have no data with which to compare an item's sensitivity to its falsification rate in the context of applying for credit—a topic that requires further studies in which participants' responses are retained and verified through more robust processes.

Acknowledgements We would like to thank Madalina Vasilache, Diana Franculescu and Jessica Colson for testing and interviewing the study's participants. We would also like to thank Conor Fisk for helping to transcribe the interview data. This research is funded partly by the Engineering and Physical Sciences Research Council through the Privacy Value Networks project (EP/G002606/1).

References

1. Acquisti, A.: Privacy in electronic commerce and the economics of immediate gratification. In: Proceedings of the 5th ACM Conference on Electronic Commerce, New York, pp. 21–29. ACM, New York (2004)
2. Adams, A., Sasse, M.: Privacy in multimedia communications: protecting users, not just data. In: Blandford, A., Vanderdonckt, J., Gray, P. (eds.) People and Computers XV – Interaction Without Frontiers: Joint Proceedings of HCI 2001 and IHM 2001, Lille, pp. 49–64. Springer, London (2001)
3. Annacker, D., Spiekermann, S., Strobel, M.: e-privacy: evaluating a new search cost in online environments. In: Proceedings of the 14th Bled Electronic Commerce Conference (BLED 2001), Bled, 2001, pp. 292–308
4. Barnett, J.: Sensitive questions and response effects: an evaluation. J. Manag Psychol **13**(1/2), 63–76 (1998)
5. Belsky, E., Calder, A.: Credit matters: low-income asset building challenges in a dual financial service system. Joint Center for Housing Studies, Harvard University. http://www.jchs.harvard.edu/sites/jchs.harvard.edu/files/babc_04-1.pdf (2004)

6. Berendt, B., Günther, O., Spiekermann, S.: Privacy in e-commerce: stated preferences vs. actual behavior. Commun. ACM **48**(4), 101–106 (2005)
7. Braun, V., Clarke, V.: Using thematic analysis in psychology. Qual. Res. Psychol. **3**(2), 77–101 (2006)
8. Buchanan, T., Paine, C., Joinson, A.N., Reips, U.: Development of measures of online privacy concern and protection for use on the internet. J. Am. Soc. Inf. Sci. Technol. **58**(2), 157–165 (2007)
9. Business Week, Harris Poll: Online insecurity. Technical report, Business Week/Harris Poll. http://www.businessweek.com/1998/11/b3569107.htm (1998)
10. Collard, S., Kempson, E.: Affordable Credit: The Way Forward. The Policy Press, Bristol (2005)
11. Credit Action: UK debt statistics from credit action. http://www.creditaction.org.uk/helpful-resources/debt-statistics.html (2011)
12. Culnan, M.J.: "How did they get my name?": an exploratory investigation of consumer attitudes toward secondary information use. MIS Q. **17**(3), 341–363 (1993)
13. Glaeser, E.L., Laibson, D., Scheinkman, J.A., Soutter, C.L.: What is social capital? The determinants of trust and trustworthiness. Technical report, National Bureau of Economic Research. http://www.nber.org/papers/w7216 (1999)
14. Grossklags, J., Acquisti, A.: When 25 cents is too much: an experiment on willingness-to-sell and willingness-to-protect personal information. In: Workshop on Economics of Information Security, Pittsburgh, 2007
15. Hand, D.: Modelling consumer credit risk. IMA J. Manag. Math. **12**(2), 139 (2001)
16. Hand, D., Brentnall, A., Crowder, M.: Credit scoring: a future beyond empirical models. J. Financ. Transform. **23**, 121–128 (2008)
17. Hann, I., Hui, K., Lee, S., Png, I.P.L.: Online information privacy: measuring the cost-benefit trade-off. In: Applegate, L., Galliers, R.D., DeGross, J.I. (eds.) Proceedings of the Twenty-Third International Conference on Information Systems, Barcelona, 2002, pp. 1–10
18. Hann, I., Hui, K., Lee, T.S., Png, I.P.L.: The value of online information privacy: evidence from the USA and Singapore. In: International Conference on Information Systems, Barcelona (2002)
19. Hine, C., Eve, J.: Privacy in the marketplace. Inf. Soc. **14**, 253–262 (1998)
20. Horne, D.R., Norberg, P.A., Ekin, A.C.: Exploring consumer lying in information-based exchanges. J. Consum. Mark. **24**(2), 90–99 (2007)
21. Hui, K., Teo, H., Lee, S.: The value of privacy assurance: an exploratory field experiment. MIS Q. **31**(1), 19–33 (2007)
22. Hunt, J., Fry, B.: Spendsmart: How to Tackle Debt, Know Your Money Mind & Make Your Cash Go Further. Hachette, London (2009)
23. Infosecurity Europe: Woman 4 times more likely than men to give passwords for chocolate. http://www.eskenzipr.com/page.cfm/T=m/Action=Press/PressID=202 (2008)
24. Jennett, C., Malheiros, M., Brostoff, S., Sasse, M.: Privacy for applicants versus lenders' needs for predictive power: is it possible to bridge the gap? In: European Data Protection: In Good Health? Brussels, pp. 35–51. Springer, Netherlands (2012)
25. Jentzsch, N.: Financial Privacy: An International Comparison of Credit Reporting Systems. Springer, Berlin/New York (2007)
26. Jones, P.A.: Access to credit on a low income. Technical report, The Co-operative Bank. http://www.wip.smile.co.uk/images/pdf/access_to_credit_final_report.pdf (2001)
27. Junger, M., van Kampen, M.: Cognitive ability and self-control in relation to dietary habits, physical activity and bodyweight in adolescents. Int. J. Behav. Nutr. Phys. Act. **7**, 22 (2010)
28. Jupiter Research: Security and privacy data. FTC security workshop. http://www.ftc.gov/bcp/workshops/security/020520leathern.pdf (2002)
29. Kirchler, E., Hoelzl, E., Kamleitner, B.: Spending and credit use in the private household. J. Socio-Econ. **37**(2), 519–532 (2008)
30. Kourti, I.: Project FLAME social study report. Technical report. http://tnc2009.terena.org/core/getfile2f59.pdf (2009)

31. Kumaraguru, P., Cranor, L.: Privacy indexes: a survey of Westin's studies. Technical report, Institute for Software Research, Carnegie Mellon University. http://repository.cmu.edu/isr/856 (2005)
32. Landesberg, M.K., Levin, T.M., Curtin, C.G., Lev, O.: Privacy online: a report to Congress. Technical report, Federal Trade Commission. http://www.ftc.gov/reports/privacy3/toc.shtm (1998)
33. Lin, M., Prabhala, N., Viswanathan, S.: Judging borrowers by the company they keep: social networks and adverse selection in online peer-to-peer lending. SSRN eLibrary (2009)
34. Malheiros, M., Jennett, C., Seager, W., Sasse, M.: Trusting to learn: trust and privacy issues in serious games. In: McCune, J.M., Balacheff, B., Perrig, A., Sadeghi, A.R., Sasse, A., Beres, Y. (eds.) Trust and Trustworthy Computing, vol. 6740, pp. 116–130. Springer, Berlin/Heidelberg (2011)
35. McDonald, A.M., Cranor, L.F.: Cost of reading privacy policies, the. I/S J. Law Policy Inf. Soc. **4**, 543 (2008)
36. Metzger, M.J.: Communication privacy management in electronic commerce. J. Comput. Mediat. Commun. **12**(2), 335–361 (2007)
37. Milne, G.R., Gordon, M.E.: Direct mail privacy-efficiency trade-offs within an implied social contract framework. J. Public Policy Mark. **12**(2), 206–215 (1993)
38. Nissenbaum, H.: Privacy as contextual integrity. Wash. Law Rev. **79**(1), 119 (2004)
39. O'Hara, K., Shadbolt, N.: The Spy in the Coffee Machine. Oneworld, Oxford (2008)
40. Pew Internet and American Life Project: Trust and privacy online: why Americans want to rewrite the rules. Technical report, The Pew Internet & American Life Project. http://pewinternet.org/~/media//Files/Reports/2000/PIP_Trust_Privacy_Report.pdf.pdf (2000)
41. Phelps, J., Nowak, G., Ferrell, E.: Privacy concerns and consumer willingness to provide personal information. J. Public Policy Mark. **19**(1), 27–41 (2000)
42. Pine, K., Gnessen, S.: Sheconomics. Headline, London (2009)
43. PRBC Credit Reporting Agency: Nontraditional credit report. http://www.microbilt.com/nontraditional-credit-report.aspx (2011)
44. Royal Bank of Scotland: Your loan application. http://www.rbs.co.uk/personal/loans/loan-application-info.ashx (2011)
45. Taylor, H.: Most people are "Privacy pragmatists" who, while concerned about privacy, will sometimes trade it off for other benefits. Technical report, Harris Interactive. http://www.harrisinteractive.com/vault/Harris-Interactive-Poll-Research-Most-People-Are-Privacy-Pragmatists-Who-While-Conc-2003-03.pdf (2003)
46. Thomas, L.: A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. Int. J. Forecast. **16**(2), 149–172 (2000)
47. Tsarenko, Y., Tojib, D.R.: Examining customer privacy concerns in dealings with financial institutions. J. Consum. Mark. **26**(7), 468–476 (2009)
48. Westin, A.F.: E-Commerce & Privacy: What Net Users Want. Privacy & American Business, Hackensack (1998)
49. Westin, A.F.: Social and political dimensions of privacy. J. Soc. Issues **59**(2), 431–453 (2003)
50. Whisman, M.A., Snyder, D.K.: Sexual infidelity in a national survey of American women: differences in prevalence and correlates as a function of method of assessment. J. Fam. Psychol. **21**(2), 147–154 (2007)

Part IV
Economics of Cybercrime

Chapter 12
Measuring the Cost of Cybercrime

Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel J. G. van Eeten, Michael Levi, Tyler Moore, and Stefan Savage

Abstract This chapter documents what we believe to be the first systematic study of the costs of cybercrime. The initial workshop paper was prepared in response to a request from the UK Ministry of Defence following scepticism that previous studies had hyped the problem. For each of the main categories of cybercrime we set out what is and is not known of the direct costs, indirect costs and defence costs – both to the UK and to the world as a whole. We distinguish carefully between traditional crimes that are now "cyber" because they are conducted online (such

R. Anderson (✉) · R. Clayton
Computer Laboratory, University of Cambridge, Cambridge, UK
e-mail: ross.anderson@cl.cam.ac.uk; richard.clayton@cl.cam.ac.uk

C. Barton
Security Research and Operations, Cloudmark, Inc., Reading, UK

R. Böhme
Department of Information Systems, University of Münster, Münster, Germany
e-mail: rainer.boehme@uni-muenster.de

M.J.G. van Eeten
Faculty of Technology, Policy and Management, Delft University of Technology, Delft, Netherlands
e-mail: m.j.g.vaneeten@tudelft.nl

M. Levi
School of Social Sciences, Cardiff University, Cardiff, UK
e-mail: levi@cf.ac.uk

T. Moore
Department of Computer Science and Engineering, Southern Methodist University, Dallas, TX, USA
e-mail: tylerm@smu.edu

S. Savage
Department of Computer Science and Engineering, University of California, San Diego, CA, USA
e-mail: savage@cs.ucsd.edu

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_12, © Springer-Verlag Berlin Heidelberg 2013

as tax and welfare fraud); transitional crimes whose modus operandi has changed substantially as a result of the move online (such as credit card fraud); new crimes that owe their existence to the Internet; and what we might call platform crimes such as the provision of botnets which facilitate other crimes rather than being used to extract money from victims directly. As far as direct costs are concerned, we find that traditional offences such as tax and welfare fraud cost the typical citizen in the low hundreds of pounds/euros/dollars a year; transitional frauds cost a few pounds/euros/dollars; while the new computer crimes cost in the tens of pence/cents. However, the indirect costs and defence costs are much higher for transitional and new crimes. For the former they may be roughly comparable to what the criminals earn, while for the latter they may be an order of magnitude more. As a striking example, the botnet behind a third of the spam sent in 2010 earned its owners around $2.7 million, while worldwide expenditures on spam prevention probably exceeded a billion dollars. We are extremely inefficient at fighting cybercrime; or to put it another way, cyber-crooks are like terrorists or metal thieves in that their activities impose disproportionate costs on society. Some of the reasons for this are well-known: cybercrimes are global and have strong externalities, while traditional crimes such as burglary and car theft are local, and the associated equilibria have emerged after many years of optimisation. As for the more direct question of what should be done, our figures suggest that we should spend less in anticipation of cybercrime (on antivirus, firewalls, etc.) and more in response – that is, on the prosaic business of hunting down cyber-criminals and throwing them in jail.

12.1 Introduction

As countries scramble to invest in information security, governments want to know how large that investment should be, and what the money should be spent on. This creates a demand among rational policy-makers for accurate statistics of online/electronic crime and abuse. However, many of the existing surveys are carried out by organisations (such as antivirus software vendors or police agencies) with a particular view of the world and often a specific agenda. This chapter therefore sets out to collate what is known, and what is not, as of the beginning of 2012.

It builds on a report written by four of us in 2008 for the European Network and Information Security Agency, ‘Security Economics and the Single Market’ [2]. There we analysed the statistics available at the time, their shortcomings, and the ways in which they could lead to incorrect policy decisions.

For example, the number of phishing websites, of distinct attackers and of different types of malware is persistently over-reported, leading some police forces to believe that the problem is too large and diffuse for them to tackle, when in fact a small number of gangs lie behind many incidents and a police response against them could be far more effective than telling the public to fit anti-phishing toolbars or purchase antivirus software. This is part of a much wider problem of attributing risks to patterns of offending.

There are over 100 different sources of data on cybercrime, yet the available statistics are still insufficient and fragmented; they suffer from under- and over-reporting, depending on who collected them, and the errors may be both intentional (e.g., vendors and security agencies playing up threats) and unintentional (e.g., response effects or sampling bias). The more prominent sources include surveys (from Eurostat, CSI and consultancies); security breach disclosure reports; direct observations of attack trends (e.g., from Symantec, McAfee and Microsoft); and reports by trade bodies (from banking trade associations, or the Anti-Phishing Working Group). We compared and analysed the CSI, Eurostat and Symantec statistics in the aforementioned ENISA report [2].

The proximate motivation for this chapter was a request from the Chief Scientist at the UK Ministry of Defence, Sir Mark Welland, for an update on the analysis we produced in the 2008 report. This was driven in turn by the publication in February 2011 of a report [9] commissioned by the UK Cabinet Office from Detica (part of BAE plc) which estimated cybercrime's annual cost to the UK to be £27 billion (about 1.8 % of GDP). That report was greeted with widespread scepticism and seen as an attempt to talk up the threat; it estimated Britain's cybercrime losses as £3 billion by citizens, £3 billion by the government and a whopping £21 billion by companies. These corporate losses were claimed to come from IP theft (business secrets, not copied music and films) and espionage, but were widely disbelieved both by experts and in the press. The Ministry of Defence asked us to set out what figures are known, what can reasonably be estimated and what can only be guessed.

We begin by setting out a framework for analysing the costs of cybercrime in Sect. 12.2, differentiating cybercrimes from physical ones and decomposing cost categories. Next, in Sect. 12.3 we review available information on costs for all substantial categories of cybercrime. We discuss the costs of the common infrastructure facilitating many types of cybercrime in Sect. 12.4, and we present the costs together in summary form in Sect. 12.5 before concluding in Sect. 12.6.

12.2 A Framework for Analysing the Costs of Cybercrime

Even before computers started to make things more complicated, it was already hard to define and measure white-collar crimes. While it is clearly a crime to set up a fly-by-night mail-order firm, collect payments and ship no goods, the situation is less clear when goods are mis-described or defective. Periodic scandals (McKesson & Robbins in 1938, IOS and Equity Funding in 1973, Enron in 2001, the banking crisis in 2008) raise questions about the boundary between business and crime, leading to changes in definitions as well as regulations. These shifts are associated with changes in social attitudes and political discourse; for a discussion see [36,37].

While tying down fraud was hard enough a decade ago, globalisation and technology are making the problem harder still today. Many corporations are transnational, as are many cybercrimes. If a Chinese gang steals secrets from BAE, is this a UK crime as BAE has its primary stock-market listing in London, or a

US one as it does more business there? Furthermore, while there are some online and electronic crimes for which we have UK figures (such as card fraud) there are others for which we have only global figures (such as the incomes of gangs selling fake pharmaceuticals or operating botnets). In these circumstances, the sensible way forward is to estimate global figures. We will work from the fact that the UK accounts for about 5 % of world GDP to scale our national estimates up or down as appropriate. Where there is reason to believe that the UK figures are out of line with other countries, we will say so and make an appropriate allowance.

12.2.1 Differentiating Cybercrime from Other Crime

In May 2007 the European Commission issued a Communication "towards a general policy on the fight against cyber crime", noting that there was not even an agreed definition of cybercrime [11]. It proposed a threefold definition:

1. Traditional forms of crime such as fraud or forgery, though committed over electronic communication networks and information systems;
2. The publication of illegal content over electronic media (e.g., child sexual abuse material or incitement to racial hatred);
3. Crimes unique to electronic networks, e.g., attacks against information systems, denial of service and hacking.

We propose to follow this definition here, despite the fact that the boundary between traditional crime and cybercrime remains fluid. Advances in information technology are moving many social and economic interactions from the physical world to cyberspace, so a moving boundary between cyber and physical is inescapable. For example, the UK will move all claims for welfare payments online in 2013, and most claims are made that way already; welfare fraud is 0.8 % of the £152 billion expenditure of the Department of Work and Pensions, or a tad over £1.2 billion. Income tax fraud (evasion as opposed to avoidance) adds about a further £3 billion. These sums dwarf the amounts stolen by 419 fraudsters or even carders. What is important is to have a yardstick with which to measure changes. For that reason, we have to decompose fraud figures into different categories.

We must also not lose sight of the big picture. The reduced transaction costs and economies of scale brought by the Internet have unleashed substantial productivity gains [3]. We may hope that the overall costs of crime will go down, in the sense that the value of the electronic versions of old-fashioned crimes will decrease by more than the value of new crimes made possible by new technology. Nevertheless, we should bear in mind that even if the costs of crime go up, there may still be a substantial net gain for society.

12.2.2 Decomposing the Cost

As for measuring costs, the Detica report considered four categories:

1. Costs in anticipation of cybercrime, such as antivirus software, insurance and compliance;
2. Costs as a consequence of cybercrime, such as direct losses and indirect costs such as weakened competitiveness as a result of intellectual property compromise;
3. Costs in response to cybercrime, such as compensation payments to victims and fines paid to regulatory bodies;
4. Indirect costs such as reputational damage to firms, loss of confidence in cyber transactions by individuals and businesses, reduced public-sector revenues and the growth of the underground economy.

We will not use this methodology as it stands, as the second heading includes both direct and indirect costs of which the former might at least in principle be measured accurately while the latter are harder to assess. The third item we view as being composed entirely of direct costs: if a bank designs an insecure website and has to pay compensation to customers whose accounts are debited without their mandate, these are clearly direct costs. We therefore use a more straightforward approach, in which we simply split direct costs from indirect costs. We will also have some things to say about the costs of security (though these cannot always be allocated to specific types of crime) and about the social and opportunity costs of reduced trust in online transactions.

Where possible we will decompose the costs of crime still further. Just as a thief who steals the lead from a church roof, or copper wire from a railway signalling system, may earn a handful of cash while doing damage that costs tens of thousands to repair (and disrupts the lives of thousands), so there can also be large asymmetries in costs and revenues.

Figure 12.1 visualizes our framework. We define and discuss its cost categories in the following paragraphs.

12.2.2.1 Criminal Revenue

Criminal revenue is the monetary equivalent of the gross receipts from a crime. We do not include any lawful business expenses of the criminal.[1] For example, an illicit online pharmacy may purchase hosting services from a legitimate provider and pay the market price. This reduces the criminal's profits, but contributes to the gross product of the economy in which the provider is located.

[1]The UK Proceeds of Crime Act does not allow an offender's costs to be deducted from the amount he is deemed to owe the state.

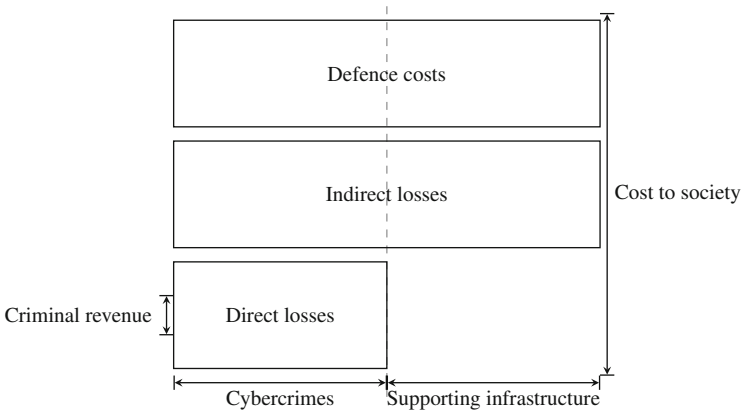

Fig. 12.1 Framework for analysing the costs of cybercrime

But consider phishing advertised by email spam. The phisherman's criminal revenue is the sum of the money withdrawn from victim accounts. If spamming is also a crime, and is carried out using a botnet (a network of subverted PCs, see Sect. 12.4.1), then the revenue of the spammer, possibly split with the owner of the botnet, must also be accounted as part of the criminal-revenue contribution to GDP.

12.2.2.2 Direct Losses

Direct loss is the monetary equivalent of losses, damage, or other suffering felt by the victim as a consequence of a cybercrime.

Example – Direct losses include:

- Money withdrawn from victim accounts;
- Time and effort to reset account credentials (for banks and consumers);
- Distress suffered by victims;
- Secondary costs of overdrawn accounts: deferred purchases, inconvenience of not having access to money when needed;
- Lost attention and bandwidth caused by spam messages, even if they are not reacted to.

As a practical matter we will generally disregard distress; victims are not generally entitled to sue for it, it is hard to measure, and it is generally worst when exacerbated by secondary victimisation (such as banks disbelieving complaints from

victims). Even if we chose to include distress (as has been done in Home Office studies of the costs of violent crime) there is limited data available about the costs of the time spent repairing stolen[2] identities.[3]

12.2.2.3 Indirect Losses

Indirect loss is the monetary equivalent of the losses and opportunity costs imposed on society by the fact that a certain cybercrime is carried out, no matter whether successful or not and independent of a specific instance of that cybercrime. Indirect costs generally cannot be attributed to individual victims.

Example – Indirect losses include:

- Loss of trust in online banking, leading to reduced revenues from electronic transaction fees, and higher costs for maintaining branch staff and cheque clearing facilities;
- Missed business opportunity for banks to communicate with their customers by email;
- Reduced uptake by citizens of electronic services as a result of lessened trust in online transactions;
- Efforts to clean-up PCs infected with malware for a spam sending botnet.

Observe in Fig. 12.1 that indirect losses is the first category to span both cybercrimes and its supporting infrastructure. The whole idea of distinguishing criminals' profit centres from the common infrastructure being employed in the crimes is to avoid allocating the collateral damage caused by the infrastructure to the actual types of cybercrimes, where they would show up as indirect losses. Since the means (e.g., botnets) would not be around if there were not ends (e.g., phishing victims), we consider losses caused by the cybercriminal infrastructure as indirect

[2]Banks may like to describe impersonation as "identity theft" as it carries with it an implied liability shift – that it wasn't the bank's money that was stolen but the customer's identity. This is controversial; as well as dumping liability it increases the fear of crime. Victimisation studies, such as the BCS and GetSafeOnline, show considerable public anxiety about card fraud and impersonation.

[3]The US Identity Theft Resource Center studies the time taken by victims to repair damage caused by identity theft. In 2004, it was more than 300 h; in 2008, 76 h; and in 2009, 68 h [16]. This survey, although interesting, is rather limited in that there were just 203 victims reporting their experiences and they were all assisted by the ITRC, and so might have been reasonably efficient at dealing with their problems.

by nature; irrespective of whether or not the legal framework formally criminalizes the means.

12.2.2.4 Defence Costs

Defence costs are the monetary equivalent of prevention efforts. They include direct defence costs, i.e., the cost of development, deployment, and maintenance of prevention measures, as well as indirect defence costs, such as inconvenience and opportunity costs caused by the prevention measures.

Example – Defence costs include:

- Security products such as spam filters, antivirus, and browser extensions to protect users;
- Security services provided to individuals, such as training and awareness measures;
- Security services provided to industry, such as website take-down services;
- Fraud detection, tracking, and recuperation efforts;
- Law enforcement;
- The inconvenience of missing messages falsely classified as spam.

Defence costs, like indirect losses, are largely independent of individual victims. Often it is even difficult to allocate them to individual types of cybercrime. Defences can target the actual crimes or their supporting infrastructure, and the costs can be incurred in anticipation of or reaction to crimes, the latter being to deter copycats.

12.2.2.5 Cost to Society

The cost to society is the sum of direct losses, indirect losses, and defence costs.

12.2.3 Discussion of the Framework

As we shall see, criminal revenue is in practice significantly lower than direct losses and much lower than direct plus indirect losses. In a classical analysis, policy may be guided by looking at direct or indirect losses; a company may invest in protection to the extent that this reduces its direct costs, while a government might consider indirect losses and invest in collective defence efforts (such as policing) so long as

every extra unit of defence spending reduces the sum or direct and indirect losses by at least as much.

It is possible to spend too much on defence. The 9/11 Commission estimated that the September 2001 attacks cost no more than $500,000 to carry out [46], but in 2008 it was estimated that the USA had already spent over $3 trillion on defence costs and on the wars in Afghanistan and Iraq [51]. Criminologists examine the widely different expenditure on preventing deaths from terrorism, road traffic accidents, and domestic violence in terms of signal crimes [24] where the symbolic dimensions reach into the general psyche and demand action. One can put a price on this – using willingness-to-pay models – but there is not a simple cost equation.

Such behavioural theories may not be sufficient to explain all irrational expenditures in the case of cybercrime. For example, it has been reported that the botnet behind a third of all the spam sent in 2010 earned its operators a mere $2.7 million in profit from sales of knockoff pharmaceuticals, while the cost of spam to ISPs and users worldwide is in the billions [32]. Thanks to spam filtering, spam is no longer salient to most citizens in the way that terrorism is.

The misallocation of resources associated with cybercrime results mostly from economic and political factors rather than from behavioural ones. Globalisation means that for much online crime, the perpetrators and victims are in different jurisdictions. This reduces both the motivation and the opportunity for police action. In the case of spam, for example, it is not socially optimal for ISPs to be spending hundreds of millions of dollars on coping with floods of spam; a more rational policy would be to arrest the criminals. Yet the Russian state has not co-operated sufficiently for this to happen. The long-term winners may be firms like Google and Microsoft as people are driven to webmail services with good spam protection.

We will return to this complexity in the conclusions. In the next two sections we collect what is known about the actual costs, and add our own estimates where appropriate. Section 12.3 iterates through all relevant types of cybercrimes, the cybercriminals' profit centres. Many of these activities rely on a universal infrastructure, based on botnets. These are set up and operated for criminal purposes, and specialized to support cybercrimes, launch attacks, and cover up traces. This infrastructure is a cost centre for cybercriminals, yet the criminals do not pay the full price as the subverted PCs are paid for by their owners. Negative externalities are borne by society in the form of indirect and defence costs. These costs cannot for the most part be attributed to individual crimes and are discussed separately in Sect. 12.4.

12.3 What We Know

Few of the existing measures of cybercrime try to unbundle the different types of crime and categories of cost described above. In the following two sections, we summarize what is known. We also comment on the strength of evidence, and then

pull together the numbers below in Sect. 12.5. In our summary table we will only record crimes that impose costs in excess of $10 million per annum worldwide.

12.3.1 Online Payment Card Fraud

Bad things that happen on the Internet most commonly have a direct effect on real citizens when a charge they don't agree with appears on their credit card statement or bank statement. The UK Payments Administration, a payment industry trade association, publishes annual reports. Their most recent figure for 2010 puts card-not-present fraud in Internet transactions at £135 million for UK-issued credit and debit cards. The £135 million is a lower bound for the direct costs because even if these are accurate accounts of all the banks' costs and losses, the banks are not the only victims. There is unnoticed and unreported fraud, and there are wrongly denied claims (a hotly contested area). In these cases, the losses stay with the cardholder. Disputed transactions where a PIN was not used are routinely charged back to the merchant, and merchants may wrongly claim that a PIN was used to pass liability back to the bank (and ultimately the cardholder). Moreover, some fraud attempts get recouped; as well as cutting the banks' losses, this may add to the defence costs, while leaving the direct losses with middlemen (who are potentially victims, as in the case of money mules[4]). Overall, there are some grounds for suspicion that some of the drop in reported bank losses since 2008 is not due solely to better fraud prevention but also to more vigorous dumping of liability on merchants and cardholders.

Indirect losses are very difficult to quantify. They will have two major components: losses due to lack of confidence by consumers, and business foregone by merchants out of the fear of fraud. A rough proxy for the former is Eurostat's ICT survey, according to which 14 % of the UK consumers stated in 2010 that they refrained from buying goods or services online because of security concerns.Yet we must not scale the £27 billion online retail sales (9.5 % of total) in the UK[5] to the 8.7 million of lost online customers, because lost online sales are largely retained as offline sales. The benefits of online over offline sales include reduced consumer search costs and reduced distribution costs, both leading to lower prices as elasticity of demand increases and competition intensifies [39]. Only reduced search costs have a net effect on the economy and may guide estimates of indirect losses – disregarding other possible effects such as wider product differentiation leading to higher prices in the long run [33]. Considering that online merchants or distributors may more often be sited abroad than their offline counterparts, the domestic loss caused by forgone online sales is even lower. This would lead

[4]Money mule is a term for people hired by criminals to cash out stolen assets or engage in money laundering, often without knowing that their activity is illegal.

[5]Source: Office for National Statistics, 2011, total excluding automotive fuel

us to estimate the indirect costs of loss of confidence by consumers at perhaps $700 million.

As for caution on the part of merchants, there is a regular survey of online merchants carried out by Cybersource, a Visa company that does credit card processing [30]. Merchants reported lost revenues of 1.8 % of turnover, mostly to chargebacks, of which 32 % were ascribed to fraud; at the same time, merchants rejected 4.3 % of orders out of fear of fraud. This is double the figure of 30 basis points that capable online service firms expect to lose to fraud. Then again, Cybersource is trying to sell credit-card fraud prevention services, so perhaps it is not surprising their numbers are at the high end of the plausible range. Suppose that a capable firm might turn down valid orders amounting to 2 % of online turnover. A 2009 BCG report assessed the UK's digital economy at £100 billion or 7.2 % of GDP [27] but of this only half was actual online shopping. So we might rate the indirect losses at 1 % of the 'digital economy', or $1.6 billion, or a bit more than double what we might infer from Eurostat for the consumers. (We will gross up the two figures by somewhat less than the usual GDP multiplier of 20 to account for the fact that the UK has relatively weak protection for bank customers compared, for example, with the USA.)

As for defence costs, we will estimate the total for card and bank fraud at the end of Sect. 3.3, as the terminals used to accept cardholder-present transactions account for the largest single slice of the investment, and perhaps a third of the whole.

12.3.2 Online Banking Fraud

Online banking fraud is often conflated with payment card fraud, since both target the financial system and affect banks. However, we distinguish them for several reasons. First, the fraud is perpetrated differently. In online banking fraud, a customer's credentials (e.g., username and password) are obtained by a criminal, who then logs in to the account and initiates transfers to an intermediary who is cooperating with the criminal. Second, the fraud figures are collected separately although, unfortunately, the figures on online banking fraud are less reliable than for card fraud. Third, the parties affected are different: banks and their customers (both consumer and business) suffer online banking fraud, whereas with card fraud consumers, merchants and banks are impacted, and suffer significant administrative costs dealing with disputed transactions.

Online banking fraud is primarily carried out in two ways. In a phishing attack, criminals impersonate bank websites in order to get unsuspecting users to provide their login credentials. Several reports have investigated the revenues available to criminals for phishing. In one study, Moore and Clayton estimated that between 280,000 and 560,000 people gave away their credentials to phishing websites each year [44]. To arrive at a rough estimate of criminal revenue, the authors multiplied this figure by an estimate from Gartner that identity theft costs an average of $572

per victim.[6] Thus, their 2007 estimate was that criminals could earn $320 million per year from phishing. In a separate study, Florêncio and Herley studied when passwords were entered at unexpected websites and estimated that 0.4 % of the Internet population is phished annually [14].

The other modus operandi of online banking frauds is to install keystroke-logging malware. In October 2010, the FBI arrested a crime ring alleged to be using the prolific Zeus malware, which harvests credentials from many banks. The FBI claims that the criminals attempted to steal $220 million, and successfully stole $70 million, though it is not clear what the time frame was for the thefts [13].

In addition to large botnet-based malware such as Zeus, some criminals have started using spear-phishing to install targeted malware on machines used by small to medium sized businesses, typically targeting the CFO, the payroll department or the accounts payable department. According to the FBI, as of September 2011, they were investigating around 400 such cases of what might be called a corporate account takeover where criminals stole $85 million [50]. As these figures relate specifically to the USA rather than global losses (as was the case with Zeus) and as spear-phishing seems to be more developed against US targets than in Europe, we will estimate global losses at $300 million.

Many defence costs also apply to other forms of online crime – antivirus, malware remediation programs, and so forth. There are some online-banking specific efforts, though, most notably the take-down industry whose firms contract with banks to remove phishing websites that impersonate the real thing. There are also vendors of chip authentication programme calculators, systems for generating one-time passwords via mobile phones, and so on, which might account for two dozen booths at a trade fair like RSA. Their collective turnover might be estimated at $500 million globally. Adding in a similar amount for the banks' internal security development costs gives us an estimate of $1,000 million globally, or $50 million for the UK, for securing online banking.

Regarding indirect costs, Eurostat's survey suggest that security concerns keep 16 % of all individuals in the UK from carrying out online banking activities. An unofficial but plausible estimate puts the annual reduction of support cost at $70 per new online banking customer.[7] Combining both figures in a back-of-the-envelope calculation gives us a point estimate of £450 million in indirect cost for 2010, or $700 million, shared between UK consumers and banks. This estimate is highly uncertain. It might be an upper bound because we cannot rule out that a fraction of the 16 % stays away from online banking for more than one reason, and therefore might not adopt online banking even if security concerns were not an issue. Conversely, the cost savings seem to account for marginal costs only, i.e., assuming the bank maintains a network of branches anyway. If online banking

[6] This number is now considered extremely suspect. Florêncio and Herley have shown that this type of estimate has been regularly over-estimated by using small sample sizes and failing to deal appropriately with outliers [15].

[7] http://snarketing2dot0.com/2007/03/27/the-economics-of-online-banking/, 2007.

became pervasive, the savings from closing branches could be much higher than the losses, especially if some of the losses can be socialised.

Indeed, an important research question here is whether countries with stronger consumer protection laws actually have more profitable banks, because of higher trust leading to more online banking leading to reduced costs of branches, staff and cheque handling, as well as increased transaction fees. From this year, the 17 countries of the Eurozone are expected to publish uniform fraud statistics, and perhaps it will become feasible to get even bank CEOs to support stronger consumer protection. As UK consumer protection is lower than in the USA and the better Eurozone countries, we will scale up this notional economic loss from diminished confidence from $700 million in the UK to $10 billion globally rather than to $14 billion.

12.3.3 In-Person Payment Card Fraud

The cost of physical payment card fraud can be estimated by subtracting Internet fraud from Financial Fraud Action UK's total fraud figure for 2010, that is, £230 million for UK-issued credit and debit cards. However some doubt remains about certain categories over whether they should be considered – at least partly – to be online crimes or not. Some frauds are clearly physical, for example the £67.4 million ($106 million) due to face-to-face retail fraud; but this is still electronic (at least in the UK) as the great majority of card transactions are authorised online and use EMV. What is more, the common fraud mechanisms are technical in nature; villains use tampered PED terminals or ATM skimmers to capture card data and make forged cards that operate in fall-back mag-stripe mode. That accounts for the growth of counterfeit fraud in 2005–2008; the deployment of EMV led to many more terminals accepting PINs which increased the opportunity for rogue devices to steal card and PIN data. The industry's response was that fewer and fewer UK ATMs accept mag-stripe fall-back transactions. This in turn led crooks to cash out overseas; the fall in counterfeit UK transactions since 2008 has been matched by a growth in fraudulent overseas transactions, which amounted to £93.9 million ($147 million) by 2010. In most cases, these frauds are facilitated by online communications between colluding fraudsters across the UK and abroad. So it seems sensible to account for online and electronic fraud, a category that includes card fraud perpetrated in person.

The defence costs of EMV deployment are much harder to estimate. Retailers incurred significant expenditures in upgrading their terminal fleets; we have public figures from market leader Ingenico, which with 38 % of the retail terminal market booked $907 million of sales in 2010. This gives $2.4 billion for the total market, and from experience we would hazard a guess that the total cost of the systems are about three times that (when one adds the cost of integration, back-end systems and everything else). On the other hand, much of the cost of such systems can

be ascribed to providing functionality rather than security. We will estimate the defence cost of preventing card fraud as about equal to the cost of the terminals alone, namely $2.4 billion.

12.3.4 Fake Antivirus

Some cyber criminals altered the page contents of large numbers of web servers in such a way as to cause a visiting user to be presented with a pop-up window warning that their computer has been infected with malware, and that they should click OK to run antivirus software. When a user does click on the warning, fake antivirus software is installed that first disables any installed antivirus software and then issues repeated requests for payment. The only way to make these warnings go away is to pay up.

One group of researchers managed to get access to several internal databases run by three different criminal gangs perpetrating these crimes in 2008–2010 [52]. They found that the criminal groups had kept detailed records of conversion rates from installations to sales, along with the prices paid. The authors estimate that these three groups collectively earned $97 million per year. While there are probably more groups than the ones they studied, it is quite possible that these groups received the vast majority of the revenue attained from fake antivirus sales since the revenues from online crime are often concentrated among the most successful criminal groups.

12.3.5 Infringing Pharmaceuticals

Advertising is one of the key avenues for monetizing online criminal platforms such as botnets. Criminals use a range of advertising vectors, including sending unsolicited bulk email (i.e., spam), manipulating search engine results (e.g., so-called black-hat search engine optimization), and abusing social communications platforms (e.g., Twitter spam, blog spam, etc.) to attract users to sites selling a range of goods and services. Typically, these criminal advertisers are loosely organized independent contractors who are paid on a commission basis for any customers they bring in, by sponsoring so-called affiliate programs [49].

Unlicensed pharmaceuticals are perhaps the goods most widely promoted using criminal advertising.[8] Pfizer sells a genuine Viagra pill for several dollars while factories in India will provide a generic version to merchants for under a dime; this

[8] In one recent report Symantec's MessageLabs division reported that pharmaceutical advertising represented roughly 80 % of all spam email in June of 2010 [53] which is matched by similar data published by M86 Security from the same time period [40].

large margin for arbitrage has made "counterfeit" (more accurately, brand and/or patent-infringing) pharmaceuticals one of the best organized among underground businesses. Internet mail-order vendors support several thousand advertising affiliates and several dozen sponsoring affiliate programs.

Such activity incurs a range of indirect costs, many of them challenging to reason about in isolation. For example, pharmaceutical advertising dominates unsolicited bulk email and thus is a driver for the considerable expenses for anti-spam and content filtering products and services (a 2005 Frost and Sullivan report placed "World Content Filtering Revenues" at $1.31 billion [18]). But these are not the only threats prevented by these technologies; even if counterfeit pharmaceuticals were removed from the market (say by better IP enforcement in China and India) it is not clear that the advertising channel would not be repurposed, and hence still drive the anti-spam industry. Similarly, estimates of lost time and productivity due to unwanted email or poor search quality vary widely; they are both challenging to validate and to assign to a particular product category.

Looking at the direct risks to consumers, unregulated Internet pharmaceutical sales pose two potential liabilities: first, there is a risk of fraud – that a customer might order a drug and either not receive it or suffer subsequent charges to their credit card – and second there could be health risks due to poor quality and adulterated drugs. To the first point, there is very little empirical evidence to support the fraud hypothesis. Indeed, one of the authors' research groups has placed hundreds of pharmaceutical orders and has received products in all but a handful of cases and no unexplained fraud against the credit cards used [29, 35]. On the issue of health risks, while there are certainly concrete documented cases of individual harms [58], we are unaware of any systematic study of these risks or their overall costs to consumers. Moreover, recent studies of large-scale underground pharmaceutical programs documents that as much as a third of revenue is derived from *returning customers* — which seems unlikely if these customers were experiencing significant adverse reactions [41]. Of course, pervasive online availability of drugs such as opiates without prescription might exacerbate devastating addictions that impose substantial social costs.

Ironically, we possess better data about the criminal revenue brought in to the infringing vendors than about their direct or indirect costs to society. In one 2008 study, researchers manipulated the command and control channel of the Storm botnet to drive estimates of the underlying consumer conversion rate [28]. Extrapolating from this data, they suggested that this one botnet could drive revenues of $3.5 million per year for the pharmaceutical programs it advertised.

A recent 2011 study provides a broader scope and a tighter bound on such sales using a technique to infer order volume based on the allocation of customer service identifiers over time [29]. Whilst placing a series of undercover purchases from each affiliate program, the researchers discovered that each customer was given a unique order number and that this number was incremented for each new order, across all orders from all advertisers for the affiliate program. They inferred a monthly order volume exceeding 82,000 across 7 pharmaceutical affiliate programs. Using a

mostly conservative approximation for order size, they further extrapolated that monthly revenue from these programs is roughly $6 million.

However, a number of biases may affect this result. First, these seven programs clearly do not represent all organizations sponsoring infringing pharmaceutical sales online. For instance, researchers have found that other advertising vectors besides email spam, notably the manipulation of web search results, are also widely used to promote unauthorized pharmacies [34]. Nonetheless, the study estimating revenue does include the four largest affiliate programs, which represented over two-thirds of all email-advertised pharmaceutical URLs in a large-three month study of spam-based advertising [35]. Second, it does not account for differences in drug formulary. For example, sites selling restricted drugs such as steroids, opiates and stimulants can command a far larger revenue per order (one recent study demonstrated that carrying such drugs can double the revenue of a pharmaceutical program [41]). Finally, the methodology does not account for shopping cart abandonment just before the credit card is entered nor does it account for credit card declines (e.g., for fraud or insufficient funds).

Recently, a series of conflicts between major actors in the pharmaceutical affiliate program space has led to broad leakage of underlying financial records and transaction databases [31]. A recent 2012 study analyzing this data shows that, at its peak (2009), one of the largest pharmaceutical affiliate programs had annual gross revenues of $67 million (of which 6.4 % could be attributed to sales inside Great Britain) [41]. Moreover, by the contemporaneous order-number analysis mentioned earlier with this ground truth data for the same program, the authors were able to calibrate for biases due to declines and abandonment and establish an average revenue-per-order of $125. If we assume that these factors are consistent across the industry, then we estimate that the monthly revenue of counterfeit pharmaceutical sales from these top programs in 2010 was $8 million per month. Moreover, based on activity in both email and web advertising channels we believe that these programs represent at very least a third of organized counterfeit pharmaceutical activity online at the time. Thus, we believe that 2010 revenues from sales of online pharmaceutical was likely bounded by $24 million per month; $288 million for the year.

These same two studies, include geo-located customer order data, placing the fraction of such orders originating in the UK at between 3 % [29] and 6 % [41]. Thus, despite the unique characteristics of the large US market, it seems robust to continue using our estimate that the UK share is 5 %, reflecting the UK's share of world GDP.

Putting these measures together, we can estimate that UK consumers provided roughly $400,000 to the top counterfeit pharmaceutical programs in 2010 and perhaps as much as $1.2 million per-month overall.[9] To summarize, we will estimate

[9]Note, this revenue is split among a number of actors. Roughly 35–40 % is typically paid to advertisers (driving investments in botnets and other cybercrime support infrastructure), 20 % goes to suppliers and shipping, and 10–15 % goes to bank discount and agent fees for payment

UK-originated criminal revenue at no more than $14 million a year, and global revenue at $288 million.

12.3.6 Copyright-Infringing Software

The for-profit sale of counterfeit software, as with pharmaceuticals, is an advertising-based enterprise. The costs of distribution are negligible (particularly for online distribution) so the direct costs to criminals are primarily in sales acquisition (i.e., email spam, search engine optimisation, etc.) The societal costs are borne in the form of lost licensing revenues to copyright and brand holders (with the same confounding effects that we find in valuing losses due to counterfeit drug sales) while there are social benefits from people gaining value by using software that they would not have purchased at all at the full list price.

In 2004, the Business Software Alliance engaged Forrester Research to poll 1,000 Internet users in each of the US, UK, France, Germany, Canada and Brazil concerning their attitudes towards email spam advertising [17]. Over 20 % of UK respondents (versus 27 % for the entire group) responded affirmatively that they had purchased software advertised in this manner. If we took all such advertisements to represent counterfeit software organizations and the survey to be representative and accurate, then the online UK market for counterfeit software in 2004 would have been close to 12 million users.

A more recent 2011 study, using the empirical order number technique described earlier, estimated that three of the top five leading counterfeit software organizations together produced over 37,000 sales per month [29]. This metric is imperfect, since undoubtedly some of those orders did not complete and others were declined or refunded. Moreover, we do not know the monetary value of each such sale. However, if we assume the order rate estimate is correct and that the average software sale was $50, then this reflects an annual turnover of $22 million worldwide for these organizations. Given that software prices have fallen in the past 7 years (Microsoft's Office now costs tens of dollars rather than hundreds) and that ever more software is available free through cloud services, this fall should not be surprising.

12.3.7 Copyright-Infringing Music and Video

Disputes over the value of copyright-infringing music and video have been many and vociferous, with the music industry blaming the Internet for declining CD sales.

---

card processing. After a range of indirect costs, net revenue for operators is typically 10–20 % of gross [41].

We have to treat such claims with caution. First, copyright infringement performed by individuals (as opposed to for profit) is a civil matter in the UK and most other countries, so does not fall under the definition of cybercrime. Second, there has long been debate about whether illicit online copying actually depresses CD sales; an early study by Felix Oberholzer-Gee and Koleman Strumpf concluded that the people who did most file-sharing also bought the most CDs [47], while a thorough study by the Dutch government [23] concluded that copyright infringement through downloading pirated entertainment products gives a net social gain (for each dollar lost by the music industry, consumers gained two dollars' worth of value). On a broader scale, the transformation brought about by technology has meant that instead of people buying music an album at a time from a record company for £15 they now buy it a track at a time from Apple for 79p. Consumers get more music for their money and more musicians make a living. Job losses among music company middle managers are just the creative destruction inherent in technological progress.

As a result we do not think it is prudent to count the multibillion dollar claims of indirect losses made by music company advocates, but only the criminal gains made directly by gangs that operate downloading hubs. These are only in the hundreds of millions; for example, the recent raids on the Megaupload gang in Auckland who were claimed to be the world's largest led to asset seizures of the order of $50 million [10]. That site had 150 million users and 50 million daily visits; if we believe reports saying they had roughly a third of the market and that the $50 million represented a year's profits, then we get a global figure for proceeds of crime of $150 million.

12.3.8 Stranded Traveller Scams

Compromised webmail accounts are often used to send spam to the account owner's friends, a list of whom will be to hand in the address book. Spam blocking is often less stringently applied when there is regular communication, and the spam may leverage the social link, perhaps by providing a personal recommendation of a product. That aside, one of the most common uses of these compromised accounts is to operate the stranded traveller scam.

In this scam, an email is sent along the lines of:

> I write this with tears in my eyes. I had to travel to London at short notice and last night I was mugged at gun point. They have stolen all my cash, credit cards and mobile phone. Fortunately my passport and airline ticket was in my hotel room, but the manager will not let me check out until I settle my bill. Please will you spare me $1,900 to pay the hotel, I will reimburse you as soon as I get back.

If the recipient of the email is taken in, they will be instructed to send the money by Western Union. The sender may be reassured because they believe the money can only be picked up in London by the holder of an appropriate passport, but in practice the dollar amount will be below the limit for which government identity documents

are needed, and the money can be picked up anywhere in the UK. A detailed account of the scam, from a victim's viewpoint, is given by Fallows [12].

The scammers exploit all the main webmail platforms, AOL, Gmail, Hotmail and Yahoo! along with Facebook. One of this chapter's authors has unpublished 2010 data obtained by examining customer support reports to one of these companies. On average, the criminals were receiving one or two payments a day. Scaling up across the five platforms, and assuming (fairly arbitrarily) that only one in two losses was mentioned to the webmail company, means the annual turnover for this scam is approximately $10 million.

Assigned a loss to the UK is complex. Most of the victims were from the US and in 2010 most of the money was flowing towards the UK – although there are clear reasons to suppose its final destination was West Africa. So although this scam is relatively high profile, and makes for a good anecdote, the loss to the UK is most unlikely to exceed $1 million per annum.

12.3.9 Fake Escrow Scams

Another widely discussed but relatively uncommon scam is 'fake escrow'. Here the victim believes that they have won an online auction for a car or motorbike. The seller proposes that to safeguard both their interests they should use a third party escrow agent, which conveniently for the purchaser will also deliver the vehicle to their door. The seller will give the vehicle to the escrow/delivery company and pay the fees. The purchaser will pay money to the escrow company who will pay the seller and deliver the vehicle. However, despite having a convincing website with an online delivery tracking system, the escrow company is a sham and the putative purchaser will be perhaps $10,000 out of pocket.

There are around 100 active fake escrow websites at any given time [26], and the more competent fraudsters are believed to be faking the sale of one car a week ("Sodano": Personal communication 2007). This puts the overall turnover in the region of $200 million per annum. The scam operates to some extent in North America, but rather more in Europe. UK losses may be of the order of $10 million a year.

12.3.10 Advanced Fee Fraud

Advanced Fee Fraud (AFF) is sometimes called *419 fraud* after the relevant article of the Nigerian criminal code. It comes in a large number of formats, from the deceased dictator's family who want to smuggle millions of dollars, to scams where people win millions in lotteries they have never entered. The common feature of all of these frauds is that the victim must pay out a small amount of money (a tax, a bribe or just a bank account opening fee) in the expectation that this will release

the large sum to them. If they pay out once then some other obstacle will arise and they will need to provide another advance fee – in extreme cases until they are personally bankrupt, or if they are repurposing their employer's funds, until their own fraud becomes apparent.

There are very strong links historically between AFF and West Africa, particularly Nigeria, going back to the days when it was conducted by letter and then fax. Email has made communications simpler, although the higher-value scams often involve face to face meetings, and occasionally even kidnapping – so at the top end, this is not purely a cybercrime.

As usual, figures are hard to come by, but in a 2006 Chatham House report [48], Peel set out a detailed account of Nigerian financial crime, which extends far beyond what happens in cyberspace. He quotes a 2004 CIFAS web page setting the cost to the UK economy at £150 million, but consulting the original (at archive.org) shows that the data on average loss (£31,000) came from a 2001 NCIS press release and – according to CIFAS – the same 2001 release gave the £150 million figure as the losses for 2003 (which is presumably a typo).

Although it is possible to identify from press reports many individual cases of large losses, these are mainly in the US and they are scattered over many different years. There does not appear to be any reliable data at all on the overall loss to the UK in any particular year, with what figures there are being a summation of all cases worked by a particular police force, and perhaps then multiplied up to speculatively account for under-reporting.

In practice we suspect that the majority of losses in purely cyber frauds are relatively small, having ourselves seen initial demands in lottery frauds of only £800. The prevalence and diversity of AFF emails in spam does however suggest that many criminals consider this a worthwhile line of activity. The higher profile of this fraud generally also makes it seem likely that it is more lucrative than the stranded traveller and fake escrow scams just discussed. So to avoid a gap in our tables we will pick a number of $50 million for UK losses, but we would be the first to admit that this figure is merely indicative and we have no real evidence to support it. We suspect it is rather on the high side.

12.3.11 PABX Fraud

The Communications Fraud Control Association (CFCA) publishes data on fraud losses associated with telephony, both fixed and mobile. Their methodology is to survey experts from within the industry as to what proportion of turnover is lost to fraud, and – with some statistical adjustments to account for company size – thereby estimate the size of the problem. Their headline figure for 2011 is $40 billion [7]. Their methodology leads to some bizarre results – the overall loss is down by a third from their previous 2008 report, but 98 % of the people they surveyed believed that fraud was static or increasing. However, their members do report real losses, and

just 34 members stated that their companies had collectively lost $2 billion in the previous year.

The CFCA data distinguishes a range of crimes, from manipulation of the SS7 signalling system to hide the identity of a caller, through clip-on fraud (physically connecting to someone else's phone line), to straightforward subscription fraud – failing to pay the bill. Of particular interest is PABX fraud. The criminals reconfigure a company's telephone system (Private Automatic Branch Exchange) to accept incoming calls and relay them onward. They then sell phone cards (which can be little more than instructions on how to dial) to expat workers, who then call home at the company's expense. This crime is decades old, and was once done by accessing a modem within the PABX that was used for remote maintenance, but PABXs are now placed on the Internet – and are often left with weak or default passwords.

The CFCA estimate for PABX fraud is $4.96 billion worldwide, $1.28 billion of which they believe occurs in Western Europe (so the UK share would be in the region of $185 million). The CFCA does not set out whether this is the wholesale or retail cost of the calls – defrauded companies can often renegotiate the actual payment they make to settle their unexpected bill.

12.3.12 Industrial Cyber-espionage and Extortion

Following the Detica report, UK government spokespersons have talked up the risk of espionage. One of them warned at a conference in Cambridge that the university had better invest more in cybersecurity or see its priceless intellectual property stolen. But Cambridge does not own priceless intellectual property; academics own the copyright in their own publications and software, and while the university does have the right of first refusal on patents that staff members choose to file, such patents cannot be stolen once filed.

Similar comments can be made of other companies with valuable IP: drug firms' new products may be vulnerable prior to filing, yet we are unaware of any case where a filing has been spoilt by unauthorised prior exposure. As for firms with valuable software, it is common for source code to be very widely available. Microsoft has tens of thousands of engineering staff with access while large numbers of outside organisations (from the Chinese government to some researchers at the University of Cambridge) have access to source code for Windows under a non-disclosure agreement. The Detica claim of £9.6 billion of annual losses by "companies that create significant quantities of IP or whose IP is relatively easy to exploit" has no obvious foundation.

The second part of Detica's claim is £7.6 billion involving the theft and exploitation of non-IP related data such as companies involved in open tendering competititons or which can be affected by large share price movements. Again, there is no obvious foundation for this; however stock markets do have mechanisms to detect suspicious trades in advance of price-sensitive announcements, and if a leak from one company to another causes a tender for a public-sector IT project to be

priced more keenly than would have otherwise been the case, then it is entirely unclear that the public is thereby harmed. Indeed there are frequent complaints about the oligopoly of large firms that win most public-sector business and there are officials at the Cabinet Office – the co-author of the Detica report – whose job it is to promote SME competition for public-sector supply and service contracts.

A third part is the claim that £2.2 billion per annum is lost to extortion, with the comment that "we believe this type of cybercrime goes largely unreported." This is a very old and persistent claim made by security salesmen. One of us (Anderson) recalls working for a bank a quarter century ago and hearing it; even when it was truthfully denied, the salesmen persisted "we know it happens but you're not allowed to tell anyone" until escorted to the door for impertinence.

Extortion does occasionally happen – there was a widely reported case in 2004 when DDoS was used against online casinos and $4 million was paid before the gang was arrested [38] – but like kidnapping, extortion is a hard crime to get away with, as money-laundering is not trivial when the sender of the funds wishes to track down the recipient and is supported by the police.

In sum, because there is no reliable evidence of the extent or cost of industrial cyber-espionage and extortion, we do not include any figures for these crimes in our estimates.

12.3.13 Fiscal Fraud

The Detica report also includes "fiscal fraud committed against the Government" in its assessment of cybercrime. Certainly much tax and welfare fraud is committed by citizens who misrepresent their circumstances, and the UK (like other countries) is moving both tax filing and welfare claims online. Detica claims £2.2 billion of fraud across tax, welfare, pensions, the NHS, other central government functions, and local government. They ascribed all such fraud to the cyber category given that so many claims are now made online. That figure may well be justified but such fraud is nothing new.

In the USA, the IRS has made a determined effort to crack down on phishing gangs that impersonate it in order to trick people out of tax refunds, while in the UK, Her Majesty's Revenue and Customs (HMRC) appears to have made rather limited efforts. Following a news report that tax refund fraud was costing HMRC £600 million a year [54], which was largely stolen online by foreign cyber criminals, we persuaded the MP for Cambridge, Dr. Julian Huppert, to ask the following Parliamentary Question:

> Dr. Huppert: To ask the Chancellor of the Exchequer what estimate he has made of the cost to the public purse of payments for tax refunds being fraudulently redirected as a result of websites that impersonate Government websites in the last three financial years. [86764]

The response, on behalf of the Chancellor of the Exchequer, was unhelpful:

> Mr. Gauke: HM Revenue and Customs do not have an estimate of the cost of tax refund payments being fraudulently redirected as a result of websites that impersonate Government websites.

Rather than impersonate the tax office, criminals in the USA have been impersonating citizens by electronically filing fraudulent tax returns using stolen lists of names and Social Security numbers. The IRS claims the problem is widespread and growing. By their own estimation, in 2010 around 1.5 million tax returns were fraudulently filed, garnering refunds totalling $5.2 billion [1]. By contrast, British tax authorities have taken steps to authenticate citizens who file returns electronically, physically mailing out passwords to the address on record that must be produced in order to file online.

In an off-the-record conversation with a senior civil servant, we learned that welfare cheating in Britain is 0.8 % of the total expenditure, which is over £160 billion. This figure is robust; they have done frequent drill-downs. It is overwhelmingly about misrepresentation of circumstances (undeclared partner/income/capital) and some of it is not even malicious (particularly elderly people who had put away a nest egg for a specific purpose and did not think it part of their capital). But most is fraud, and the rates vary widely from the state pension (under 0.1 %) to means-tested benefits (over 4 %). All of it will be computer crime from 2013 when all claims will be done online; most of it already is. But it is almost unchanged from a few years ago when claims were in person. So welfare fraud adds £1.2 billion to Britain's fraud figures.

Tax evasion is more slippery. Her Majesty's Revenue and Customs believe it is 2 % but the figure is not robust and has been the subject of much internal debate in the civil service: different departments try to play it up, or down. As a result the numbers keep getting referred to the Office of National Statistics, who keep quibbling. This too is all rapidly "becoming computer crime." That will add a further £8 billion. In passing, to put some of the numbers into context, we will note that so-called Missing Trader VAT fraud or Carousel VAT fraud[10] is estimated to have cost the UK £3 billion in 2005/2006 [22]. At that time, this fraud had no substantial cyber component at all. Yet almost all VAT returns are online from this year (2012). Similarly, most income tax fraud is about misrepresentation of circumstances, income or capital, just like welfare fraud; the cyber component about which Mr. Gauke has no information might amount to a few hundred million. For consistency we will put down the cost of tax fraud as £8 billion or $12 billion, which includes income tax, VAT and corporate taxes too – but we will put both welfare fraud and tax fraud separately at the bottom of the table to remind the reader that

[10]In a VAT fraud goods such as mobile phones are imported into the UK. They are sold on within the UK, VAT paid on this sale (15 to 20 % at various times) which should be paid to the taxman is pocketed by criminals who shut down the company and disappear. The recipient can now re-export the phones and claim back the VAT (doubling the criminal's take) – and the same batch of goods can then be cycled round again and again, hence the carousel term.

these are the figures for largely traditional frauds that now, because of electronic filing, fall within the EU definition.

12.3.14 Other Commercial Fraud

There are other types of commercial fraud from insider trading to embezzlement which will eventually, like fiscal fraud, be undeniably cyber. Perhaps the largest category is control fraud, where the executives in charge of a company (or the ministers in charge of an economy) abuse their authority to loot it. There have been cases of control fraud with a cyber element, such as the Equity Funding affair, but we have decided to exclude such frauds from this report for brevity as they are mostly concerned with the exercise of power in interpersonal and institutional relationships rather than by means of claims made relatively formally and mechanically through automated systems. What is more, in the case of control fraud (at least) the known countermeasures are not technical but concerned with institutional mechanism design – such as awarding company executives sufficient stock options to align their incentives with shareholders, and encouraging less developed countries to hold formal second-price auctions for access to their natural resources rather than permitting ministers to strike deals privately with foreign companies.

12.4 The Infrastructure Supporting Cybercrime

We now review the infrastructure supporting cybercrime. While these activities are often referred to directly as cybercrime, in fact they are used to enable lots of different crimes. Consequently, we estimate the infrastructure's costs separately so as to avoid double counting.

12.4.1 Botnets

Botnets are a key part of the infrastructure for cybercrime. A botnet is a network of thousands, sometimes even millions, of machines that have been infected with malware, putting them under the remote control of criminals. The botnet herders who assemble these collections of machines may either use them for crime directly or rent them out to others to operate. The operator will typically send instructions to his infected machines to download further malware to implement specific attacks. Botnets provide a versatile platform for a variety of criminal business models, including sending spam, committing click fraud, harvesting account credentials, launching denial-of-service attacks, installing scareware and phishing.

It is the criminal business models that generate the revenue, but because the botnets provide a platform for them, they have value within the criminal economy, as we can see in underground markets where botnets are sold and rented. The cost of the criminal business models that use botnets are discussed elsewhere in this chapter; as for the botmasters' turnover, Herley and Florêncio estimated in 2009 an upper bound to botnet herders' income of 50c per machine per annum, so that a 20,000-machine botnet earned its herder some $190 a week [20]. That was an upper bound, and botnets seem to have become bigger since then; if we assume 50 million bots worldwide, we will estimate annual herder income in the single millions per year, and that is too low a figure for us to include it in our summary table.

There are also the costs the botnets themselves inflict on society. These losses occur first and foremost in the cost of dealing with the infected machines. The costs are distributed across different actors, most notably the owners of those machines, ISPs (in the sense of access providers) and hosting providers. Recent empirical analysis showed that around 80 % of all infected machines are located in the networks of ISPs [56]. Of the remaining 20 %, the largest share could be attributed to infected servers at hosting providers. This picture is evolving, however, and may increasingly include mobile devices.

12.4.2 Botnet Mitigation by Consumers

There are only snippets of evidence about the costs associated with botnet mitigation. Currently, there is no single authoritative data source to identify the total population of infected machines around the world, and each source has its own strengths and weaknesses.[11] Two robust and independent estimates both suggest that a little over one million British households have had a machine in a botnet at least once per year. The first estimate comes from Microsoft, which they say is based on telemetry from over 600 million machines worldwide. In the first half of 2010, their Malicious Software Removal Tool cleaned up around 500,000 bots in the UK [43]. They did not produce a similar number for the second half, but overall malware infections in the UK did rise slightly compared to the first half, so it is likely that the total for the year lies just over one million [42]. Following a completely different measurement approach, a study of the Dutch market, done in close collaboration with the ISPs, compared Dutch infection levels to those in the UK and other countries [57]. In those metrics, the UK is in the same ball park as the Netherlands: in 2010, around 6 % of the 19 million UK broadband subscribers had

[11]Many estimates of the total number of machines that are affected have relied on counting the number of unique IP addresses that show up in botnet activity. We now know that this overestimates the size of the problem, often by an order of magnitude. Dynamic IP address allocation can cause the same infected machine to show up under many different IP addresses. When the Dutch police took down Bredolab, it was claimed that the botnet had infected 30 million machines in about 2 years. In reality, the evidence suggests that it was around three million machines.

a machine in a botnet at some point during the year. That translates to 1.1 million subscribers.

It is much less clear, however, what cleaning up these one million infected machines costs. Many infections are cleaned up by the generic countermeasures of automatic security updates and anti-virus software, without much user involvement. Those would fall under indirect costs. We do not really know how often users clean up machines themselves and at what cost. In a 2007 survey by Consumers Union, a US consumer protection organization, residential users reported to their total repair cost from malware were around US $5 billion in the past year – or around $100 per household/broadband subscriber [8]. In addition to clean-up, this includes the cost of replacing a poorly functioning, malware-ridden computer with a new one, even though it seems incorrect to fully attribute the price of a new machine to the problem of malware. Furthermore, the reliability of these surveys is highly questionable, a problem which is exacerbated by the fact that underlying data and methodology are not made public. Furthermore, the cost of PCs has fallen sharply in the last 5 years, from almost a thousand pounds to a few hundred. We will therefore be ultra-cautious and estimate a rough current equivalent for the UK of $500 million. To put this in perspective, it sets the average clean-up cost per household per year at less than the cost of 1 h of end-user time valued against the UK GDP per capita.

12.4.3 Botnet Mitigation by Industry

Another loss is borne by ISPs and hosting providers, who may have to act against infected machines in their networks.[12] The scale at which such mitigation efforts take place is highly variable. Finnish ISPs, for example, act on most infected machines that are detected, whereas in UK, as in most other Western countries, ISPs only tackle a small fraction of the infected population. We have no data on how much these efforts cost. To some extent they can be automated, but the largest cost is customer support. A recent German initiative set up a national call centre to deal with botnet mitigation. In the first year of its operation, it notified 315,518 end users that they owned an infected machine [19] – a fraction of the infected population. Most notified users ran an automatic clean-up tool; less than 1 % of them needed customer support over the phone, with the average call lasting for about 15 min. Operating the centre cost about 2 million euro in funding for its setup and first year of operations. Note that this cost does not cover the whole problem, as comparative infection measurements suggest the centre notified around a quarter of the infected population. If all infected customers were to be notified, the cost would rise, but still remain well under 10 million, as we expect that the marginal costs of notifying and

[12]Some of these costs are directly related, and proportional, to the infection of machines. ISPs also bear indirect losses and more general defence costs, which do not vary with the outbreak of botnet infections.

helping additional users to diminish. While there are no comparable figures for the UK, it seems reasonable to assume that the costs there are similar or lower, in other words: in the single millions per year. A more fully developed mitigation strategy, like the one in Finland, is probably in the same ballpark. It may cost more at first, but automation can take over part of the process and, once the infection rate diminishes, so do the support costs of customers.

A further cost is that of botnet mitigation by commercial firms other than service providers (and banks and retailers, whose $7 billion anti-fraud measures we account for above). Figures for the total information security industry are difficult, with some reports suggesting of the order of $20 billion [6] or even more. Symantec alone has turnover of $6 billion, but that is everything, not just their antivirus business; we will estimate antivirus expenditure worldwide at $3.4 billion. It is a bit harder to reckon how much of industry's infosec costs to ascribe to generic defences (over and above the specific payment-system defences deployed by banks and merchants); sysadmins mostly do other things than security, and internal controls have other purposes than limiting the damage an infected machine can do. In order to be ultra-cautious, we will ascribe a global figure of $10 billion to generic cybercrime defences by companies worldwide. This sets the corporate costs of mitigating botnets to be equal to the costs borne by individuals. We don't know whether this is right; corporates are more concerned about security, but also more efficient at providing it. So we will hazard a figure of $10 billion, bearing in mind that only the order of magnitude is probably right.

12.4.4 Other Botnet Mitigation Costs

The indirect losses due to botnets are much more dispersed; many losses mentioned elsewhere in the chapter are relevant here as well. A more specific effect is that malware may motivate end users, platform owners and service providers to move away from general purpose computers towards more controlled platforms, such as Apple's model of iOS and the App Store. This may help innovation in the short run, but in the end it might reduce innovative capacity via locked-down devices, new concentrations of market power and less user autonomy [60]. The economic effects of this are unpredictable, though; if Apple wins at the expense of Microsoft, it is best to see this as the normal operation of capitalism rather than trying to interpret it as a security issue. After all, Android in turn is eating some of Apple's market. That said, if the security issues support an overall trend towards more restrictions on end users devices, this may reduce the potential for innovation and generate substantial societal opportunity costs.

Defence costs are more straightforward; many actors in the ecosystem bear them. End users, ISPs and software vendors all invest in technical measures to protect against infection. While numerous problems remain, end-user adoption of antivirus software is actually very high. The cost of this software may be borne by end users, by access providers (who bundle it with the subscription) or by platform vendors like

Microsoft, who incorporate it in the platform. There are also vendors of antivirus solutions that provide them for free to home users, and cross-subsidise this from corporate licenses. The total cost of antivirus defence mechanisms is unknown, but we could assume from the Eurostat 2010 ICT survey that 88 % of all households with a broadband subscription use at least one of these products. A conservative estimate would put the worth of a single license at $10, ignoring for a moment which actor actually bears this cost. For the UK, these assumptions estimate the total cost of antivirus countermeasures at around $170 million.

Other general defence costs are more difficult to estimate. Software vendors constantly patch their products against vulnerabilities that can be exploited by malware. Anecdotal evidence suggests that for mission-critical software, such as enterprise databases, the cost of a single patch development cycle can run up to a million dollars [55]. Similarly, the deployment of the patch within companies is also costly. By way of illustration: every time Google patches the kernel on the workstations of its employees, the subsequent reboot of the machines costs them over $1 million in lost productivity [4]. Further deployment costs include testing and assuring the patches before rolling them out. All of this suggests that for the whole software market, the cost of patching will be in the hundreds of millions, probably more. However, the part of this cost that can be attributed to the UK will probably be at most its share in global GDP, as its software industry is proportionally smaller than that in the USA. If we assume, for illustrative purposes, that the global cost of patching is $1 billion per year, this would mean the UK bears $50 million of this. This does not include the costs of deployment, which are borne by the end users.

Finally, we should also mention the cost of law enforcement as a general defence cost. Investigations aimed at taking down botnets and prosecuting the criminals behind them are very time-consuming and require costly specialists. That being said, few cases are investigated in depth. In the last 2 years, there have been only a handful of botnet takedowns: Kelihos, DNS Changer, Rustock, Pushdo/Cutwail, Bredolab, Coreflood and Waledec. None of these seem to have had a substantial involvement from UK law enforcement agencies. We understand that the UK police have received an extra £30 million for the next 4 years which will let them open three regional centres to support cybercrime and forensic investigations: say another $10 million raising UK police cyberbudgets to $15 million a year. Meanwhile the US spends about $100 million at the federal level (FBI, Secret Service, FTC and NCFTA) and we may assume the same again at state level. The US is by far the major player in cyber enforcement, and seems to do about half the work; so we will estimate global law-enforcement expenditures at $400 million.

12.4.5 Pay-per-Install

Wondracek and colleagues studied links between the pay-per-install business and the porn industry [59]. A pay-per-install operator is a criminal who infects PCs to

order; for example, they charge $130 to install malware on 1,000 PCs in the USA (the prices for Asia are much lower – as low as $3 for China). This study unearthed a whole ecosystem of shady services, with business links between adult pay sites, free sites, link collections, traffic brokers, search engines and redirector services. As another example, $160 bought 49,000 visitors, of whom more than 20,000 were vulnerable to at least one known vulnerability. They concluded that although not all porn sites are crooked, many are; the underground economy is a major financier of the adult business. A further study by Caballero, Grier, Kreibich and Paxson found that 12 of the world's top 20 malware families used PPI services for distribution [5].

If 50 million machines are in botnets, with an average infection duration of 6 months, then 100 million machines are infected every year. If half of these are done by PPI firms at an average cost of $50 per thousand, that is a turnover of $2.5 million which lies below our reporting threshold. In fact, Caballero and colleagues caution that there's little point in cleaning up a botnet if the herder can rebuild his asset by very modest payments to PPI services.

12.5 Fitting the Estimates into the Framework

Previous studies of cybercrime have tended to study quite different things and were often written by organisations (such as vendors, police agencies or music industry lawyers) with an obvious agenda. The subject is difficult because definitions are hard; much fraud that used to be conducted on paper or face-to-face (such as tax and welfare fraud) is now online and these traditional frauds are much larger in volume and value terms than the new purely computer frauds. Also, there is a significant amount of fraud in between the traditional and the new, such as payment card fraud. This type of fraud began to change 20 years ago, but the move online and the transition from magnetic-stripe to EMV technology have quite changed the modus operandi. We have called this *transitional* fraud for want of a better name.

In this report we have gone through the main types of fraud, whether traditional, transitional or modern. For each modus operandi we have collected the best figures from current research, and where none were available we have done what we could to provide neutral estimates. We collate our estimates here in Table 12.1. The numbers in bold are the ones we observed or estimated directly, whether UK figures or global ones; where we have only one of the two, the other is scaled on the basis that the UK is 5 % of the world by share of GDP. In some cases, the scale is different for specific reasons that are mentioned in the text.

While we are relatively happy to scale down global fraud figures to obtain UK-specific estimates, we urge caution in interpreting the global estimates where we could only extrapolate from UK figures. Extrapolating from a sample representing 5 % of world GDP can exaggerate the impact of local variation in fraud. The media claim, for example, that tax fraud is higher in Greece, and medical benefits fraud higher in the USA; we do not try to investigate such variance here. Ideally, global

Table 12.1 Judgement on coverage of cost categories by known estimates

Type of cybercrime	UK estimate	Global estimate	Ref. period	Criminal revenue	Direct losses	Indirect losses	Defence cost
Cost of genuine cybercrime							
Online banking fraud							
– Phishing	$16 million	**$320 million**	2007	$\times^{?}$	$\times^{?}$		
– Malware (consumer)	$4 million	**$70 million**	2010	$\times^{\downarrow}$	$\times^{\downarrow}$		
– Malware (businesses)	$6 million	**$300 million**		$\times^{\downarrow}$	$\times^{\downarrow}$		
– Bank tech. countermeasures	$50 million	**$1,000 million**	2010				$\times^{?}$
Fake antivirus	$5 million	**$97 million**	2008–2010	×	×		
Copyright-infringing software	**$1 million**	**$22 million**	2010	×	×		
Copyright-infringing music etc.	$7 million	**$150 million**	2011	$\times^{\downarrow}$			
Patent-infringing pharma	**$14 million**	**$288 million**	2010	×			
Stranded traveller scam	$1 million	**$10 million**	2011	$\times^{\downarrow}$			
Fake escrow scam	$10 million	**$200 million**	2011	$\times^{\downarrow}$			
Advance-fee fraud	**$50 million**	$1,000 million	2011	$\times^{\downarrow}$			
...							
Cost of transitional cybercrime							
Online payment card fraud	**$210 million**	$4,200 million	2010		(×)		
Offline payment card fraud							
– Domestic	**$106 million**	$2,100 million	2010		$\times^{\downarrow}$		
– International	**$147 million**	$2,940 million	2010		$\times^{\downarrow}$		
– Bank/merchant defence costs	$120 million	**$2,400 million**	2010				$\times^{\downarrow}$
Indirect costs of payment fraud							
– Loss of confidence (consumers)	**$700 million**	$10,000 million	2010			$\times^{?}$	
– Loss of confidence (merchants)	**$1,600 million**	$20,000 million	2009			$\times^{?}$	
PABX fraud	**$185 million**	**$4,960 million**	2011	×	$\times^{\downarrow}$		
...							
Cost of cybercriminal infrastructure							
Expenditure on antivirus	**$170 million**	$3,400 million	2012				×
Cost to industry of patching	$50 million	**$1,000 million**	2010				$\times^{?}$
ISP clean-up expenditures	**$2 million**	$40 million	2010			$\times^{?}$	
Cost to users of clean-up	**$500 million**	$10,000 million	2012			$\times^{?}$	
Defence costs of firms generally	$500 million	**$10,000 million**	2010				$\times^{?}$
Expenditure on law enforcement	**$15 million**	$400 million	2010				×
...							

(continued)

Table 12.1 (continued)

Type of cybercrime	UK estimate	Global estimate	Ref. period	Criminal revenue	Direct losses	Indirect losses	Defence cost
Cost of traditional crimes becoming 'cyber'							
Welfare fraud	**$1,900 million**	$20,000 million	2011	×	(×)		
Tax fraud	**$12,000 million**	$125,000 million	2011	$\times^{?}$	(×)		
Tax filing fraud	–	**$5,200 million**	2010	×	(×)		
...							

Estimating costs and scaling: Figures in boldface are estimates based on data or assumption for the reference area. Unless both figures in a row are bold, the non-boldface figure has been scaled using the UK's share of world GDP unless otherwise stated in the main text. Extrapolations from UK numbers to the global scale should be interpreted with utmost caution. A threshold to enter this table is defined at $10 million for the global estimate

Legend: ×: included, (×): partly covered; with qualifiers $\times^{\uparrow}$ for likely over-estimated, $\times^{\downarrow}$ for likely underestimated, and $\times^{?}$ for very high uncertainty

surveys could be undertaken for the categories where we lack global estimates. Nonetheless, we include the extrapolated figures in the table when necessary to aid policymakers while such data is not available.

Readers may be wondering why the table does not include any totals. It is after all a simple matter to add up the "Cost of genuine cybercrime" section to give a $170 million figure. But since this value is less than the reported cost of card fraud in the very next row of the table, and that in turn is dwarfed by many other figures in later rows – and many of these are extremely rough estimates – we believe it is entirely misleading to provide totals lest they be quoted out of context, without all the caveats and cautions that we have provided.

Our work has its limitations. Inter alia, it gives a static view of the economics of cybercrime while the dynamics also matter. The development of the dark market in carding data, crimeware, botnet rental and other illegal services has, as we have just noted, made only a small contribution to the total figure we have presented. Nevertheless, it has enabled significant growth in other crime categories in the period since such markets got organised in the mid-2000s. However we believe that our work is a principled start to being able to measure the cost of cybercrime. We propose to continue updating our estimates, and to produce new versions of this chapter every few years.

12.6 Conclusion

The data we have collected indicates that, in terms of the measurable costs:

- Traditional frauds such as tax and welfare fraud cost each of us as citizens a few hundred pounds/euros/dollars a year.[13] With such crimes, the costs of defences, and of enforcement, are much less than the amounts stolen.
- Transitional frauds such as payment card fraud cost each of us as citizens a few tens of pounds/euros/dollars a year. Online payment card fraud, for example, typically runs at 30 basis points, or 0.3 % of the turnover of e-commerce firms. Defence costs are broadly comparable with actual losses, but the indirect costs of business foregone because of the fear of fraud, both by consumers and by merchants, are several times higher.
- The new cyber-frauds such as fake antivirus net their perpetrators relatively small sums, with common scams pulling in tens of cents/pence per year per head of population. In total, cyber-crooks' earnings might amount to a couple of dollars per citizen per year. But the indirect costs and defence costs are very substantial – at least ten times that. The clean-up costs faced by users (whether personal or corporate) are the largest single component; owners of infected PCs may have to spend hundreds of dollars, while the average cost to each of us as citizens runs in the low tens of dollars per year. The costs of antivirus (to both individuals and businesses) and the cost of patching (mostly to businesses) are also significant at a few dollars a year each.

This brings us to an interesting question. Traditional acquisitive crimes, such as burglary and car theft, tend to have two properties. The first is that the impact on the victim is greater in financial terms than either the costs borne in anticipation of crime, or the response costs afterwards such as the police and the prisons. For example, Canada estimated victim costs of $47 billion, criminal justice system costs of $13 billion and defence costs of $10 billion across its economy as a whole in 2003 [25]. Other countries use different measures but the broad picture is similar.

Drilling down into the victim costs, we find that for nonviolent crimes the value of the property stolen or damaged is much greater than the cost of lost output, victim services or emotional impact. With the new cybercrimes, the pattern is much more like robbery, where (according to UK Home Office 2005 figures) only £109 is stolen but the additional costs include £483 for health services, £1,011 for lost output, and a whopping £3,048 for distress [21].

It is worth noting that the criminal justice system recognises the quite disproportionate social costs of robbery as opposed to burglary; the typical robbery incurs £2,601 of criminal justice system costs compared with a typical burglary, where a much larger amount (£846) is stolen and yet less than half as much (£1,137) is spent

[13]The precise choice of currency isn't important given the accuracy of the figures available to us; we can be reasonably sure we have got the orders of magnitude right, and often the binary order of magnitude, but not much beyond that.

on justice. Yet while robbers get longer sentences than burglars do, cyber-crooks get shorter ones. This is probably because cyber-crimes, being impersonal, evoke less resentment and vindictiveness. Indeed, the crooks are simply being rational: while terrorists try to be as annoying as possible, fraudsters are quite the opposite and try to minimise the probability that they will be the targets of effective enforcement action.

Why does cyber-crime carry such high indirect and defence costs? Many of the reasons have been explored in the security-economics literature: there are externalities, asymmetric information, and agency effects galore. Globalisation undermines the incentives facing local police forces, while banks, merchants and service providers engage in liability shell games. We are also starting to understand the behavioural aspects: terrorist crimes are hyper-salient because the perpetrators go out of their way to be as annoying as possible, while most online crooks go out of their way to be invisible. The possible policy remedies have also been discussed at length, from better statistics to better international cooperation [2, 45]. But what is the priority?

The straightforward conclusion to draw on the basis of the comparative figures collected in this study is that we should perhaps spend less in anticipation of computer crime (on antivirus, firewalls etc.) but we should certainly spend an awful lot more on catching and punishing the perpetrators.

A final point is that, according to the British Crime Survey, some 2 % of respondents reported suffering a traditional acquisitive crime such as burglary or car theft, while more than double that number suffered fraud. The survey did not disambiguate the online and electronic frauds of interest here from the door-to-door and boiler-house variety, but the former probably accounted for most of it. A special module on IT security of the 2010 Eurostat ICT survey completes this picture. Ranking all 27 EU countries by online user's concerns, the UK ranks sixth for virus infections, fourth for spam, and second behind Latvia for the three remaining threats: personal data abuse and privacy violation; financial losses caused by phishing and pharming; and financial losses due to fraudulent payment card use. When looking at the self-reported actual experience of threats, the picture becomes more differentiated. On the one hand, UK residents exactly match the EU average for virus infections and privacy threats, and they seem to receive less spam than the average European. On the other hand, the UK ranks second for financial losses caused by phishing and pharming attacks, and first for payment card fraud, which affected 5 % of the UK's online population.

If this interpretation is correct, then cybercrime is now the typical volume property crime in the UK, and the case for more vigorous policing is stronger than ever.

References

1. Alvarez, L.: With personal data in hand, thieves file early and often. The New York Times. http://www.nytimes.com/2012/05/27/us/id-thieves-loot-tax-checks-filing-early-and-often.html (2012)

2. Anderson, R., Böhme, R., Clayton, R., Moore, T.: Security economics and the internal market. http://www.enisa.europa.eu/act/sr/reports/econ-sec/economics-sec (2008)
3. Brynjolfsson, E., Saunders, A.: Wired for Innovation: How Information Technology Is Reshaping the Economy. MIT, Cambridge (2009)
4. Bushnell, T.: How Google developers use Ubuntu. http://www.ubuntuvibes.com/2012/05/how-google-developers-use-ubuntu.html (2012)
5. Caballero, J., Grier, C., Kreibich, C., Paxson, V.: Measuring pay-per-install: the commoditisation of malware distribution. In: Proceedings of the 20th USENIX Conference on Security, SEC'11, Berkeley. USENIX Association (2011)
6. Canalys Inc.: Enterprise security market to exceed $22 billion in 2012. http://www.canalys.com/static/press_release/2011/canalys-press-release-201211-enterprise-security-market-exceed-22-billion-2012.pdf (2011)
7. Communications Fraud Control Association: 2011 Global fraud loss survey. http://www.cfca.org/fraudlosssurvey/ (2011)
8. Consumers Union: State of the 'net' survey '07. Consum. Rep. **9**, 28–34 (2007)
9. Detica and Office of Cyber Security and Information Assurance: The cost of cyber crime. http://www.cabinetoffice.gov.uk/resource-library/cost-of-cyber-crime (2011)
10. Eliot, C.: Kim Dotcom – pirate or enabler? http://www.nzherald.co.nz/auckland-region/news/article.cfm?l_id=117&objectid=10784190 (2012)
11. European Commission: Towards a general policy on the fight against cyber crime. COM(2007) 267 final. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0267:FIN:EN:PDF (2007)
12. Fallows, J.: Hacked! The Atlantic. http://www.theatlantic.com/magazine/archive/2011/11/hacked/8673/ (2011)
13. Federal Bureau of Investigation: International cooperation disrupts multi-country cyber theft ring. Press release. http://www.fbi.gov/news/pressrel/press-releases/international-cooperation-disrupts-multi-country-cyber-theft-ring (2010)
14. Florêncio, D., Herley, C.: Evaluating a trial deployment of password re-use for phishing prevention. In: Cranor, L.F. (ed.) Proceedings of the Anti-phishing Working Groups 2nd Annual eCrime Researchers Summit, Pittsburgh, vol. 269, pp. 26–36, 4–5 Oct 2007. ACM (2007)
15. Florêncio, D., Herley, C.: Sex, lies and cyber-crime surveys. In: 10th Workshop on the Economics of Information Security, Fairfax (2011)
16. Foley, L., Barney, K., Foley, J., Leeii, J., Fergerson, J., Sarrel, M., Nelson, C., Frank, M.: Identity theft: the aftermath 2009. http://www.idtheftcenter.org/artman2/uploads/1/Aftermath_2009_20100520.pdf (2010)
17. Forrester Data: Consumer attitudes toward spam in six countries. http://www.bsacybersafety.com/files/Forrester_Consumer_Spam.pdf (2004)
18. Frost and Sullivan: Increasing security needs of enterprises to fuel growth in the world content filtering market. Press release. http://www.frost.com/prod/servlet/press-release.pag?Src=RSS&docid=84071018 (2006)
19. Gözenoglu, M., Morawe, R.: The German Anti-Botnet Advisory Center. Presentation at 'Internet Security Days', 13–15 Sept 2011, Brühl. http://www.internet-security-days.com/templates/downloads/session-2011/110913_Goezenoglu_Morawe_ABBZ.pdf (2011)
20. Herley, C., Florêncio, D.: Nobody sells gold for the price of silver: dishonesty, uncertainty and the underground economy. In: Proceedings (online) of the Workshop on Economics of Information Security. http://research.microsoft.com/pubs/80034/nobodysellsgoldforthepriceofsilver.pdf (2009)
21. Home Office: The economic and social costs of crime against individuals and households 2003–2004. http://webarchive.nationalarchives.gov.uk/20110218135832/http://rds.homeoffice.gov.uk/rds/ecom_soc_cost.html (2005)
22. House of Lords European Union Committee: Stopping the carousel: missing trader fraud in the EU. 20th Report of Session 2006–2007 (2007)

23. Huygen, A., Rutten, P., Huveneers, S., Limonard, S., Poort, J., Leenheer, J., Janssen, K., van Eijk, N., Helberger, N.: Ups and downs – economic and cultural effects of file sharing on music, film and games. TNO report 34782 http://www.ivir.nl/publicaties/vaneijk/Ups_And_Downs_authorised_translation.pdf (2009)
24. Innes, M.: Signal crimes and signal disorders: notes on deviance as communicative action. Br. J. Sociol. **55**, 335–355 (2004)
25. Institute for the Prevention of Crime: Cost of the criminal justice system. http://www.socialsciences.uottawa.ca/ipc/eng/cost_of_the_criminal_justice_system.asp (2012)
26. Irish, H.: Machine learning to classify fraudulent websites. 3rd Year Project Report, Computer Laboratory, University of Cambridge (2012)
27. Kalapesi, C., Willersdorf, S., Zwillenberg, P.: The connected kingdom: how the Internet is transforming the U.K. economy. http://www.connectedkingdom.co.uk/the-report (2010)
28. 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 ACM Conference on Computer and Communications Security, Alexandria (2008)
29. Kanich, C., Weaver, N., McCoy, D., Halvorson, T., Kreibich, C., Levchenko, K., Paxson, V., Voelker, G.M., Savage, S.: Show me the money: characterizing spam-advertised revenue. In: Proceedings of the USENIX Security Symposium, San Francisco (2011)
30. Khan, A., Hunt, J.: UK online fraud report 2012. http://forms.cybersource.com/forms/FraudReport2012UKUKwebwww2012 (2012)
31. Krebs, B.: SpamIt, Glavmed pharmacy networks exposed. Krebs on security blog. http://krebsonsecurity.com/2011/02/spamit-glavmed-pharmacy-networks-exposed/ (2011)
32. Krebs, B.: Who's behind the world's largest spam botnet? Krebs on security blog. http://krebsonsecurity.com/2012/02/whos-behind-the-worlds-largest-spam-botnet/ (2012)
33. Kuksov, D.: Buyer search costs and endogenous product design. Mark. Sci. **23**(4), 490–499 (2004)
34. Leontiadis, N., Moore, T., Christin, N.: Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In: Proceedings of the USENIX Security, San Francisco (2011)
35. Levchenko, K., Chachra, N., Enright, B., Felegyhazi, M., Grier, C., Halvorson, T., Kanich, C., Kreibich, C., Liu, H., McCoy, D., Pitsillidis, A., Weaver, N., Paxson, V., Voelker, G.M., Savage, S.: Click trajectories: end-to-end analysis of the spam value chain. In: Proceedings of the IEEE Symposium on Security and Privacy, Oakland (2011)
36. Levi, M.: Social reactions to white-collar crimes and their relationship to economic crises. In: Deflem, M. (ed.) Economic Crisis and Crime, pp. 87–105. The JAI Press/Emerald, London/Bingley (2011)
37. Levi, M., Burrows, J.: Measuring the impact of fraud in the UK: a conceptual and empirical journey. Br. J. Criminol. **48**, 293–318 (2008)
38. Leyden, J.: Russian bookmaker hackers jailed for eight years. http://www.theregister.co.uk/2006/10/04/russian_bookmaker_hackers_jailed/ (2006)
39. Lieber, E., Syverson, C.: Online vs. Offline Competition. In: Peitz, M., Waldfogel, J. (eds.) The Oxford Handbook of the Digital Economy. Oxford University Press, New York (2012)
40. M86 Security Labs: Canadian pharmacy no longer king. http://www.m86security.com/labs/traceitem.asp?article=1316 (2010)
41. McCoy, D., Pitsillidis, A., Jordan, G., Waver, N., Kreibich, C., Krebs, B., Voelker, G.M., Savage, S., Levchenko, K.: PharmaLeaks: understanding the business of online pharmaceutical affiliate programs. In: Proceedings of the USENIX Security Symposium, Bellevue (2012)
42. Microsoft Inc.: Microsoft security intelligence report, vol. 10 (2010). http://www.microsoft.com/security/sir/
43. Microsoft Inc.: Microsoft security intelligence report, vol. 9 (2010). http://www.microsoft.com/security/sir/
44. Moore, T., Clayton, R.: Examining the impact of website take-down on phishing. In: Cranor, L.F. (ed.) Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit, Pittsburgh, vol. 269, pp. 1–13, 4–5 Oct 2007. ACM (2007)

45. Moore, T., Clayton, R., Anderson, R.: The economics of online crime. J. Econ. Perspect. **23**(3), 3–20 (2009)
46. National Commission on Terrorist Attacks Upon the United States: The 9/11 Commission Report: Final Report of the National Commission on Terrorist Attacks upon the United States. W.W. Norton, New York (2004)
47. Oberholzer-Gee, F., Strumpf, K.: File-sharing and copyright. Harvard Business School Working Paper 09–132. http://www.hbs.edu/research/pdf/09-132.pdf (2009)
48. Peel, M.: Nigeria-Related Financial Crime and its links with Britain. Chatham House Report, London (2006)
49. Samosseiko, D.: The Partnerka – What is it, and why should you care? In: Proceedings of the Virus Bulletin Conference, Geneva (2009)
50. Snow, G.: Cyber security: threats to the financial sector. Testimony before the House Financial Services Committee. http://financialservices.house.gov/UploadedFiles/091411snow.pdf (2011)
51. Stiglitz, J.E., Bilmes, L.J.: The Three Trillion Dollar War: The True Cost of the Iraq Conflict. W.W. Norton, New York (2008)
52. Stone-Gross, B., Abman, R., Kemmerer, R.A., Kruegel, C., Steigerwald, D.G., Vigna, G.: The underground economy of fake antivirus software. In: 10th Workshop on the Economics of Information Security, Fairfax (2011)
53. Symantec: MessageLabs Intelligence Report. http://www.symanteccloud.com/mlireport/MLI_2010_06_June_FINAL.pdf (2010)
54. Taylor, J.: Overseas cyber-crimewave taking £600 million a year from the taxman. The Independent (2011). http://www.independent.co.uk/news/uk/crime/overseas-cybercrimewave-taking-600m-a-year-from-thetaxman-6271552.html
55. Van Eeten, M., Bauer, J.M.: Economics of malware: Security decisions, incentives and externalities. Tech. Rep. OECD STI Working Paper 2008/1, OECD, Paris. http://www.oecd.org/dataoecd/53/17/40722462.pdf (2008)
56. Van Eeten, M., Bauer, J.M., Asghari, H., Tabatabaie, S.: The role of Internet service providers in Botnet mitigation: an empirical analysis based on spam data. Tech. Rep. s0 STI Working Paper 2010/5, OECD, Paris. http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/DOC(2010)5&docLanguage=En (2010)
57. Van Eeten, M., Asghari, H., Bauer, J.M., Tabatabaie, S.: Internet service providers and Botnet Mitigation: a fact-finding study on the Dutch market. The Hague: Ministry of Economic Affairs. http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2011/01/13/internet-service-providers-and-botnet-mitigation.html (2011)
58. Vancouver Sun: Online drugs can prove deadly: Coroner. http://www.canada.com/vancouversun/news/story.html?id=ddadbf8a-bdac-45c4-a566-36acd8ffd72b (2007)
59. Wondracek, G., Holz, T., Platzer, C., Kirda, E., Kruegel, C.: Is the Internet for porn? An insight into the online adult industry. In: Proceedings (online) of the 9th Workshop on Economics of Information Security, Cambridge http://weis2010.econinfosec.org/papers/session2/weis2010_wondracek.pdf (2010)
60. Zittrain, J.: The Future of the Internet: And How to Stop It. Allen Lane, London (2008)

Chapter 13
Analysis of Ecrime in Crowd-Sourced Labor Markets: Mechanical Turk vs. Freelancer

Vaibhav Garg, Chris Kanich, and L. Jean Camp

Abstract Research in the economics of security has contributed more than a decade of empirical findings to the understanding of the microeconomics of (in)security, privacy, and ecrime. Here we build on insights from previous macro-level research on crime, and microeconomic analyses of ecrime to develop a set of hypotheses to predict which variables are correlated with national participation levels in crowd-sourced ecrime. Some hypotheses appear to hold, e.g. Internet penetration, English literacy, size of the labor market, and government policy all are significant indicators of crowd-sourced ecrime market participation. Greater governmental transparency, less corruption, and more consistent rule of law lower the participation rate in ecrime. Other results are counter-intuitive. GDP per person is not significant, and, unusually for crime, a greater percentage of women does not correlate to decreased crime. One finding relevant to policymaking is that deterring bidders in crowd-sourced labor markets is an ineffective approach to decreasing demand and in turn market size.

13.1 Introduction

The new school of ecrime [49] is both organized and driven by profits [10, 35, 51]. This is reflected by the technical and policy proposals to fight ecrime, all grounded in deterrence or rational choice theory [43]. As Anderson demonstrated in his

V. Garg (✉)
Indiana University, Bloomington, IN, USA
e-mail: gargv@umail.iu.edu

L.J. Camp
School of Informatics and Computing, Indiana University, Bloomington, IN, USA
e-mail: ljcamp@indiana.edu

C. Kanich
University of Illinois at Chicago, Chicago, IL, USA
e-mail: ckanich@uic.edu

R. Böhme (ed.), *The Economics of Information Security and Privacy*,
DOI 10.1007/978-3-642-39498-0_13,

canonical work, technical measures will never be the silver bullet [2]. Policy efforts must complement technical measures [30]. Thus, both security researchers as well as practitioners must address different stakeholders' economic incentives (or disincentives) to invest in security [49]. To the extent that security is a market, there are legitimate stakeholders such as service providers and end-users. Economic analysis of the incentives that drive service providers [18] and end-users [27] have offered academic insights [41] that inform practical solutions, e.g., the Google Vulnerability Reward program.

International efforts have been narrowly focused on ecrime itself, e.g., the European Convention of Cybercrime. A macro-level approach offers a set of complementary tactics for decreasing the threat; that is, potential measures to reduce the motivations for attackers [28]. A broader response focuses on creating environments where ecrime would be unlikely to flourish. This would go beyond the immediate term effort to deter individual attackers; instead it seeks to build a long term structure so that attackers become legitimate market participants.

To the extent that all market participants are driven by profits, why do some choose to become legitimate stakeholders, while other resort to criminal activity? Often we observe that criminal activity online clusters in specific countries, e.g., Romania, Nigeria. This phenomenon indicates that such determination is informed at least in part by macroeconomic factors. In previous research, we developed a macroeconomic model of organized ecrime [19] based on the economics of smuggling [4]. We concluded that organized ecrime can be welfare increasing in local jurisdictions. We found that ecrime markets may exist in one of the two possible equilibria: (1) high enforcement low crime or (2) low enforcement, high crime. Despite the near term increase in social welfare, a low enforcement, high crime equilibrium is not ideal, as thriving ecrime potentially acts as a prohibitive tariff against a legitimate market.

In this chapter, we empirically examine the theoretical findings of our previous work. We use the geographic locations of ecrime crowd-sourced labor to evaluate the existence of nations in the dual equilibria. Specifically, we identify the macro-level variables that encourage participation in legitimate crowd-sourcing markets, and distinguish them from variables that appear to facilitate illegitimate ecrime activities. Section 13.2 is the background and related work. The methodology is described in Sect. 13.3. Section 13.4 details the results. Section 13.5 presents the discussion. We conclude in Sect. 13.6.

13.2 Background and Related Work

Anderson notes the economic nature of security markets [2]. Both attackers [10, 35, 51] and defenders [54] are economically incentivized. Franklin et al. [17] note the shift from 'hacking for fun' to 'hacking for profit'. Furthermore, this market allows participants to offer specialized services, increasing efficiency [36]. The goods being traded range from 0 day vulnerabilities [34] to human CAPTCHA solvers [37].

Increasing activity in underground markets of information goods to bypass security measures has led to financial loss for individuals and institutions alike. The annual loss due to phishing, and possible gain to phishers, has been estimated to be as big as $178.1 million annually [35].

Motoyama et al. [38] examine one such market in Freelancer. Freelancer is a crowd-sourced labor market. Participants in the market are either bidders or buyers. Buyers create demand by posting jobs that are difficult to automate, but relatively easy to do for humans, e.g., transcription, translation. Bidders bid on the jobs of their choice and are compensated based on their performance or other criteria stated by the buyer. The authors of this study estimate that only 65.4 % of the jobs posted on Freelancer are for legitimate tasks. The remainder ask the bidders to do tasks that thwart security mechanisms, e.g. solve CAPTCHAs or send spam.

Web service abuse is not limited to crowd-sourced labor markets, nor is the abuse limited to online crime. Thomas and Martin [51] found that Internet Relay Chat (IRC) channels are being used to trade credit card data and other financial information. They also found evidence of physical crime. Holz et al. analyze the underground market with the instance of keyloggers and dropzones [21]. Threats with externalities, such as malware, may have a greater impact even when a subset of end-users are conscientious [29]. With fake antivirus, attackers have found a way of duping conscientious end-users who might not be technically adept. Stone-Gross et al. [50] estimated that the combination of merely three fake antivirus businesses generated a revenue of approximately $130 million.

Simultaneously, economics has also informed defender strategies to alleviate cybercrime. On the technical side, the goal has been to make attacks more expensive and decrease the rational imperative to attack through diminishing returns [30]. Such deterrence-based approaches are potentially successful [20], but the impact may be limited to a small time frame [43].

Thus, many previous investigations have either been microeconomic or game theoretic, the former investigating attackers' motivations and the latter suggesting defender strategies. Researchers have targeted specific markets such as IRC channels [51], Freelancer [38], malicious Chinese websites [59]. Complementary work using macroeconomics, however, has been limited.

However, insights grounded in macroeconomics are much needed. For example, the policy solutions to massive copyright violations based in deterrence theory have resulted in misguided regulatory proposals such as SOPA/PIPA. However, macroeconomic modeling of software copyright violation notes that violations in many instances would lead to an increase in revenue due to externalities [40]. Furthermore, Osorio found that massive copyright violations are driven by lack of access and economic resources [40]. Arguably, then Netflix has been more effective than SOPA would be [47]. Simultaneously, there is evidence that price cuts [9] have been more effective than DMCA [16].

Macro-level analysis has been used to study a diverse set of problems from smuggling to Olympic gold medals. Bernard and Busse [3] developed a regression-based empirical model that can predict the number of Olympic medals won by every country. Bhagwati and Hansen [4] on the other hand made a theoretical model of

smuggling. Counterintuitively, they found that smuggling is social welfare increasing. While smuggling denies trade gain, it engenders production and consumption gain. As long as the sum of production and consumption is greater than trade loss, smuggling would be social welfare increasing.

Macro-level investigations have been used to study other organized criminal activities offline. There are seven theories in criminology that have been empirically validated for crime in the physical world [44]. Social disorganization theory suggests that crime is a manifestation of neighborhood dynamics rather than that of individual motivations. Influential factors include urbanism, poverty, residential transience, heterogeneity as well as family disruption [46]. While investigations into this theory are relatively new, the findings have been encouraging [44]. Measuring informal social control can, however, be difficult, especially over the Internet.

Anomie/Strain theory underlines the disconnect between culturally driven individual aspirations and the social structures that facilitate achievement. When opportunities are rare there is incentive for individuals to deviate from cultural norms to achieve culturally desirable goals. For example, Messner and Rosenfeld [33] modeled crime as a function of the American dream, which is driven by an emphasis on economic success and provided for by an institutional structure built on the economy. While macro-level assessment of this theory is rare, they measure the strength of non-economic institutions. Some influential factors include family structure, religious participation, political involvement, education, and access to welfare, income-replacement value of welfare, as well as its comprehensiveness [7, 48]. While there is significant support for this theory, directly measuring the strength of relevant macro indicators is difficult [44].

Resource/economic deprivation theory analyses both the impact of poverty as well as economic inequality, e.g., income disparity [5, 57]. Thus, deprivation can be relative or absolute. Both perspectives have been extensively tested and demonstrate a strong and reliable ability to impact crime [44, 57]. This theory has support in cybercrime, for example previous research has identified software piracy as a function of Gross Domestic Product (GDP) per capita[1] [40].

Routine activity theory assumes motivated offenders and examines the macro-level indicators that engender opportunities for the offenders to exploit. Convergence of offender, target, and absent guardianship drives deviant behavior. The key measures here are household activity ratio and aggregate unemployment. Empirical validation of this theory, with most studies concentrating on the lack of guardianship or the lack of informal social control [12]. Weakly protected online targets are always available and proximate to motivated attackers. However, this theory has received little support in the domain of cybercrime [58].

Deterrence/rational choice theory analyses the impact of deterrence initiatives on crime, e.g., incarceration, criminal justice system, regulation, prosecution,

[1]Gross Domestic Product (GDP) indicates the aggregate worth of goods and services produced by a country in a specific time frame, typically annually. GDP per capita is an indicator of the average standard of living in a country.

etc. While the impact of this theory has weak empirical validation for crime in general [44], it has support for cybercrime [43].

Social support/altruism theory looks at the inverse relationship between state-sponsored support [13] or community altruism [8] with crime rates. This theory as well has limited empirical validation. It is also difficult to demonstrate the difference between state sponsored support (welfare) vs. private altruism (charity).

Subcultural theory examines if certain cultures are predisposed towards deviant behavior. For example, Colin Powell famously called Nigeria 'a nation of scammers'. The underlying determinants of predisposition towards crime may be other factors, such as large urban population. This is possibly the weakest theory in terms of empirical support. To the extent that culture is reflected by legal frameworks, there is some support for this theory in massive copyright infringement [40].

A similar cohesive theory of ecrime is missing. Our ability to fight ecrime is limited by our understanding of the actors involved. While the microeconomic investigations have provided an insight into the structure of ecrime markets and how they function [38,50], they have been limited in their ability to explain the evolution of organized ecrime. Why are certain markets more conducive to deviant behavior than others? Which markets would be more conducive to what kind of ecrime? In this chapter, we present the first such investigation for crowd-sourced markets.

13.3 Methodology and Data Collection

In previous research, we developed a theoretical macroeconomic model of ecrime [19]. Our model was based on the macroeconomic analysis of smuggling. We assumed that illegal goods are smuggled analogues of legal goods and are therefore perfectly substitutable. We found that organized ecrime can be profit increasing in local jurisdictions. Further, we argue that the success of illegal goods can act as a prohibitive tariff for the development of local legal markets. As such, the market exists in one of the two equilibria: (1) high enforcement, low crime or (2) low enforcement, high crime. Our findings are generalizable to markets where legal and smuggled goods coexist [42].

In this chapter, we conduct an empirical analysis of the macro-level factors that drive the market towards either equilibrium. We consider the specific instance of crowd-sourced labor markets. High enforcement and low crime is represented by Amazon's Mechanical Turk. Low enforcement and high crime is represented by Freelancer. Both Mechanical Turk and Freelancer provide functionally equivalent services, i.e., the ability to crowd-source tasks that are difficult or expensive to automate through computers, but require relatively less effort for human agents.

Amazon's Mechanical Turk service is used as an example of a high enforcement, low crime crowd-sourced market. Mechanical Turk is usually used for legitimate purposes, such as for survey-based research by academics [31]. Simultaneously, Mechanical Turk can potentially be used for illegitimate activities, e.g., CAPTCHA solving [6] or malware installations [11, 24]. The demographic distribution of

Mechanical Turk workers has been studied by Ross et al. [45] and Ipeirotis [23]. Both studies provide a demographic analysis of Turkers who participate by bidding for Human Intelligence Tasks (HITs). The analysis in this chapter uses a country-based distribution of Mechanical Turk workers from the publicly available database by Ipeirotis.[2]

Freelancer is an example of a market that has low enforcement and high crime. Motoyama et al. provide a distribution of two kinds of Freelancer participants: bidders and buyers [38]. Bidders are more akin to Mechanical Turk workers, in that they bid on tasks provided by other participants. Buyers, however, are the participants that are responsible for the market to exist as they create the demand by posting jobs that need to be completed. Around 65.4 % of Freelancer jobs are legitimate. However, Freelancer provides a market similar to Mechanical Turk and can be used for solving CAPTCHAs [37]. Additionally, it is used for other undesirable activities, such as account creation, social networking link generation and search engine optimization support. Motoyama et al. [38] identified 22 distinct job types of Freelancer classified in six categories: (1) legitimate, (2) accounts, (3) search engine optimization (SEO), (4) spam, (5) online social network, and (6) miscellaneous. The analysis in this chapter considers four of the identified categories:

1. Accounts: CAPTCHA solvers, basic accounts, and verified accounts.
2. SEO: white hat links, grey hat links, and miscellaneous.
3. Spam: bulk email, and bulk advertisement.
4. Online social network: create social network links.

Note that we assume that Mechanical Turk and Freelancer are perfect substitutes, due to assumptions of the underlying theoretical model [19]. However, this assumption is made for simplicity and the results would be applicable even when the same individual participates in both markets. A second assumption is that the Mechanical Turk market is primarily honest, while Freelancer is illegal. This too is an artifact of the underlying theoretical model [4, 19]. However, Pitt has shown that the results hold even when such a distinction is not clear, i.e., when honest and illegal markets coexist [42].

We consider participation in either crowd-sourced markets, Mechanical Turk or Freelancer, as a function of several macro-level factors. Osorio showed that macroeconomic indicators such as GDP per capita are predictors of massive copyright infringement [40]. Osorio considered a three-dimensional model: (1) accessibility, (2) affordability, and (3) legal framework. Accessibility was operationalized as the ability of the software to fit local needs, presence of after-sales support and corporate presence. Affordability was operationalized as GDP per capita. Legal framework was operationalized using the work of Easterly and Sewadeh [14]. Osorio's paper empirically examined the theoretical assertions of prior research [1, 25, 53].

[2] http://hdl.handle.net/2451/29585, retrieved on 24 February 2012

We begin by considering the theories that have found support in Osorio's analysis of copyright infringement. A key determinant in Osorio's model was Gross Domestic Product (GDP) per capita. The wages afforded to either Mechanical Turk or Freelancer workers are much lower than the minimum wage requirements in USA [22]. This indicates that there is an economic imperative to participate. Mason et al. found that financial incentives do increase the quantity of participation for Mechanical Turk [32]. While the wages are low, the corresponding value in local markets might be high based on purchasing power parity (PPP). Thus, another variable to consider would be GDP per capita by Purchasing Power Parity (PPP).[3] Both GDP per capita and GDP per capita by PPP are available from World Bank Development Indicators (WDI).[4]

Accessibility and affordability were the other two measures in Osorio's model. These are driven by the extant conditions of local ICT markets. To the extent that ICT investment, both public and private, is available, it would make participation in either Mechanical Turk or Freelancer easier. We operationalize this by using Digital Economy Rankings, produced by the Economist Intelligence Unit in collaboration with the IBM Institute for Business Value [52]. These rankings capture more than eReadiness of the country, evaluating quality as well as quantity. For example, they measure Internet penetration as well as speed, connectivity, affordability, etc.

The Digital Economy Rankings are a linear combination of six factors; *connectivity and technology infrastructure* indicates access to affordable connectivity, for both broadband and mobile, measuring assurance quality, reliability, and security; *business environment* indicates the degree to which development in private sector is facilitated by the economy, political stability, taxation, competition policy, the labour market, and openness to trade and investment; *social and cultural environment* measures both formal education as well as Internet literacy and associated technical skills; *legal environment* quantifies the progressive nature of the local legislative framework, pertaining to Internet commerce, to combat ecrime, spam, etc., as well as abuses and non-competitive behavior; *government policy and vision* indicates technology adoption by the government to facilitate citizen participation as well as access to information; *consumer and business adoption* of existing digital channels by businesses and individuals.

We also considered the export of Information and Communication Technology (ICT) services as well as the percentage export of ICT services from WDI. These measure the net worth of the ICT goods exported, excluding software. Percentage is computed as a ratio with the net worth of all goods exported. This measures the success of the business environment evaluated in the digital economy rankings. Legal environment is complemented with *rule of law* from Worldwide Governance Indicators (WGI) [26]. Rule of law indicates the degree to which

[3]Ideally, identical goods cost the same in two different markets, when priced in the same currency. However, transaction costs lead to different prices. Purchasing Power Parity measures the difference between prices in two different markets for identical goods and services.

[4]http://data.worldbank.org/indicator, retrieved on 24 February 2012.

a legal framework is implemented. A legal framework can also be thwarted by corruption or perceptions thereof. The former is measured by the corruption index from Transparency International (TI) as well as *control of corruption* from WGI. TI's corruption index as well as WGI's *control of corruption* measure perceptions of corruption, where corruption is defined as misuse of public power for private gain.

Other measures from WGI are also found to complement digital economy rankings. Both *government effectiveness* and *regulatory quality* are considered along with government policy and vision. Government effectiveness measures the perceived quality of public services, quality of civil service and the degree to which it is independent from political manipulation, the quality of policy formulation and implementation, and the perceived credibility of the government to commit to said policies. Regulatory quality quantifies the perceived ability of the government towards sound policy/regulations formulation and implementation that encourage private sector development. *Voice and accountability* is complementary to social and cultural environment. It measures intellectual and political freedom.

We also consider additional WDI indicators related to the availability of labor, measured by: (1) population, (2) population percentage of women, (3) percentage of urban population, and (4) number of Internet users. Population and number of Internet users are indicators of the available labor pool. Simultaneously, participation in crowd-sourced labor markets is made possible by Internet adoption. However, gender-based differences in adoption preferences may shift the equilibrium towards Mechanical Turk or vice versa [55]. Women are also less likely to commit crime offline [15, 39]. Thus, a higher ratio of women may shift the equilibrium toward Mechanical Turk. An urban population is more likely to have better access to technological infrastructure. Thus, a higher proportion of urban population would lead to higher Internet adoption, and therefore higher participation in crowd-sourced markets. Its impact on market equilibria would be insightful. Both the percentage of women as well as the percentage of the urban population are available from WDI.

We consider language-proficiency skills, specifically English-language proficiency, as another macro indicator. English-language proficiency is different from formal education as measured by social and cultural factors. Legitimate crowd-sourced tasks such as survey participation or proofreading require a degree of fluency with the specific language, mostly English as most tasks on Amazon are in English. However, illegal tasks such as CAPTCHA solving at best requires a mechanical pattern recognition that comes easier to human agents than automated ones [56]. For Mechanical Turk, a minimum level of English proficiency would be required to be able to understand the job solicitations, which are typically in English as requests can only be submitted from the United States. English-language proficiency is operationalized using the TOEFL's ranking of countries on reading, speaking, listening, and writing.[5]

[5] https://www.ets.org/toefl/research/topics/candidates_and_populations, retrieved on 24 February 2012.

Table 13.1 List of macro-level variables

Variable name	Provider	Year
Affordability (AFF)		
GDP per capita	WDI	2010
GDP per capita by PPP	WDI	2010
Consumer and business adoption	Economist	2010
Accessibility (ACC)		
Digital economy rankings	Economist	2010
Connectivity and technology	Economist	2010
Business environment	Economist	2010
Export of ICT services	WDI	2010
% export of ICT services	WDI	2010
Social and cultural environment	Economist	2010
Voice and accountability	WGI	2010
Legal (LEG)		
Legal environment	Economist	2010
Rule of law	WGI	2010
TI corruption index	Transparency International	2011
Control of corruption	WGI	2010
Government policy & vision	Economist	2010
Government effectiveness	WGI	2010
Regulatory quality	WGI	2010
Population (POP)		
Population	WDI	2010
Population density	WDI	2010
Population % women	WDI	2010
% urban population	WDI	2010
Number of Internet users	WDI	2009
English (ENG)		
English reading	TOEFL	2010
English listening	TOEFL	2010
English speaking	TOEFL	2010
English writing	TOEFL	2010
Security (SEC)		
Secure Internet servers (SIS)	WDI	2010
SIS by population	WDI	2010

Finally, Security of ICT infrastructure is also an indicator: (1) number of secure Internet servers (SIS), and (2) number of SIS by population. SIS and SIS by population would encourage market investment by providing assurance of security. These indicators were also procured from WDI.

A list of all variables considered and respective sources is given in Table 13.1. The final regression equation is given by Eq. 13.1, where N corresponds to number of workers; AFF, ACC, LEG, POP, ENG, and SEC refer to measures of affordability,

accessibility, legal framework, availability of labor, English-language proficiency, and security, respectively:

$$N = \beta_0 + \beta_1 \cdot \text{AFF} + \beta_2 \cdot \text{ACC} + \beta_3 \cdot \text{LEG} + \beta_4 \cdot \text{POP} + \beta_5 \cdot \text{ENG} + \beta_6 \cdot \text{SEC}. \tag{13.1}$$

13.4 Results

In this chapter, we empirically examine a macro-economic model of organized ecrime by considering the specific example of crowd-sourced labor markets. The model posited that information communication technology markets would tend to exist in one of two equilibria: (1) high enforcement, low crime, and (2) low enforcement, high crime [19]. Amazon's Mechanical Turk is an instance of high enforcement, low crime, while Freelancer represents low enforcement, high crime.

Participation in either of these markets is a function of macro-level indicators, including macroeconomic indicators. The research question that we address is here is two-fold. Which macro-level indicators encourage participation in crowd-sourced labor markets? Secondly, which specific indicators inform the market equilibrium, by either encouraging or alleviating criminal activity online?

We present the results of the linear regression model for the five independent variables as give by Eq. 13.1. The independent variables were normalized by subtracting the mean and dividing by the standard deviation. This allowed for more legible results while having no impact on the variance explained by the regression model. The regression for Freelancer bidders and buyers were treated differently. Freelancer bidders are more akin to Mechanical Turk workers, as they bid on tasks posted by other market participants. Freelancer buyers, however, are more similar to those who post HITs on Mechanical Turk. While bidders represent supply, buyers create demand. Table 13.3 presents the results of the regression for all three dependent variables: (1) Mechanical Turk workers, (2) Freelancer bidders, and (3) Freelancer buyers. The cells in the table represent the p-value for which the specific indicator was significant in the best-fit model.[6] NS indicates that the indicator was in the best-fit model but was not statistically significant.

The first regression model had Mechanical Turk workers as the dependent variable. The model's adjusted $R^2 = 0.95$. The best fit for the model was given by digital economy ranking, connectivity and technology infrastructure, social and cultural environment, legal environment, government policy and vision, consumer and business adoption, English proficiency scores (reading, listening, speaking, and writing), number of Internet users, population, population percentage of women, voice and accountability, secure Internet servers, and secure Internet servers by

[6]Best-fit model indicates the subset of indicators for which the corresponding linear regression obtained the highest adjusted R^2 value.

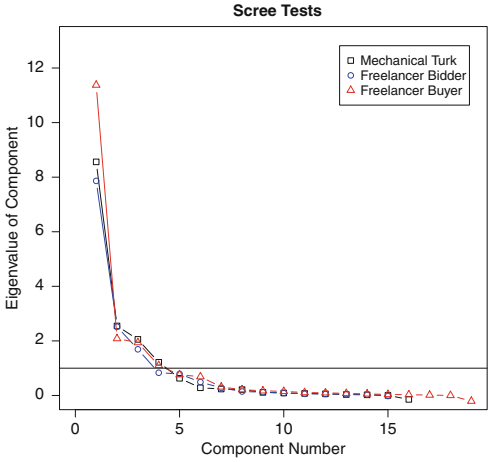

Fig. 13.1 Scree tests for EFA variables

population; adjusted $R^2 = 0.97$. The F-statistic for the model was statistically significant; $p < 2.2 \times 10^{-16}$.

The second run of the regression model had Freelancer bidders as the dependent variable. This regression gave an adjusted R^2 value of 0.93. The best fit was given by connectivity and technology infrastructure, social and cultural environment, government policy and vision, English proficiency scores (reading, writing), number of Internet users, population, export of ICT services, percentage export of ICT services, rule of law, government effectiveness, control of corruption, secure Internet servers, and secure Internet servers by population; adjusted $R^2 = 0.95$. The F-statistic for the model was statistically significant; $p < 2.2 \times 10^{-16}$.

The last run of the regression model had Freelancer buyers as the dependent variable. This model gave an adjusted R^2 value of 0.9538. The best fit was given by digital economy ranking, connectivity and technology infrastructure, business environment, social and cultural environment, government policy and vision, consumer and business adoption, TI corruption index, English proficiency scores (reading, listening), number of Internet users, population, population percentage of women, percentage of urban population, rule of law, regulatory quality, voice and accountability, control of corruption, secure Internet servers, secure Internet servers by population; adjusted $R^2 = 0.96$. The F-statistic for the model was statistically significant; $p < 2.2 \times 10^{-16}$.

Many of the indicators were highly correlated. Thus, the number of macro-level indicators can be condensed to a smaller subset using exploratory factor analysis (EFA). We conducted a scree test to identify the optimum number of factors which appears in Fig. 13.1. Only the indicators that were present in the respective best-fit models were considered. We considered the number of factors that give an eigenvalue greater than 1. Thus, there were four factors for Mechanical Turk and Freelancer Buyers, while they were three factors for Freelancer Bidders. The factor

loadings for Mechanical Turk, Freelancer Bidders, and Freelancer Buyers are given in Table 13.2.

13.5 Discussion

In this chapter, we empirically examine a macroeconomic model of ecrime in crowd-sourced labour markets. We investigate two research threads. First, we addressed the macro-level difference between two market equilibria, for crowd-sourced labor markets: (1) high enforcement, low crime, (2) and low enforcement, high crime. While Amazon's Mechanical Turk service is an example of the former, Freelancer represents the latter. Secondly, we examined the macro-level differences between stakeholders in the Freelancer market, i.e., those who bid on tasks, suppliers, and those who created the tasks or demand.

We begin by identifying the macro-level indicators that encourage overall participation in crowd-sourced labor markets. We conducted an exploratory factor analysis to identify how the different indicators for the best-fit models related to each other, Table 13.2. The factor loadings for Mechanical Turk as well as Freelancer were similar. While for Mechanical Turk and Freelancer buyers there were four factors, for Freelancer bidders there were three. The factor loadings on the fourth factors as well as the variance explained was low; 0.024 for Mechanical Turk and 0.027 for Freelancer Buyer. Thus, we assume that there are typically three factors that drive participation in either of these markets. The least amount of variance is explained by the third, and last, factor that constitutes population, number of Internet users, and SIS. This factor essentially indicates a cyber-ready labor force. Thus, availability of a labor force with the essential skill set is required but is not an adequate predictor of participation in crowd-sourced labor markets.

The second factor explains more variance and is characterized mostly by English proficiency. English proficiency is required for participation, both by the tasks themselves and the solicitations for work because most posts are in English. The degree of English proficiency corresponds to the tasks afforded. A specific level of proficiency is not required for Freelancer bidding, ability to understand spoken English is required for Freelancer buying, and proficiency in listening and speaking is needed for Mechanical Turk.

The first factor explained most of the difference in the variance. This factor constitutes both the quality and quantity of affordable Internet access. Higher Internet penetration allows a greater proportion of the population to get online. However, Internet literacy is not facilitated just by the ability to get online, but also by the quality of the bandwidth available, e.g., speed, security, reliability. Available bandwidth must be utilized, thus adoption both by individuals and businesses must be facilitated by public policies and private enterprise.

There were differences on specific indicators that constitute the three big factors. We examined these differences by comparing the best fit models for the three distinct groups.

Table 13.2 Exploratory factor analyses

	Mechanical turk factors					Freelancer buyer factors					Freelancer bidder factors			
Variable name	#1	#2	#3	#4	Uniq	#1	#2	#3	#4	Uniq	#1	#2	#3	Uniq
Digital economy ranking	0.973	0.220			0.005	0.961	0.197		0.184	0.005				
Connectivity and technology	0.934	0.138		0.187	0.067	0.910	0.146		0.205	0.102	0.926	0.119		0.126
Business environment						0.916	0.115			0.141				
Social and cultural envmt.	0.930	0.196		0.108	0.080	0.896	0.227		0.235	0.085	0.914	0.205		0.113
Legal environment	0.823	0.414		−0.212	0.105						0.918	0.157		0.129
Govt. policy & vision	0.938	0.193		−0.174	0.053	0.921	0.154		0.192	0.089				
Consumer and business adoption	0.959	0.221			0.028	0.942	0.194		0.219	0.025	0.954	0.173		0.060
TI corruption index						0.946				0.100				
English reading	0.294	0.860			0.173	0.312	0.907			0.078	0.317	0.923		0.046
English listening	0.347	0.900	−0.177	0.180	0.005	0.414	0.818	−0.185	−0.100	0.115				
English speaking	0.287	0.768	−0.219	0.304	0.188									
English writing	0.303	0.864			0.161						0.351	0.803		0.228
# Internet users		−0.133	0.988		0.005		−0.124	0.949		0.076			0.997	0.005
Population	−0.278		0.778		0.312	−0.250		0.817	−0.204	0.228	−0.297	0.128	0.753	0.328
Pop. % women		0.404		−0.107	0.825		0.451			0.794				
% urban pop.						0.554		−0.172	0.292	0.578				
Export ICT serv.											0.281	0.459	0.330	0.601
% export ICT serv.												0.444		0.798
Rule of law						0.961	0.146		−0.110	0.043	0.951	0.142		0.075
Govt. effectiveness											0.947	0.166		0.075
Regulatory quality						0.914	0.199			0.117				
Voice and accountability	0.730	0.463	−0.112		0.235	0.747	0.458			0.221				
Control of corruption						0.977				0.029	0.969			0.053
Secure Internet servers	0.384		0.490		0.593	0.313		0.456	0.328	0.579	0.328		0.519	0.617
SIS by population	0.760	0.201		0.322	0.278	0.752	0.187		0.150	0.374	0.761	0.156		0.393
Proportion variance	0.43	0.231	0.121	0.024		0.558	0.117	0.099	0.027		0.485	0.141	0.131	
Cumulative variance	0.43	0.661	0.782	0.805		0.558	0.675	0.774	0.801		0.485	0.625	0.757	

GDP per capita and GDP per capita by PPP were not in any of our best-fit models. This finding is different from that observed for massive copyright infringement [40]. This indicates the importance of the direct financial rewards, or the degree to which they supplement or complement participants' incomes is equally relevant for both markets.

Table 13.3 also shows that population, number of Internet users, and number of SIS are statistically significant indicators of participation in crowd-sourced markets. Participation, in either Mechanical Turk or Freelancer, was positively correlated with these three indicators. The three indicators together form an indicator of a cyber-ready labor force. All three load together as the third factor, see Table 13.2. Thus, while a good indicator of participation, they do not account much for the variance of participation in either market.

Connectivity and technology, social and cultural environment, government policy and vision, and English proficiency scores (reading) were other indicators that were common in the best-fit models of these three markets. These factors were not statistically significant for all three, however, these factors indicate a necessary, if not sufficient, indicator for increasing participation. These factors account for most of the variance in the best-fit models for both Mechanical Turk and Freelancer, bidders as well as buyers. Connectivity and technology, social and cultural environment, and government policy and vision loaded on the first, factor while English proficiency loaded on the second, as seen by the results of EFA in Table 13.2.

There were no statistically significant indicators that were common between Mechanical Turk workers and Freelancer bidders. This finding is counterintuitive, as Mechanical Turk workers as well as Freelancer bidders are both suppliers in respective markets, responding to a demand created by buyers. The expectation is that since the services provided are similar, the macro-economic backgrounds are as well. However, for Mechanical Turk workers the overall state of the digital economy was important, mostly driven by Internet penetration, via broadband as well as mobile. Affordable access to high bandwidth that is reliable and secure is important. Freelancer bidders may not have lower quality or quantity of access. Participation as a Freelancer bidder is driven by social and cultural environment, lower technical expertise and English proficiency than their Mechanical Turk counterparts. Social and cultural environment extends to the legal framework and the ability of the government to implement it. This result is different from perceptions of corruption, this is rather an indicator of general perceptions of rules being followed in society and penalties being imposed for deviations.

Statistically significant common indicators between Freelancer bidders and buyers are also limited. The one common indicator was social and cultural environment. Both bidding and buying, then, require a minimum level of technical expertise and general education. However, the switch from bidder to buyer demands several other resources. Participation as buyers requires access to affordable bandwidth. Solicitation of jobs would probably be a part of the larger business strategy. Thus, business environment is also important and so is consumer and business adoption. English-language proficiency is again more important for buyers than for bidders.

Table 13.3 Regression models

	Mechanical turk				Freelancer buyer				Freelancer bidder			
Variable name	Estimate	Std. Err.	Pr(> |t|)		Estimate	Std. Err.	Pr(> |t|)		Estimate	Std. Err.	Pr(> |t|)	
Intercept	35.71	59.77	0.555		3,009	5,832	0.609		42,300	51,980	0.420	
Digital economy ranking	55.78	24.91	0.034	*	9,168	2,941	0.003	**				
Connectivity and technology	−18.26	6.405	0.008	**	−2, 651	667.5	≈0	***	5,361	5,100	0.299	
Business environment					−2, 179	774.3	0.007	**				
Social and cultural envmt.	11.85	7.431	0.123		−1, 417	641.9	0.033	*	−23, 330	8,337	0.007	**
Legal environment	9.148	5.706	0.121									
Govt. policy & vision	−13.58	7.302	0.074		−1, 534	676.0	0.029	*	7,139	6,209	0.257	
Consumer and business adoption	−24.99	9.189	0.011	*	−2, 490	931.1	0.011	*	14,940	9,029	0.105	
TI corruption index					71.81	40.79	0.086					
English reading	−7.021	3.631	0.064		−390.5	248.2	0.124		5,980	3,437	0.089	
English listening	16.62	4.542	0.001	**	727.8	232.9	0.003	**				
English speaking	−7.782	3.527	0.036	*								
English writing	−6.019	4.082	0.152						−5,294	3,905	0.182	
# Internet users	−57.68	4.109	≈0	***	−2,403.25	373.26	≈0	***	−97,883.6	6,788.9	≈0	***
Population	61.11	3.773	≈0	***	2,765.01	354.94	≈0	***	115,491.5	6,308.8	≈0	***

(continued)

Table 13.3 (continued)

	Mechanical turk				Freelancer buyer				Freelancer bidder			
Variable name	Estimate	Std. Err.	Pr(> |t|)		Estimate	Std. Err.	Pr(> |t|)		Estimate	Std. Err.	Pr(> |t|)	
Pop. % women	−10.31	8.461	0.234		−1,782.62	790.26	0.023	*				
% urban population					917.47	366.06	0.017	*				
Export ICT serv.									5,298.2	4,992.8	0.295	
% export ICT serv.									1,305	668.3	0.057	
Rule of law					45.21	40.00	0.266		1,463	613.2	0.022	*
Govt. effectiveness									−1,413	571.4	0.017	*
Regulatory quality					75.77	39.84	0.066					
Voice and accountability	−0.4775	0.246	0.064		−23.51	18.49	0.211					
Control of corruption					−112.3	53.04	0.041	*	−831.8	594.5	0.167	
SIS	68.90	2.398	≈0	***	6,427.72	230.69	≈0	***	58,685.9	3,550.5	≈0	***
SIS by population	−30.56	10.530	0.008	**	−1,632.79	1,034.02	0.123		−19,872.4	12,758.3	0.127	
Residual std. err.	14.84 on 26 DF				1,434 on 38 degrees freedom				22,760 on 43 degrees freedom			
Multiple R^2	0.98				0.98				0.96			
Adjusted R^2	0.97				0.96				0.95			
F-statistic	87.72 on 16 and 26 DF				78.02 on 19 and 38 DF				71.62 on 15 and 43 DF			

Significance codes: 0 $\leq$ *** $\leq$ 0.001 $\leq$ ** $\leq$ 0.01 $\leq$ * $\leq$ 0.05

Percentage of urban population is also an indicator of buying. Urban areas are likely to have higher quality and quantity of Internet penetration, so this is not surprising.

However, Mechanical Turk participation, which also requires access to affordable bandwidth, is not dependent on urban population. This difference becomes clearer when we consider that buying is also a function of population percentage of women and corruption. These indicators together suggest that buying in Freelancer is explained by social disorganization theory in criminology [46]. Social disorganization theory considers crime to be a function of the breakdown of community structures, as measured by corruption, heterogeneity of population, and urbanization.

This is different from participation as a bidder, which is dependent on rule of law and the effectiveness of the government to enforce the rules. Thus, for bidders, deterrence-based efforts, such as increasing penalties on increased enforcement, might be successful at decreasing participation [20]. However, the long-term impact of such measures would be limited [43]. Without the demand for Freelancer services being alleviated, there is likely to be a displacement effect, where bidders from countries with a lax legal framework would increasingly participate [43].

13.6 Conclusion and Future Work

This work examines the macro-level factors that encourage participation in crowd-sourced labor markets online. We differentiate between crowd-sourced labor markets with a lower number of illegitimate tasks and those with a high number of illegitimate tasks. In previous work we posited that ICT markets would tend to exist in one of the two equilibria: high enforcement and low crime or low enforcement and high crime. We considered Amazon's Mechanical Turk as an example representing high enforcement, low crime and Freelancer representing low enforcement, high crime. We identified the macro-level factors that appear to facilitate either of these two market equilibria.

While this methodology provides insight into the factors underlying participation in cybercrime, it is by no means comprehensive. A complementary approach could identify high enforcement, low crime and low enforcement, high crime countries, and then examine the proportion of participation in Mechanical Turk by those countries as opposed to in Freelancer. We wish to perform this analysis in future research. Adoption of Mechanical Turk/Freelancer may also be driven by the dynamics of social networks. A macro-level investigation examines variables that facilitate such dynamics. This, however, does not directly address why certain networks may become more popular in specific countries. To the degree that adoption is driven by trust and just plain awareness, physical social networks may in fact influence the decision to participate in a specific crowd-sourced labor market. However, given that crowd-sourced labor markets are not like typical social networks, such as Facebook or telephone networks, we assume that the impact of network effects is trivial. We also do not address the intentionality of

participants from either a rational microeconomic or a boundedly rational behavioral perspective. Individual motivations can be pursued through a survey-based study, and should be addressed by future research.

Participation in crowd-sourced labor markets requires three factors: (1) affordable access to reliable and secure Internet, (2) English-language proficiency, and (3) availability of a cyber-ready labor force. A high enforcement, low crime equilibrium has several additional characteristics. The digital economy of the participants' local jurisdictions must be thriving so as to provide access to affordable, reliable, and secure bandwidth. Participants must have high technical as well as language skills. Adoption of ICTs by individuals as well as businesses in the local jurisdiction must be sufficient and facilitated by government policies.

There were several statistically significant indicators that were common between Mechanical Turk and Freelancer buyers: digital economy ranking, connectivity and technology, consumer and business adoption, and English proficiency scores (listening). So why do more Mechanical Turk workers not gravitate towards Freelancer? A possible explanation is that Mechanical Turk workers have higher levels of English proficiency as well as formal education and Internet literacy. Simultaneously, Freelancer buyers come from jurisdictions with higher urban population, more corruption, as well as those conducive to a business enterprise that is not necessarily ethical.

A low enforcement, high crime equilibrium is facilitated by a poor legal framework. Freelancer bidders' local jurisdictions often do not have an effective or comprehensive legal framework. Even when the legal framework is adequate, prosecution is denied due to corruption in the local jurisdictions of Freelancer buyers. Participation through bidding in these markets can be alleviated, in the short term, through deterrence-based strategies, such as increasing penalties or higher enforcement. Yet the very factors of corruption and lack of consistent rule of law limits the efficacy of these short-term efforts. The long-term success of the deterrence would be limited. Thus, short-term policies must be complemented by long-term strategies. Two strategies are addressing participation through buying (i.e., by alleviating the demand for criminal goods), and changing the underlying macro-level factors. The second requires shifting the underlying macro-level factors of underdeveloped ICT markets, inadequate language skills (in English) as well as Internet literacy.

Acknowledgements We would like to thank Prof. Panagiotis G. Ipeirotis who made the demographic data publicly available. We also thank Prof. Stefan Savage's research group at UCSD for providing us with the data on Freelancer. Finally, we thank the Stat/Math Center at Indiana University for their insight on the statistical analysis. Any mistakes in this chapter are the authors' own responsibility.

This presentation of this research was made possible by funding from the Volkswagen Foundation. This material is based upon work supported by the National Science Foundation (NSF) under award number 0916993. Any opinions, findings, and conclusions or recommendations expressed in this presentation are those of the author(s) and do not necessarily reflect the views of the NSF.

References

1. Al-Jabri, I., Abdul-Gader, A.: Software copyright infringements: an exploratory study of the effects of individual and peer beliefs. Omega **25**(3), 335–344 (1997)
2. Anderson, R.: Why information security is hard – an economic perspective. In: Proceedings of the 17th Annual Computer Security Applications Conference, New Orleans (2001)
3. Bernard, A., Busse, M.: Who wins the Olympic Games: economic resources and medal totals. Rev. Econ. Stat. **86**(1), 413–417 (2004)
4. Bhagwati, J., Hansen, B.: A theoretical analysis of smuggling. Q. J. Econ. **87**, 172–187 (1973)
5. Bursik, R., Jr., Grasmick, H.: Economic deprivation and neighborhood crime rates, 1960–1980. Law Soc. Rev. **27**(2), 263–283 (1993)
6. Bursztein, E., Bethard, S., Fabry, C., Mitchell, J.C., Jurafsky, D.: How good are humans at solving CAPTCHAs? A large scale evaluation. In: Proceedings of the 2010 IEEE Symposium on Security and Privacy, Berkeley/Oakland (2010)
7. Chamlin, M., Cochran, J.: Assessing Messner and Rosenfeld's institutional anomie theory: a partial test. Criminology **33**(3), 411–429 (1995)
8. Chamlin, M., Cochran, J.: Social altruism and crime. Criminology **35**(2), 203–226 (1997)
9. Chen, Y., Png, I.: Software pricing and copyright enforcement: private profit vis-a-vis social welfare. In: Proceedings of the 20th International Conference on Information Systems, Charlotte (1999)
10. Choo, K.K., Smith, R.: Criminal exploitation of online systems by organised crime groups. Asian J. Criminol. **3**(1), 37–59 (2008)
11. Christin, N., Egelman, S., Vidas, T., Grossklags, J.: It's all about the Benjamins: an empirical study on incentivizing users to ignore security advice. In: Proceedings of the 16th International Conference on Financial Cryptography and Data Security, Kralendijk. Financial Cryptography and Data Security Lecture Notes in Computer Science, vol. 7035, pp. 16–30 (2012)
12. Cohen, L., Felson, M.: Social change and crime rate trends: a routine activity approach. Am. Soc. Rev. **44**, 588–608 (1979)
13. Colvin, M., Cullen, F., Ven, T.: Coercion, social support, and crime: an emerging theoretical consensus. Criminology **40**(1), 19–42 (2002)
14. Easterly, W., Sewadeh, M.: Global development network growth database. Technical report, World Bank Group (2001)
15. Edlund, L., Li, H., Yi, J., Zhang, J.: Sex ratios and crime: evidence from China's one-child policy. Technical report 3214, Forschungsinstitut zur Zukunft der Arbeit (IZA) (2007)
16. Fortunato, M.: Let's not go crazy: why Lenz vs. Universal Music Corp. undermines the notice and takedown process of the Digital Millennium Copyright Act. J. Intellect. Prop. Law **17**, 147–445 (2009)
17. 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 14th ACM Conference on Computer and Communications Security, Alexandria (2007)
18. Gal-Or, E., Ghose, A.: The economic consequences of sharing security information. In: Proceedings of the 2nd Annual Workshop on the Economics of Information Security, University of California, Berkeley (2003)
19. Garg, V., Husted, N., Camp, J.: Smuggling theory approach to organized digital crime. In: Proceedings of the 6th Annual APWG eCrime Researcher's Summit, San Diego (2011)
20. Higgins, G., Wilson, A., Fell, B.: An application of deterrence theory to software piracy. J. Crim. Justice Pop. Cult. **12**(3), 166–184 (2005)
21. Holz, T., Engelberth, M., Freiling, F.: Learning more about the underground economy: a case-study of keyloggers and dropzones. In: Proceedings of the 14th European Symposium on Research in Computer Security, Saint-Malo (2009)
22. Horton, J., Chilton, L.: The labor economics of paid crowdsourcing. In: Proceedings of the 11th ACM Conference on Electronic Commerce, Harvard (2010)

23. Ipeirotis, P.: Demographics of Mechanical Turk. Technical report CEDER-10-01, Stern School of Business, New York University (2010)
24. Kanich, C., Checkoway, S., Mowery, K.: Putting out a hit: crowdsourcing malware installs. In: Proceedings of the 5th USENIX Workshop on Offensive Technologies, San Francisco (2011)
25. Katz, M., Shapiro, C.: Technology adoption in the presence of network externalities. J. Pol. Econ. **94**(4), 822–841 (1986)
26. Kaufmann, D., Kraay, A., Mastruzzi, M.: The worldwide governance indicators: methodology and analytical issues. Hague J. Rule Law **3**(2), 220–246 (2011)
27. Krishnamurthy, S., Tripathi, A.: Bounty programs in free/libre/open source software. In: Bitzer, J., Schröder, P.J.H. (eds.) The Economics of Open Source Software Development, pp. 165–183. Elsevier, Amsterdam/Boston (2006)
28. Kwon, J., Johnson, M.: An organizational learning perspective on proactive vs. reactive investment in information security. In: Proceedings of the 10th Annual Workshop on Economics of Information Security, Fairfax (2011)
29. Lelarge, M.: Economics of malware: epidemic risks model, network externalities and incentives. In: Proceedings of the 47th Annual Allerton Conference on Communication, Control, and Computing, Monticello (2009)
30. Li, Z., Liao, Q., Striegel, A.: Botnet economics: uncertainty matters. In: Johnson, M.E. (ed.) Managing Information Risk and the Economics of Security, pp. 245–267. Springer, New York/London (2009)
31. Mason, W., Suri, S.: Conducting behavioral research on Amazon's Mechanical Turk. Behav. Res. Methods **44**, 1–23 (2012)
32. Mason, W., Watts, D.: Financial incentives and the "performance of crowds". In: Proceedings of the ACM SIGKDD Workshop on Human Computation, Washington, DC (2010)
33. Messner, S., Rosenfeld, R.: Crime and the American Dream. Wadsworth Publishing Co., Belmont (1994)
34. Miller, C.: The legitimate vulnerability market: inside the secretive world of 0-day exploit sales. In: Proceedings of the 6th Annual Workshop on the Economics of Information Security, Pittsburgh (2007)
35. Moore, T., Clayton, R.: An empirical analysis of the current state of phishing attack and defence. In: Proceedings of the 6th Annual Workshop on the Economics of Information Security, Pittsburgh (2007)
36. Moore, T., Clayton, R., Anderson, R.: The economics of online crime. J. Econ. Perspect. **23**(3), 3–20 (2009)
37. Motoyama, M., Levchenko, K., Kanich, C., McCoy, D., Voelker, G., Savage, S.: Re: CAPTCHAs–understanding CAPTCHA-solving services in an economic context. In: Proceedings of the 19th USENIX Security Symposium, Washington, DC (2010)
38. Motoyama, M., McCoy, D., Levchenko, K., Savage, S., Voelker, G.M.: Dirty jobs: the role of freelance labor in web service abuse. In: Proceedings of the 20th USENIX Security Symposium, San Francisco (2011)
39. Nagel, I., Hagan, J.: Gender and crime: offense patterns and criminal court sanctions. Crime Justice **4**, 91–144 (1983)
40. Osorio, C.: A contribution to the understanding of illegal copying of software: empirical and analytical evidence against conventional wisdom. In: Program on Internet and Telecoms Convergence (2002). http://hdl.handle.net/1721.1/1479
41. Ozment, A.: Bug auctions: vulnerability markets reconsidered. In: Proceedings of the 3rd Workshop on the Economics of Information Security, Minnesota (2004)
42. Pitt, M.: Smuggling and price disparity. J. Int. Econ. **11**(4), 447–458 (1981)
43. Png, I., Wang, C., Wang, Q.: The deterrent and displacement effects of information security enforcement: international evidence. J. Manag. Inf. Syst. **25**(2), 125–144 (2008)
44. Pratt, T., Cullen, F.: Assessing macro-level predictors and theories of crime: a meta-analysis. Crime Justice **32**(2005), 373–450 (2005)

45. Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers? Shifting demographics in Mechanical Turk. In: Extended Abstracts on Human Factors in Computing Systems, Atlanta (2010)
46. Sampson, R., Groves, W.: Community structure and crime: testing social-disorganization theory. Am. J. Soc. **94**(4), 774–802 (1989)
47. Sanchez, J.: SOPA, internet regulation and the economics of piracy. http://www.cato.org/publications/commentary/sopa-internet-regulation-economics-piracy (2012)
48. Savolainen, J.: Inequality, welfare state, and homicide: further support for the institutional anomie theory. Criminology **38**(4), 1021–1042 (2000)
49. Shostack, A., Stewart, A.: The New School of Information Security. Addison-Wesley Professional, Upper Saddle River (2008)
50. Stone-Gross, B., Abman, R., Kemmerer, R., Kruegel, C., Steigerwald, D., Vigna, G.: The underground economy of fake antivirus software. In: Proceedings of the 10th Annual Workshop on the Economics of Information Security, Fairfax (2011)
51. Thomas, R., Martin, J.: The underground economy: priceless. USENIX; login **31**(6), 7–16 (2006)
52. Unit, E.I.: Digital economy rankings 2010 beyond e-readiness. Technical report, Economist Intelligence Unit and The IBM Institute for Business Value (2010)
53. Varian, H.: Economics of Information Technology. University of California, Berkeley (2001)
54. Varian, H.: System reliability and free riding. In: Camp, L.J., Lewis, S. (eds.) Economics of Information Security, pp. 1–15. Kluwer, Boston (2004)
55. Venkatesh, V., Morris, M.G.: Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. MIS Q. **24**(1), 115–139 (2000)
56. von Ahn, L., Blum, M., Hopper, N., Langford, J.: CAPTCHA: using hard AI problems for security. In: Proceedings of Eurocrypt, Warsaw (2003)
57. Walker, I., Pettigrew, T.: Relative deprivation theory: an overview and conceptual critique. Br. J. Soc. Psychol. **23**(4), 301–310 (1984)
58. Yar, M.: The novelty of 'cybercrime'. Eur. J. Criminol. **2**(4), 407–427 (2005)
59. Zhuge, J., Holz, T., Song, C., Guo, J., Han, X., Zou, W.: Studying malicious websites and the underground economy on the Chinese Web. In: Proceedings of the 7th Annual Workshop on the Economics of Information Security, Hanover (2008)