Estimating the costs of consumer-facing cybercrime: A tailored instrument and representative data for six EU countries

Riek, Markus; Boehme, Rainer; Ciere, Michael; Hernandez Ganan, Carlos; van Eeten, Michel

Publication date
2016

Document Version
Final published version

Published in
Proceedings of Workshop of Economics of Information Security

Estimating the costs of consumer-facing cybercrime: A tailored instrument and representative data for six EU countries

Working Paper

Markus Riek*, Rainer Böhme
University of Innsbruck, Department of Computer Science
Innsbruck, Austria
markus.riek@uibk.ac.at
Michael Ciere, Carlos Gañán, Michel van Eeten
TU Delft, Faculty of Technology, Policy and Management
Delft, The Netherlands

Abstract. While cybercrime has existed for many years and is still reported to be a growing problem, reliable estimates of the economic impacts are rare. We develop a survey instrument tailored to measure the costs of consumer-facing cybercrime systematically, by aggregating different cost factors into direct losses and expenses for protection measures. We use our instrument to collect representative primary data on the prevalence of seven different types of consumer-facing cybercrime in six European countries. Our results show that cybercrime rather causes losses of time than money and that the losses of victims are dwarfed by the expenses for preventive protection. We identify scams to be the worst type of cybercrime in terms of losses. While identity thefts associated with financial accounts cause high initial losses for the victims, most of them receive substantial compensation. We find that loss distributions are skewed to the left, bearing the risk of overestimating costs when looking at figures summarized by the arithmetic mean.

Keywords. *Costs of cybercrime, Consumer research, Empirical measurement*

*Corresponding author.

1 Introduction

Like other cyber threats, consumer-facing cybercrime is around for many years. The globally increasing use of the Internet and in particular the uptake of online services which require financial transactions, such as online banking and shopping, attracts profit-oriented criminals [e. g., ITU, 2015]. While the problem seems to receive increasing attention from the media, economic estimates of the impact on consumers are still rare. Economic cost estimates are needed to inform policy, decisions on security investments in the private sector, or the messages to be conveyed in public awareness campaigns.

Ryan and Jefferson [2003] already remark that security decision are often poor because there is no reliable data upon which to base them. Even worse, there is unreliable data that is masquerading as reliable data. In 2004, the US Congressional Research Service assigned high priority to the question whether society devotes enough resources to information security. They add that part of the answer must come from economic analysis. However, "[n]o one in the field is satisfied with our present ability to measure the costs and probabilities of cyber-attacks." Cashell et al. [2004, p.1].

Unfortunately, more than a decade later current studies still seem to fall short on providing reliable economic estimates on the costs of cybercrime for consumers. While critique regarding existing estimates [e. g., Florêncio and Herley, 2013, Hyman, 2013] and proposals for improved measurement [Anderson et al., 2013] exist, progress remains very slow. Part of the reason is that the majority of studies focuses on the costs of cybercrime for businesses, neglecting consumers. The more important aspect might be that estimating the costs of consumer-facing cybercrime is a challenging and laborious endeavor. We set out to fill this gap with the three following contributions:

- **Development of a measurement instrument.** We develop an instrument to measure the costs of cybercrime for consumers grounded on a review of existing work in the context of cybercrime surveys and general approaches for loss estimation.
- **Representative measurement of cybercrime.** We use our instrument to collect primary data on the prevalence of seven different types of cybercrime among the adult population of Internet users in six European countries.
- **Estimation of costs to consumers.** We derive economic cost estimates for different cost categories including losses of victimization and expenses for protection.

To the best of our knowledge, we are the first to provide economic cost estimates based on a detailed breakdown into different types of cybercrime and different cost categories with representative data for multiple countries. While our empirical results focus on the costs for consumers, we also derive implications for financial services, payment, and online shopping providers.

The study is structured along our contributions. Section 2 introduces cybercrime measurement by reviewing existing studies and methods. Section 3 describes the development of our measurement instrument. Section 4 presents the empirical results of the survey. Section 5 reports the cost estimation. The final Section 6 concludes the study with a discussion.

2 Measuring the costs of cybercrime

Measuring cybercrime has always been a complicated endeavour. This Section provides a brief overview of the *status quo* with a focus on the challenges of estimating costs. Section 2.1 reviews available data sources. Section 2.2 discusses the estimation and aggregation of cost measures. Finally, Section 2.3 briefly presents results of existing cybercrime surveys.

2.1 Data collection

Police-recorded crime statistics. Traditionally, the prevalence and costs of crime have been measured based on police-recorded crime statistics. The approach works well for many traditional crimes, in particular if a police report is required for victims to receive insurance payments. In the context of cybercrime a number of limitations, quirks and caveats put police-recorded crime statistics to doubt. The first reason is a lack of consensus what constitutes a *cybercrime*. As there is no authoritative definition [Arief et al., 2015], some offenses may be classified as cybercrime when in fact they were not, while others may be concealed within other statistics [Kerr, 2003]. Things become even more difficult when statistics should be compared across countries.

The second reason is underreporting. Businesses are generally reluctant to share information on security incidents or victimization because they worrying about their reputation [Cavusoglu et al., 2004]. According to the 2013 UK Cyber crime report [McGuire and Dowling, 2013], only two percent of businesses report online crime incidents to the police. An EU-wide survey finds that while 79 % of the consumers would report online banking or bank card fraud, only 54 % would report online shopping fraud, and just 37 % the unauthorized access to their email or social media account [European Commission, 2015]. The numbers for identity theft in the US are even more alarming, showing that only 8 % of the victims reported incidents to law enforcement agencies [Harrell, 2015]. The perception that the incident was not significant enough, the belief that the police could not help, or the fact that the victim did not know how to report, are the most common reasons [Harrell, 2015, Rieckmann and Kraus, 2015].

Technical indicators. Another empirical approach to collect data is by direct observation. Security companies and academic researchers studying cybercrime have developed a wealth of tools to observe security incidents. Bilge

et al. [2014] for example have used passive DNS to identify malicious URLs. Kanich et al. [2008] took control over a portion of the spam-sending Storm botnet to measure its size and understand its modes of operation. While these sources are helpful to analyze cybercrime, they present several limitations when it comes to cost estimation. The first and perhaps most important limitation is that these tools are often designed for tracking attack trends rather than the actual impacts. Take phishing for example. The fact that the volume of phishing attacks increases can mean two things: either more people fall victim to these attacks or the attackers are increasing the volume in response to lower success rates [Herley and Florêncio, 2008].

Another set of studies, analyzes the business models of criminals more comprehensively. Levchenko et al. [2011] for example, provide an analysis of the whole spam value chain. McCoy et al. [2012] analyze the business models of online pharmaceutical affiliate programs. While these studies provide better information on the impacts, they are typically tailored to a particular type of cybercrime and only provide a limited view on the bigger picture. Other researchers have observed underground markets to obtain price quotes for criminal artifacts or study criminals' communication channels [e. g., Franklin et al., 2007, Thomas et al., 2013]. While prices may indicate how much money criminals handle, the impact on victims cannot be observed.

A last set of limitations is more practical. Many sources, especially those of commercial vendors, are inaccessible for independent research. That makes it difficult to use them for impact assessment, unless one is willing to simply trust the aggregate statistics that can be gathered from the public reports.

Surveys. In the absence of or in addition to other indicators surveys can be used to measure cybercrime. Since 1996, organizations have been conducting surveys to quantify the diversity and amount of threats that appear when using computers [Computer Security Institute (CSI) and Federal Bureau of Investigation (FBI), 1996]. These surveys vary with regard to the entity in charge, the method, the questionnaire, the surveyed population, and the statistical techniques employed to produce the results.

A few exceptions aside, surveys of cybercrime victims are often based on small, not representative samples. Extrapolation of those results to a wider population is statistically unsound [Center for Strategic and International Studies (CSIS), 2014]. Other surveys do not clarify their methodology, making it hard to evaluate their results [e. g., Deloitte, 2013]. Even surveys with generally robust methods suffer from limitations. Some general crime surveys only spend a few questions on cybercrime and do not cover all types of cybercrime [Jansson and Office, 2007]. Other source of errors include ambiguity, nonresponse, self-selection bias, and the lack of a standard method for counting losses [Hyman, 2013].

2.2 Cost estimation

Estimating the cost of cybercrime has turned out to be equally challenging. A lot of criticism on the current measures of the costs of cybercrime concerns the methods of extrapolation. A particularly prominent example is the 2012 report from Detica, a defense contractor hired by the UK government to estimate the overall cost of cybercrime in the UK [Detica et al., 2011]. They arrived at an unbelievably large figure of 27 Billion lost per annum. Florêncio and Herley [2013] blame the methodologies in many cybercrime reports that almost always exaggerate the numbers on the high side.

When one or more data sources are available, one faces another problem: aggregation. Any measurement instrument captures only a specific class of events that the instrument can observe. Translating the observed events to the more universal population of potential events at different levels presents several challenges. Surveys of financial losses of organizations are particularly challenging to interpret, in this respect, as they always deal with a small number of data points in relation to what they are supposed to represent: all organizations. As outlined by Florêncio and Herley [2013], many of the survey-based estimates of losses are driven by the inclusion of high-value single outliers, which heavily skew and exaggerate results. A handful of respondents formulate the majority of the estimate.

Moreover, estimating individual costs and aggregating them at different levels do not always provide an accurate aggregate cost of cybercrime. Even when aggregation is performed satisfactorily, it only results in a total estimate for a specific type of impacts. For example, a survey among firms can only yield firm-level impacts. It does not take impacts on the consumers, the cost of law enforcement, or other effects into account. Remarkably many studies ignore this issue and are rightly criticized for it [Florêncio and Herley, 2013, Ryan and Jefferson, 2003]. They simply extrapolate firm-level losses to estimate the overall loss to society. But many of the firm-level losses are not losses to society.

Anderson et al. [2013] provide a framework to measure the costs of cybercrime systematically. They distinguish three main cost categories, direct losses, indirect losses and protection costs. Furthermore, they separate cybercrimes from the supporting infrastructure. They use their framework to order cost categories and provide estimates on existing data sources. However, their framework has not been used to inform a survey instrument to estimate the cost of cybercrime.

Outside the cybercrime context – although somewhat related – other parties face the problem of measuring and estimating aggregate losses. The issue is at the core of the insurance industry, but also concerns financial institutions in the context of operational risk management. Operational risks can for example arise from failure to manage employees' use of the IT and from the business practice in itself. The Loss Distribution Approach (LDA)

is a simple way to measure operational risk using frequency and severity of loss data [Frachot et al., 2004]. The LDA has three essential components: (i) a frequency distribution of the number of losses, (ii) a severity distribution of the amount of losses, and (iii) an aggregate loss distribution that combines the two. The distributions to model the losses of cybercrime are structurally comparable to the LDA. Dutta and Perry [2006] survey loss distribution methods used in operational risk management finding that common techniques are: parametric distribution fitting, a method of Extreme Value Theory, and capital estimation based on non-parametric empirical sampling. Different one- and two-parameter distributions can be used to model the loss severity, including, gamma, truncated lognormal, and Weibull.

2.3 Consumer surveys on cybercrime

While many of studies, published by consultancies [e. g., PwC, 2015], the security industry [e. g., Kaspersky Lab, 2015, Ponemon Institute, 2015], or public entities [Federation of small businesses, 2013], report costs of cybercrime for businesses, fewer studies exist for consumers. We summarizes the most important ones for the EU and the US in Table 1.

Table 1: Representative consumer surveys on cybercrime

Region	Year	Crimes	Costs	Study
US	2012, 2014	Identity theft	Yes	Identity theft [Harrell, 2012, 2015]
EU	2012, 2013, 2014	Identity theft, fraud, extortion, scam, malware	No	Special Eurobarometer on Cyber Security [European Commission, 2012, 2013, 2015]
DE	2015	Identity theft, online shopping fraud, phishing, malware	Yes	Cybercrime in Germany [Rieckmann and Kraus, 2015]
UK	2014	Online banking fraud, identity theft, extortion, phishing, malware	Yes	Cybercrime prevalence and impact in the UK [Hernandez-Castro and Boiten, 2014]

For the US, Harrell [2015, 2012] surveyed a large sample of more than 60 000 respondents regarding their costs of identity theft (IDT). While their focus is on IDT in general, some results also apply to cybercrime. They find that in 2014, 7 % of the US consumer have been a victim of identity theft. The most common types have been in the context of credit cards and bank accounts. The survey asks for direct and indirect costs to victims, separating the money stolen by the criminals from additional costs encountered by the victims, such as legal fees, bounced checks, or other miscellaneous expenses. The average financial loss of victims who experienced identity theft incident in the past 12 months is 1 343 $ (with a median of 300 $).

In the EU, the Special Eurobarometer series on Cyber Security is the most important resource on the prevalence of cybercrime [European Commission,

2012, 2013, 2015]. Representative data on different types of cybercrimehave been collected in three subsequent years (2012 – 2014) for all 28 EU member states. The most recent report covers some forms of identity theft among other types, such as online shopping fraud, scam, extortion, and malware infections. 7 % of Internet users in EU have experienced identity theft in 2014 last year [European Commission, 2015]. Note that this number is not comparable to the US survey, as the definition of identity theft differs between both studies. A major shortcoming of the Eurobarometer survey is that it does not ask for the costs of victimization.

Surveys covering the costs of victimization in Europe only exist on the national level in some countries. In 2014, Hernandez-Castro and Boiten [2014] covered a wide range of cybercrime and cyber security related issues for consumers in the UK. Though they only reported rough cost estimates, the survey was focused on extortion losses following infections with ransomware. In 2015 the German Institute for Economic Research (DIW Berlin) reported that the annual costs of cybercrime for consumers in Germany are 3.4 bn € (0.1 % of GDP or 41.5 € per citizen, Rieckmann and Kraus [2015]).

3 Instrument development

Building on the lessons learned from earlier studies we develop an instrument to measure the prevalence of cybercrime and the costs to consumers. We start theoretically, by defining a framework of cost categories in Section 3.1. The following Section 3.2 explains the modeling of individual cost factors and Section 3.3 describes their aggregation.

3.1 Framework of costs

A first step towards accurate estimates is a clear definition of the costs. Where applicable, we call intentional spending *expenses* and unintentional spending *losses*. The aggregate of both is called *costs*. Figure 1 illustrates our framework, which adapts previous work by Anderson et al. [2013] to measure the costs of cybercrime to society. We distinguish three *aggregate* cost categories: direct losses $\mathcal{L}$, indirect losses $\mathcal{I}$, and protection expenses $\mathcal{P}$. Each aggregate cost category can comprise a set of cost *factors* $\{M, T, C, S, \ldots\}$.

$\mathcal{L}$ represents direct losses of cybercrime victims. It is further broken down for different types of cybercrime c which occur with probability p_c. Accordingly, $\mathcal{L}_c$ represents the aggregate loss for one type of cybercrime. Indirect losses $\mathcal{I}$ are not associated with a particular crime, but result from the general prevalence of cybercrime. $\mathcal{I}$ includes effects of behavioral change, market distortions, and so on. Finally, protection expenses $\mathcal{P}$ represent costs for protection which are spent in anticipation of a crime.

Direct losses $\mathcal{L}_c$ primarily include monetary losses M_c and the time lost to deal with an incident T_c. Protection expenses $\mathcal{P}$ can include the money

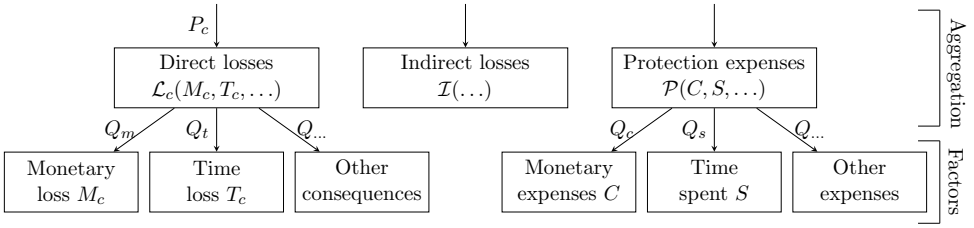

Figure 1: Cost factors and aggregate cost categories of cybercrime

spent on protection measures C and also the time S spent for example to learn about secure behavior or to select appropriate protection measures. Because incidents do not necessarily lead to a monetary loss and not every person spends time on security, we allow that every cost factor materializes with a probability Q. Cybercrimes can also have emotional, social, or even physical impacts on the victim [Arief et al., 2015], denoted as *other consequences* or *other expenses* in Figure 1. Modic and Anderson [2015] study the emotional effects of various types of Internet-related scams, finding that their perceived impact can exceed the monetary losses. Because emotional effects are difficult to quantify in monetary terms, we focus on the money and time that is lost.

We also neglect indirect costs $\mathcal{I}$ in the measurement instrument, because they are inherently different from $\mathcal{L}$ and $\mathcal{P}$ and require observation of the broader economic context which is typically not easy for consumers. Nevertheless, we provide some insights into indirect effects of victimization found in our data in Section 5.4.

3.2 Cost factors

Even though cost factors differ contextually – e. g., C represents the amount of money *intentionally* spent for security and M_c the money that is *unintentionally* lost by victims – all of them can be modeled with semi-continuous random variables. Semi-continuous random variables combine a continuous distribution with point masses at one or more locations [e. g. Min and Agresti, 2002]. They are different from left-censored or truncated variables because all zeros are valid outcomes and not merely proxies for negative or missing responses. Monetary cybercrime losses M_c for example can be modeled as a mixture of zeros, i. e., no loss occurred, and a continuous distribution of positive values, representing the losses. These mixture distributions are often zero-inflated, e. g. because many victims do not lose money.

Methods for estimating the moments of such zero-inflated random variables were first investigated by [Aitchison, 1955]. Cragg [1971] introduced the two-part model (TPM) to model such random variables, arguing that semi-continuous responses should be considered as the result of two processes, one determining whether the response is zero and the other one determining the

actual level if it is *not* zero. Various studies applied the TPM in the context of medical spending [Duan et al., 1983] or other expenditures of individuals and households [Xiao-Hua and Tu, 1999]. The benefit of the TPM is its ability to study hypotheses for both parts individually as well as the compound.

It applies naturally to (preventive) protection expenses C, which are one type of the individual's spending for durable goods, as modeled in the literature [e. g., Duan et al., 1983]. However, we propose that other cost factors $\{M_c, T_c, S \ldots\}$ in our framework can also be modeled with the TPM. We explain our approach in detail for the monetary losses M_c incurred by victims of a particular type of cybercrime. We use the random variable Y to represent the losses for an arbitrary type of cybercrime. Let $y \in [0, \infty[$ denote the realization of Y. For a set of victims v we write y_i as the loss incurred by victim $i \in \{1, \ldots, v\}$. The first part of the TPM is defined by the probability of a loss, denoted as $q = P(y > 0)$. We define an indicator function $\mathbf{1}$ that models this probability of a loss. It takes an expression as single argument. Its value is one if the expression evaluates to true; otherwise it is zero. For example, $\mathbf{1}(2 > 1) = 1$. For the second part of the model, let $z \in]0, \infty[$ be the realization of a random variable Z which models the loss amount *if* a loss has occurred. The probability density function (pdf) of Z is denoted as g_θ, where θ is a vector of the mean and dispersion parameters. This results in the following mixture pdf and maximum likelihood function for Y:

$$f(x) = (1 - q) \cdot \mathbf{1}(x = 0) + q \cdot g_\theta(x), \tag{1}$$

$$L(x) = \prod_{\{x|x=0\}} (1 - q) \prod_{\{x|x>0\}} q \cdot g_\theta(x). \tag{2}$$

Duan et al. [1983] show that the likelihood function can be factored in their use of a two-part zero-inflated regression model.

$$L(x) = \Big[\prod_{\{x|x=0\}} (1 - q) \prod_{\{x|x>0\}} q \Big] \Big[\prod_{\{x|x>0\}} g_\theta(x) \Big] \tag{3}$$

Consequently, both parts of the model can be estimated separately with the maximum likelihood method if q and Z are independent. The maximum likelihood of the first part can be simplified as the mean of the indicator function, which is simply the fraction of victims suffering a loss. The second part can be evaluated by fitting different candidate loss distributions for g. Accordingly, the expected value for Y can be written as:

$$E(Y) = E(f(x)) = q \cdot E(g_\theta). \tag{4}$$

Using the TPM, we can study the probability of monetary losses q and the loss distribution under the condition of a loss Z independently. Furthermore, the model allows us to analyze the compound expected loss $E(Y)$. In the remainder of the paper we use a binary random variable Q for the first part

of the TPM, such that $E(Q) = q$ (Note the slight notation overload). The TPM can be applied to all cost factors, including none-monetary factors, such as the time spent for protection S.

3.3 Aggregate cost categories

To aggregate the cost factors, we propose a general utility function $U(\mathbb{X})$ with realizations $u \in [0, \infty[$, which models the *disutility* or *badness* of costs, losses, and other consequences. $U(\mathbb{X})$ takes a vector of cost factors $\mathbb{X}$ as input and evaluates to positive, monetary values. The results of U are monotonically increasing in every element of the input $\mathbb{X}$. Furthermore, U is defined such that an individual is indifferent between alternative a) nothing happens and b) experiencing $U = 100$ *and* receiving 100 €.

We explain U for the aggregate protection expenses $\mathcal{P}(C, S)$. Let C and S be semi-continuous random variables modeling costs $c \in [0, \infty[$ and time $s \in [0, \infty[$ spent for protection. Both follow the structure of Y described in Section 3.2. Furthermore, let $\alpha \in [0, \infty[$ be a conversion factor which must not be related to cybercrime but converts time units to monetary values. Then we can define the aggregate expenses for protection $\mathcal{P}$ as a linear combination of C and S:

$$\text{Protection expenses} = \mathcal{P}(C, S) = C + \alpha \cdot S. \tag{5}$$

Because $\mathcal{P}$ is linear, the expected value $E(\mathcal{P})$ can be written as:

$$E(\mathcal{P}) = E(C + \alpha \cdot S) = E(C) + \alpha \cdot E(S). \tag{6}$$

To aggregate cybercrime losses $\mathcal{L}$, we take multiple types of cybercrime into account. For each type $c \in \{\mathcal{C}\}$, where $\mathcal{C}$ is a set of nominal categories, let $p_c = P(\text{victim of c})$ be the probability of being victimized. Furthermore, let $M_c \in [0, \infty[$ be the random variable modeling the monetary losses and $T_c \in [0, \infty[$ the time to deal with an incident of type c. Both, M_c and T_c follow the structure of a semi-continuous variable, such as Y in Section 3.2. The loss for one type of cybercrime follows the disutility function U and is defined as:

$$\text{Cybercrime loss} = \mathcal{L}_c(M_c, T_c) = M_c + \alpha \cdot T_c. \tag{7}$$

with an expected value $E(\mathcal{L}_c)$:

$$E(\mathcal{L}_c) = E(M_c + \alpha \cdot T_c) = E(M_c) + \alpha \cdot E(T_c). \tag{8}$$

Assuming that the processes of falling victim to different types of crime are independent, we weigh the disutility of being victimized $\mathcal{L}_c(M_c, T_c)$ with the probability of being victimized P_c. The total cybercrime losses $\mathcal{L}$ are the sum over all weighted disutilities:

$$\mathcal{L} = \sum_{c \in \{\mathcal{C}\}} P_c \cdot E(\mathcal{L}_c) = \sum_{c \in \{\mathcal{C}\}} P_c \cdot (E(M_c) + \alpha \cdot E(T_c)). \tag{9}$$

4 Descriptive results

We have instantiated our measurement instrument to collect data on the costs of cybercrime in six European countries. Section 4.1 selects cost categories from the instrument and explains how we have measured them empirically. Section 4.2 describes the sample and the fieldwork. Finally, Section 4.3 reports descriptive results of consumer-facing cybercrime in the six countries.

4.1 Measurement instrument

Translating the measurement instrument into a survey requires several decisions, concerning the sampling, coverage of cost categories, and the selection of relevant types of cybercrime. Figure 2 illustrates the instance of the measurement instrument we used in the survey. The cost factors and aggregated costs in the lower part correspond to Figure 1 (in Section 3.1). Our sampling approach is added in the upper part. White boxes represent parts of the instrument we cover in the survey, gray boxes are *not* covered, and light gray boxes implicate a coverage based on assumptions.

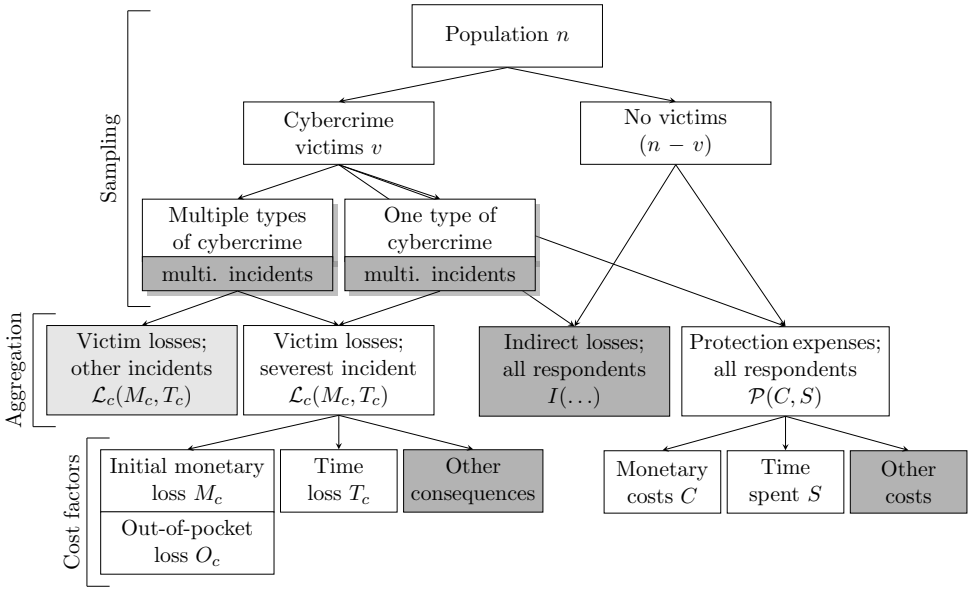

Figure 2: Instantiation of the measurement instrument used in the survey

The cost factors $\{M_c, T_c, O_c, C, S\}$ follow the structure of semi-continuous random variables, such as Y in Section 3.2. The *disutilities* of aggregated costs or losses $\{\mathcal{L}_c, \mathcal{P}\}$ follow the general disutility function U presented in Section 3.3. While our instrument allows for the inclusion of arbitrary cost factors, we only consider monetary factors $\{M_c, C\}$ and time spent $\{T_c, S\}$ and neglect other consequences (see gray rectangles in Figure 2). Monetary

losses are further broken down into *out-of-pocket losses* O_c, which are the part that is ultimately lost by the victim, and *industry losses* $M_c - O_c$, which are covered by service providers through compensation payments.

We measure the cost factors empirically using the following logic. The probabilities of incurring a loss and therefore the condition in the TPM $\{\widehat{\mathrm{Q}}_{c,m,\mu}, \widehat{\mathrm{Q}}_{c,o,\mu}, \widehat{\mathrm{Q}}_{c,t,\mu}, \widehat{\mathrm{Q}}_{c,\mu}, \widehat{\mathrm{Q}}_{s,\mu},\}$, are measured by the empirical mean (μ) of the indicator function 1. The probability of incurring a monetary loss in a cybercrime incident of type c is, for example:

$$E(\widehat{\mathrm{Q}}_{c,m}) = \widehat{\mathrm{Q}}_{c,m,\mu} = \frac{1}{v_c} \cdot \sum_{i=1}^{v_c} 1(m_{c,i} > 0), \tag{10}$$

where $m_{c,i}$ is the point estimate for the monetary loss of the i-th victim. The probability of falling victim to a type of cybercrime $\widehat{\mathrm{P}}_{c,\mu}$ is estimated accordingly, based on the overall sample n. We estimate conditional losses ($\mathrm{Z}_{c,m}$) using different methods. We compare the empirical mean ($\widehat{\mathrm{Z}}_{c,m,\mu}$) and median ($\widehat{\mathrm{Z}}_{c,m,50}$) with the theoretical mean ($\widetilde{\mathrm{Z}}_{c,m,\mu}$) and median ($\widetilde{\mathrm{Z}}_{c,m,50}$), which are based on conditional loss distributions. The parameter vector $\hat{\theta}$ for the conditional loss distribution $\widetilde{\mathrm{Z}}_{c,m}$ is estimated by fitting different candidate loss distributions g_θ to the point estimates of the costs reported in the survey. Furthermore, we calculate three different indicators for unconditional losses. First, the *expected monetary loss* indicator ($\ddot{\mathrm{M}}_{c,\mu}$), which is the mean of the conditional loss distribution scaled by the probability of the condition:

$$E(\ddot{\mathrm{M}}_c) = \ddot{\mathrm{M}}_{c,\mu} = \widetilde{\mathrm{Z}}_{c,m,\mu} \cdot \widehat{\mathrm{Q}}_{c,m,\mu}. \tag{11}$$

Notational convention: the double dots imply that the indicator combines direct empirical estimates ($\widehat{\mathrm{Q}}_{c,m,\mu}$) and estimates via theoretical distribution functions ($\widehat{\mathrm{Z}}_{c,m}$). Second, an *adjusted median loss* indicator ($\ddot{\mathrm{M}}_{c,*}$), which shifts the conditional median by the probability of a loss:

$$E(\ddot{\mathrm{M}}_c) = \ddot{\mathrm{M}}_{c,*} = \widetilde{\mathrm{Z}}_{c,m,\lambda}, \text{ with shift } \lambda_{c,m} = \frac{1 - \widehat{\mathrm{Q}}_{c,m,\mu}}{2 \cdot \widehat{\mathrm{Q}}_{c,m,\mu}}. \tag{12}$$

And third, a *harmonized loss* indicator ($\ddot{\mathrm{M}}_{c,50}$), which is the median of the conditional loss distribution ($\widetilde{\mathrm{Z}}_{c,m,50}$) scaled by the probability of the condition:

$$E(\ddot{\mathrm{M}}_c) = \ddot{\mathrm{M}}_{c,50} = \widetilde{\mathrm{Z}}_{c,m,50} \cdot \widehat{\mathrm{Q}}_{c,m,\mu}. \tag{13}$$

The results of $\ddot{\mathrm{M}}_{c,50}$ can be interpreted as expected losses of victims under the assumption of Bernoulli losses where the unknown shape of the loss distribution is simplified to its median. We discuss benefits and shortcomings of each method in Section 5.1. The same methods are used for the remaining monetary estimates: out-of-pocket losses (O_c) and protection expenses(C)}.

Unlike for monetary losses, losses of time $\{T_c, S\}$ are measured on ordinal scales. Our unconditional estimates for the time lost $\widehat{\mathrm{T}}_{c,\mu}$ and $\widehat{\mathrm{S}}_{\mu}$ are calculated as the mean of the conditional interval centers $\widehat{\mathrm{Z}}_{\mu}$ scaled by the probability of the condition $\widehat{\mathrm{Q}}_{\mu}$. Consult Table 8 in the appendix for further information on variables and cost factors.

4.2 Sampling

We collected representative data for adult Internet users in the following six European member states (in protocol order): Germany (DE), Estonia (EE), Italy (IT), the Netherlands (NL), Poland (PL), and the United Kingdom (UK).[1] This selection creates a diverse set of countries in terms of geographic location, maturity of the information and communication infrastructure, Internet usage, and cybercrime prevalence as reported in previous surveys. The fieldwork was carried out between July and October 2015 in the respective mother tongue for each country. Respondents are 18 years and older and use the Internet for personal purposes at least once per month. The sample was drawn with random digit dialing, an established technique in the target countries, with quotas set on age, gender, and region. Overall $n = 6\,394$ response sets have been collected. The demographics of the sample and the subpopulation of victims are reported in Table 10 in Appendix 6.3.

Because cybercrime victims are relatively rare, this approach leaves us with a small sample to estimate costs for the subpopulation of victims. Acknowledging earlier critique [Florêncio and Herley, 2013], 256 victims of cybercrime were included by oversampling, leading to an overall population of ($v = 1\,242$) victims. Oversampling is accounted for in the analysis by inverse probability weighting. Naturally, victims v may have experienced multiple types of cybercrime $c \in \mathcal{C}$, or experienced one type more than once. Thus, c can lead to $i \in \{1, 2, \ldots\}$ incidents. The optimal approach is an exhaustive measurement of all incidents i for all types of cybercrime c for every victim v. Because this approach is impractical, we reduced the set of incidents by asking each victim only about the severest incident. Accordingly, we do not consider multiple incidents for a single type of crime, i. e., we set $i = 1$. This decision is based on the assumption that multiple victimization of the same type of crime is rare. If multiple victimization happens across different types of crime, we recorded this fact, but asked only about the losses of the severest among all types (see the light gray rectangle in Figure 2). For the aggregation, we impute the unobserved losses with summary values obtained from all victims who reported only one or the severest incident for the respective type of crime. This rule may introduce some bias. We tend to overestimate aggregate losses, but it is safe to interpret our values as upper bounds.

[1]This survey was conducted as part of the European research project E-CRIME (`http://ecrime-project.eu/`) under grant number 607775.

4.3 Cybercrime prevalence

Types of crime analyzed. Cybercrime spans a wide range of different types, which differ with regard to the motivation of attackers and the impact on victims. These types c must be mutually exclusive to break down the losses. This is difficult as authoritative definitions or descriptions of types of cybercrime are missing [Arief et al., 2015]. Population surveys are best suited to study types of crime with a direct relationship between the victim and the criminal. Table 2 shows the seven types of profit-motivated cybercrime that we selected for our survey, along with the wording in the English version of the questionnaire. We include four types of identity theft (IDT): IDT wrt. online banking (OB), bank cards (BC), PayPal, and online shopping (OS). Furthermore, we ask for OS fraud, scams, and extortion. The wording for each type of crime may differ as it was formulated to be as comprehensible as possible for the respondents.

Table 2: Consumer-facing cybercrimes $\mathcal{C}$ with question wording

Thinking of the past 5 years, have you ever personally experienced any of the following?	
IDT wrt. OB	*Someone getting access to your bank account password (to buy something in your name, take money from your account, open a credit etc.)*
IDT wrt. BC	*Someone getting access to your bank card security numbers (to buy something in your name)*
IDT wrt. PayPal	*Someone getting access to your PayPal password (to buy something in your name, or take money from your account)*
IDT wrt. OS	*Someone getting access to your online shopping account (e. g., Amazon etc.), to buy something in your name*
OS fraud	*Products or services which you have purchased online not being delivered, being defective or of different quality than advertised*
Extortion	*Someone extorting money from you to recover access to an account or your computer*
Scams	*Someone tricking you to transfer money to a fraudulent website*
Malware	*Do the following statements apply to you? During the past 5 years, I have had malware/viruses on my computer*

Identity theft (IDT), Online shopping (OS), Online banking (OB), Bank cards (BC)

The selected types of cybercrime can be broadly categorized by third party involvement. The first three types concern IDT with the involvement of financial and payment services. The second category contains crimes related to ecommerce. And the third category crimes which typically not involve a third party. Our selection of crimes is not exhaustive. We exclude emotionally and politically motivated offenses, such as *cyber-stalking*, *cyber-bullying*, or *hacktivism*, and crimes typically not targeted against consumers, such as *denial of service attacks*. We also excluded criminal activities where consumers are merely affected indirectly or which are part of the cybercriminal

infrastructure [Anderson et al., 2013, p. 6], such as spam emails or phishing. This avoids double counting, as these activities are precursors to the selected crime types. In order to compare our data to previous surveys which report victimization rates, but did not attempt to estimate costs [e. g., European Commission, 2015], we also asked about malware.

Incidents. Table 3 shows the prevalence of cybercrime in the six surveyed countries. Each cell represents the percentage of adult Internet users who reported to have experienced any type of crime during the last five years.

Table 3: Incident rates of cybercrime by type and country

	Internet users victimized in the last 5 years					
Cybercrime type	DE	UK	NL	PL	EE	IT
IDT wrt. OB	1.4 %	3.3 %	1.4 %	1.2 %	1.0 %	1.1 %
IDT wrt. BC	3.5 %	4.8 %	2.0 %	0.9 %	1.7 %	2.7 %
IDT wrt. PayPal	2.0 %	2.3 %	0.7 %	0.8 %	0.4 %	0.9 %
IDT wrt. OS	4.3 %	4.1 %	1.1 %	0.9 %	0.8 %	1.9 %
OS fraud	8.4 %	9.0 %	10.3 %	9.7 %	9.1 %	5.0 %
Extortion	5.1 %	2.8 %	1.1 %	1.4 %	0.6 %	1.5 %
Scams	5.0 %	4.4 %	2.3 %	3.4 %	1.7 %	2.4 %
Total	22.2 %	21.6 %	15.7 %	13.9 %	13.2 %	12.1 %
For comparison:						
Malware	51.5 %	50.5 %	48.8 %	68.1 %	55.7 %	60.1 %

Germany (DE), United Kingdom (UK), Netherlands (NL), Poland (PL), Estonia (EE), Italy (IT)

Total cybercrime is most prevalent in Germany (22.2 %) and the UK (21.6 %). Italy on the other end is least affected (12.1 %). Online shopping fraud is the most prevalent type of cybercrime with incident rates of almost 10 % in all countries, except Italy, where it is only 5 %. Our results likely underestimate the real extent of online shopping fraud because victims have been identified using a proxy which added additional constraints, i. e., only victims who reported to have lost money and where not able to recover their losses completely. Section 6.2 in the appendix discusses the proxy.

IDT wrt. bank cards is comparably high in the UK (4.8 %) and Italy (2.7 %). Extortion has been mostly experienced in Germany (5 %). Malware infection has been encountered by at least twice as many respondents then all other crimes combined. In Italy and Poland the ratio is even higher. This supports our argument that malware is a precursor for many different types of cybercrime. The numbers in Table 3 include multiple victimization. With 79 %, the majority of the victims reported only one incident in the last five years. 15 % experienced two incidents of cybercrime and only 6 % fall victim to more than two types of cybercrime.

5 Results on cost estimates

This section is structured along the aggregate cost categories introduced in Section 3.1. We estimate the victims' losses $\mathcal{L}$ in Section 5.1 and protection expenses $\mathcal{P}$ among all consumers in Section 5.2. Both sections explain the data, describe the estimation procedure and present the results. Section 5.3 aggregates the cost estimates per country. Finally, Section 5.4 discusses indirect losses $\mathcal{I}$ to society.

5.1 Direct losses of cybercrime victims

The direct losses $\mathcal{L}$ are measured based on the impacts reported by the $v = 1\,242$ victims for their severest incident. Across all types of crime, the majority of victims (90.54 %) reports a loss of time to deal with the incident and a large part also reports monetary losses (62.05 %). Only a minority reports personal (13.45 %), professional (3.79 %), or other problems (10.5 %). Table 11 in Appendix 6.4 shows all impacts broken down for each type of cybercrime.

Data preparation. Part of the data preparation concerned the imputation of missing values. Overall, 712 victims reported monetary losses. These losses were reported either as point estimates (608 cases) or in one of nine ordinal categories[2], if the respondent could not recall an exact value (86 cases, 12.08 %). Instead of imputing the center of the ordinal interval, we impute the theoretical median of each interval based on fitted loss distributions. We estimate the loss distributions for every type of cybercrime individually using the approach described in the next paragraph.

Furthermore, 4 victims refused to report the amount lost and 14 victims did not know. For the refusal cases (0.66 %) we imputed the median of the loss distribution. As the victims reported a loss, but no value, we believe this is the best possible approach. For the 14 *don't know* responses (2.3 %) we imputed the median of the smallest loss category for each type of cybercrime. We do not drop the cases, because the respondents reported a loss and we assume that the losses have been small if respondents cannot recall an order of magnitude.

Cost estimation. We estimate summary statistics of the monetary losses of victimization M_c for each type of cybercrime across all six countries. We choose this approach because the total number of incidents with monetary

[2] Question: "How much money would you say you have lost due to this incident altogether (including fees you may have had to pay, etc.)?"; cost categories for €-countries and the UK in the respective currency: $[1 : 50], [51 : 100], [101 : 200], [201 : 500], [501 : 1\,000], [1\,001 : 5\,000], [5\,001 : 10\,000], [> 10\,000]$. For Poland the categories are adjusted to equivalents in Zloty.

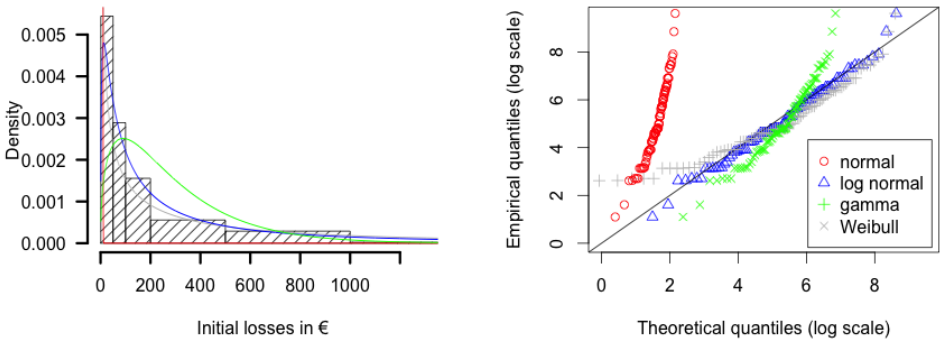

Figure 3: Monetary losses of scams (Z_c); Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate distributions on log scale

losses is too small to derive country-specific figures.[3] Consequently, seven loss distributions Z_c are fitted for the initial monetary losses, one for each type of cybercrime c.

To inform our choice of candidate distributions, we explore the data and observe that the distribution of Z_c is skewed to the left for all losses. Figure 3 for example shows a histogram of the monetary losses of scams, with the breaks based on the categorical intervals used in the questionnaire. For a better visualization the x-axis is truncated at a loss of 1 200 €, cutting off a part of the right tail (11 incidents). We fitted log normal, gamma, and Weibull distributions to the point estimates for each type of crime. These are commonly used to model monetary losses in the operational risk management literature [e. g., Dutta and Perry, 2006]. We also fitted a normal distribution for comparison. The right part of Figure 3 shows the Q–Q plot of the four different loss distributions for scams on a log scale, indicating best fit for the log normal distribution.

Table 13 in the appendix shows the parameter estimates $\hat{\theta}_c$ for all types of cybercrime along with the relative goodness-of-fit indicators AIC and BIC for each candidate distribution. According to both, AIC and BIC, the log normal distribution fits the data best for all types of crimes except IDT wrt. PayPal and extortion. For these two types the Weibull distribution performs slightly better ($\Delta\text{AIC} = +1$ for IDT wrt. PayPal and $\Delta\text{AIC} = +2$ for extortion). As the number of victims v_c is small in both cases ($v_c < 15$) and ΔAIC is not substantial, we estimate all parameters using the log normal distribution. Histograms and Q–Q plots for all types of cybercrime can be found in Appendix 6.4.

Figure 4 summarizes the distribution fitting by a Q–Q plot of the log normal loss distributions for all seven types of cybercrime. Deviations are

[3]See Table 12 in the appendix for further explanation.

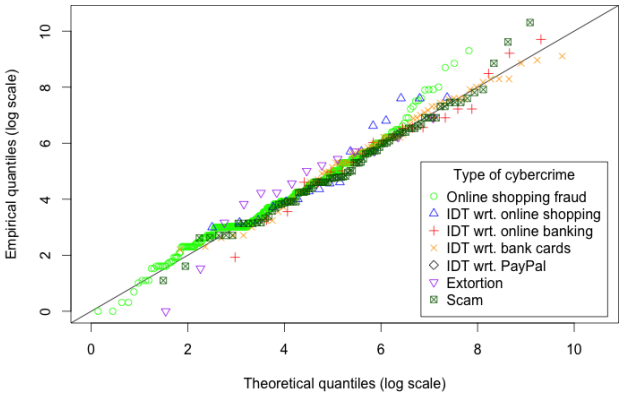

Figure 4: Q–Q plot of the all log normal loss distributions Z_c for the initial losses from all types of cybercrime

mostly in the tails. While deviations in the lower tail ($z_c < exp(3) \approx 20$ euro) are unproblematic, deviations in the upper tail need to be considered. We find that losses of online shopping fraud, IDT wrt. online shopping and online banking are likely to be underestimated by the log normal distributions.

Monetary loss estimates. Table 4 documents the monetary loss estimates along two dimensions. The first dimension compares empirical and theoretical estimates in the conditional case $\mathrm{Z}_{c,m}$. The second dimension includes the condition $\mathrm{Q}_{c,\mu}$ for the different estimators of unconditional losses M_c.

Table 4: Estimates of initial monetary losses for each type of cybercrime

	Empirical			Theoretical		Combined		
	Condition	Conditional losses (€)				Unconditional losses (€)		
Cybercrime (c)	$\widehat{\mathrm{Q}}_{c,m,\mu}$	$\widehat{\mathrm{Z}}_{c,m,\mu}$	$\widehat{\mathrm{Z}}_{c,m,50}$	$\widetilde{\mathrm{Z}}_{c,m,\mu}$	$\widetilde{\mathrm{Z}}_{c,m,50}$	$\ddot{\mathrm{M}}_{c,\mu}$	$\ddot{\mathrm{M}}_{c,*}$	$\ddot{\mathrm{M}}_{c,50}$
	(1)	(2)	(3)	(4)	(5)	(1)×(2)		(1)×(5)
IDT wrt. OB	33 %	**2106**	630	2585	466	862	0	**155**
IDT wrt. BC	35 %	1165	403	1684	329	583	0	114
IDT wrt. PayPal	24 %	2039	**1000**	**4425**	**488**	**1079**	0	119
OS Fraud	91 %	174	50	131	54	119	**45**	49
IDT wrt. OS	17 %	452	93	447	139	77	0	24
Extortion	13 %	197	131	406	74	53	0	10
Scam	45 %	1078	176	783	198	353	0	89

Estimates in €; Based on the severest incident ($v = 1\,242$)

Let us first consider the conditional losses $\mathrm{Z}_{c,m}$ to compare different loss estimates, *if* a loss occurred. The empirical mean ($\widehat{\mathrm{Z}}_{c,m,\mu}$) consistently reports higher losses than the median ($\widehat{\mathrm{Z}}_{c,m,50}$) for all types of crime. It is more than

three times bigger for IDT wrt. online shopping, IDT wrt. regard to online banking, and online shopping fraud. For scams the mean estimates are even five times larger than the median estimates. An inspection of the data shows that this is driven by a single victim reporting a loss of 30 000 €. Similarly, the theoretical mean ($\widetilde{\mathrm{Z}}_{c,m,\mu}$) is always bigger than the median ($\widetilde{\mathrm{Z}}_{c,m,50}$).

The second dimension represents unconditional losses, by including the condition ($\widehat{\mathrm{Q}}_{c,m,\mu}$). The condition shows that many severest incidents do not lead to a monetary loss, in particular for extortion and IDT wrt. online shopping. Online shopping fraud victims lose money most often (91 %).[4] These losses, however, are also the smallest across all reported cybercrimes. Comparing the combined aggregation methods shows that the *expected monetary loss* indicator ($\ddot{\mathrm{M}}_{c,\mu}$) likely overestimates the losses because it is based on the theoretical mean $\widetilde{\mathrm{Z}}_{c,m,\mu}$. While in principle more robust against outliers in the right tail, the *adjusted median loss* indicator ($\ddot{\mathrm{M}}_{c,*}$) is zero as soon as 50 % of the victims have losses. This is the case for all types of crime, except online shopping fraud. Our proposed *harmonized loss* indicator ($\ddot{\mathrm{M}}_{c,50}$) combines the best of both approaches. It is robustness against outliers and can handle data with high zero-inflation. Of course the statistical interpretation of the *harmonized loss* indicator is not straight forward and extrapolated numbers should be handled with high caution.

Table 5: Estimates of time losses for seven types of cybercrime

	Condition		C. losses	U. losses
Cybercrime	$\widehat{\mathrm{Q}}_{c,t,\mu}$	$P(Z_{c,t} > 20)$	$\widehat{\mathrm{Z}}_{c,t,\mu}$	$\widehat{\mathrm{T}}_{c,\mu}$
	(1)	(2)	(3)	(1)×(3)
IDT wrt. OB	95.24 %	14.29 %	7.29 hrs	7.11 hrs
IDT wrt. BC	**96.21** %	15.15 %	7.41 hrs	7.29 hrs
IDT wrt. PayPal	95.24 %	16.67 %	7.38 hrs	7.21 hrs
OS fraud	88.34 %	12.02 %	6.45 hrs	5.82 hrs
IDT wrt. OS	95.65 %	10.14 %	6.31 hrs	6.22 hrs
Extortion	93.42 %	17.11 %	8.16 hrs	7.62 hrs
Scams	92.20 %	**20.57** %	**8.47** hrs	**8.05** hrs

Conditional (C.), Unconditional (U.) Based on the severest incident ($v = 1\,242$)

Time lost. The time lost by victims T_c was measured in hours (hrs) using an ordinal question with five categories[5]. 57 cases are missing due to *don't know* responses and 50 victims refused to provide an answer. We impute zero for *don't know* responses (2.09 %), assuming that respondents who cannot

[4] This number is positively biased, by constraints in the proxy that identifies victims of online shopping fraud. See appendix 6.2.

[5] Question: "How much time have you spent trying to solve the problem (please think of the total number of hours you have personally spent)"; categories: [0 hrs, 1 hr],[1 hr, 10 hrs],10 hrs, 20 hrs],[> 20 hrs]

answer to a categorical question most likely only lost an insignificant amount of time. For the refusals (0.95 %) we impute the central category [1 hr, 10 hrs], assuming that some loss has happened. Table 5 shows the estimates, which are structured into conditional and unconditional losses.

The vast majority of victims experiences losses of time. For scams and extortion, the biggest number of respondents fall into the highest loss category. Every fifth scam victim has spent more than 20 hrs to deal with the incident. Accordingly, most time is lost for scams (8.05 hrs). The least average time lost is reported for online shopping fraud. Note, that as a result of the categorical mean the variation of the overall average is rather small.

Cybercrime impact maps. We jointly analyze the harmonized monetary losses and the time lost by the victims for all types of cybercrime using a *cybercrime impact map* as depicted in Figure 5. Each type of cybercrime is represented by a black circle. The average time lost ($\widehat{\mathrm{T}}_{c,\mu}$) defines the location of a crime on the x-axis and the harmonized estimate for the initial monetary loss ($\ddot{\mathrm{M}}_{c,50}$) defines the location on the y-axis. The further a crime moves to the upper right of the map, the higher is its *disutility* and, consequently, the incurred losses for the victims ($\ddot{\mathcal{L}}_c$).

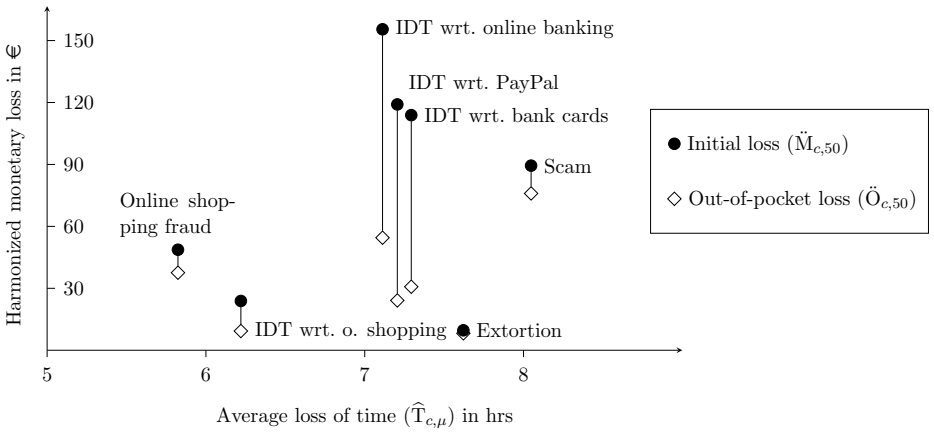

Figure 5: Cybercrime impact map

In addition to initial monetary losses, we analyze the out-of-pocket losses ($\ddot{\mathrm{O}}_{c,50}$) which represent the victim's losses after compensation payments. The out-of-pocket losses define a second location for each cybercrime on the y-axis (illustrated with white diamonds). Compensation payments were measured on an ordinal scale with six brackets representing the percentage of losses the victims were able to recover[6]. We calculate point estimates for O_c by multiplying each initial loss m_c with the center of the interval of each scale

[6]Question: "To what extent were you able to get your money back?"; scale levels: [0],[0, 25 %],[25 %, 50 %],[50 %, 75 %],[75 %, 100 %],[100 %]

level. The unconditional out-of-pocket loss ($\ddot{\mathrm{O}}_{c,50}$) is then estimated analogous to the initial losses, using the harmonized loss estimator for each type of cybercrime[7].

The cybercrime impact map illustrates that the seven types of cybercrime against consumers fall into the three categories, which are in line with the categories based on third party involvement[8]. The first category comprises incidents related to ecommerce. It is characterized by the lowest impact on consumers in terms of disutility. Online shopping fraud and IDT wrt. online shopping lead to small monetary losses (also small compensation payments) and the smallest loss of time. The second category relates to payment and financial services. It comprises IDT wrt. online banking, bank cards, and PayPal. While these crimes lead to the highest initial losses, service providers cover a large part of the costs through compensation payments. Consequently, the harmonized out-of-pocket losses for consumers are comparable to the other types of cybercrime. While we suspected that receiving compensation requires more time, we could not find evidence for this effect in our data. The third category of crimes – extortion and scams – does not involve a third party. These crimes turn out to be most time-consuming and victims do not receive any compensation. Interestingly, losses to extortion were the smallest of all crime types during the field time. Recent epidemics of ransomware might have changed this picture [Trendmicro, 2016]. According to our impact map, scams are the most *dangerous* type of cybercrime because they lead to the highest initial and out-of-pocket loss and require the longest time to deal with.

5.2 Expenses for protection

Protection expenses are estimated for all respondents in the surveyed countries ($n = 6\,394$). This section is equally structured as the estimation of initial losses and uses the same estimation procedure (see Section 5.1). We report estimates for monetary expenses ($\ddot{\mathrm{C}}_{d,50}$) and the time that consumers spend for administration ($\ddot{\mathrm{S}}_{d,\mu}$). The vast majority of consumers has protection software installed on their systems ($> 90\,\%$), a substantial part purchased commercial products ($> 62\,\%$), and $> 71\,\%$ reported to have spent time to manage protection measures.

Data preparation. 3993 respondents reported to have spend money for protection measures in the last five years. Responses are reported either as point estimate (2470 cases) or in one of eight ordinal categories (1523 cases)[9]. Point

[7] Table 14 in the appendix shows parameter estimates $\hat{\theta}$ for the distribution of out-of-pocket losses O_c

[8] As introduced in Section 4.1.

[9] Question: "Overall, during the past 5 years, how much money would you say you have spent on protection software (for example anti-virus or firewall)?"; cost categories for €-countries and the UK in the respective currency: $[1:50], [51:100], [101:200], [201:500], [501:1\,000], [1\,001:5\,000], [5\,001:$

estimates for the ordinal responses are imputed using the log normal median for each interval. For the 49 refusal cases (0.77 %) we imputed the overall median of the expense distribution. As respondents reported expenses, but no value, we believe this is the best possible approach. For the larger number of 658 *don't know* responses (10.29 %) we imputed no expenses, arguing that people are likely to know whether they spend money for a product. This is a conservative approach to estimate protection expenses. Two respondents reported expenses: $> 10\,000$ € on the ordinal scale. These were not imputed, because they seem unrealistic and substantially exceed the highest reported point estimates (5 000 €).

Expense estimation. Estimates for protection expenses $\ddot{\mathrm{C}}_{d,50}$ are derived for each country individually. Consequently, six cost distributions are fitted, one for each country d. $\ddot{\mathrm{Q}}_{d,\mu}$ denotes the percent of consumers, who spend money for protection. The empirical parameter estimates $\hat{\theta}$ are estimated by fitting different candidate cost distributions. We tried a log normal, gamma, Weibull, and normal distributions to the cost data for each country, because these are typically used to model expenses with two-part models in the literature [e. g., Duan et al., 1983, Min and Agresti, 2002]. The relative quality indicators suggest a log normal distribution for Germany, Italy, and the UK and a Gamma or Weibull distribution for Estonia, the Netherlands, and Poland. As the differences in the qualitative fit indicators are small and the Q–Q plots for Estonia, the Netherlands, and Poland show a good fit of the log normal distribution, in particular in the upper tail, we estimate the costs for all countries using the log normal distribution.[10]

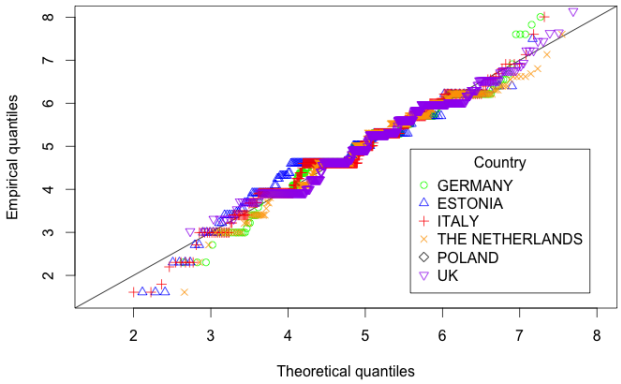

Figure 6: Q–Q plot for protection expenses

10 000], [> 10 000]. For Poland the categories are adjusted to equivalents in Zloty.

[10] Table 15 in the appendix shows the parameter estimates $\hat{\theta}_c$ for each country along with relative quality indicators AIC and BIC for each distribution. The empirical loss distributions and Q–Q plots for all country can be found in appendix 6.4

To summarize Figure 6 shows the Q–Q plot for the log normal distribution of expenses in each country. The empirical quantiles are characterized by *steps*, which are formed by common replies for round values, such $3.91 \approx log(50)$. The theoretical distributions overestimate a few values in the lower tail $< 2.5 \approx log(12)$ and underestimates slightly in the upper tail.

The time consumers spent to manage protection measures was measured in hours using a question with five ordinal categories and a time frame of one year[11]. 150 *don't know* responses and 37 refusals are imputed with zeros, i. e. the respondent did not spend any time. Results are multiplied by five, to measures all cost factors for the same time frame. The expected time spent $\widehat{\mathrm{S}}_{d,\mu}$ by consumers is estimated by the average of the interval centers of the ordinal question.

Loss estimates. Table 6 reports the expenses ($\ddot{\mathrm{C}}_{d,50}$) and the time ($\widehat{\mathrm{S}}_{d,\mu}$) spent for protection. Even though, not explicitly noted in Table 6, the conditional expenses can also be analyzed along the empirical and theoretical dimension. As for the cybercrime losses, the empirical mean $\widehat{\mathrm{Z}}_{c,d,\mu}$ is constantly higher than the empirical median $\widehat{\mathrm{Z}}_{c,d,50}$. However, the effect is smaller than for the cybercrime losses.

Table 6: Estimates of protection expenses per country

	Monetary expenses P (€)					Time spent S (hrs.)		
	Cond.	Conditional (C.)			Unc.	Cond.	C.	Unc.
Cntry	$\widehat{\mathrm{Q}}_{c,d,\mu}$	$\widehat{\mathrm{Z}}_{c,d,\mu}$	$\widehat{\mathrm{Z}}_{c,d,50}$	$\widetilde{\mathrm{Z}}_{c,d,50}$	$\ddot{\mathrm{C}}_{d,50}$	$\widehat{\mathrm{Q}}_{s,d,\mu}$	$\widehat{\mathrm{Z}}_{s,d,\mu}$	$\widehat{\mathrm{S}}_{d,\mu}$
	(1)	(2)	(3)	(4)	(1)×(4)	(6)	(7)	(6)×(7)
DE	52 %	224	150	155	80	84 %	**20.11**	**16.88**
EE	16 %	141	100	91	14	55 %	12.1	6.72
IT	42 %	192	100	118	50	78 %	14.31	11.15
NL	46 %	226	200	164	75	69 %	17.67	12.27
PL	60 %	124	86	82	49	73 %	16.05	11.78
UK	58 %	**262**	**195**	**184**	**106**	67 %	14.07	9.37

Unconditional (Unc.); Based on the full sample ($n = 6\,242$); Germany (DE), Estonia (EE), Italy (IT), Netherlands (NL), Poland (PL), United Kingdom (UK)

We only report our *harmonized loss* indicator for the unconditional expense estimates. Table 6 shows that roughly half of the respondents spend money on protection measures across all countries, except Estonia where only 16 % reported expenses. Accordingly, the percent of respondents spending time on protection is also the smallest in Estonia (55 %). While also spending only a small amount of time, consumers in the UK report the highest expenses for protection. Germans might be called most protective, as they invest

[11] Question: "And now, thinking of the past 12 months, how much time did you spend learning about and installing protection software?"; categories: [0 hrs, 1 hr],[1 hr, 10 hrs],[10 hrs, 20 hrs],[> 20 hrs]

the largest amount of time and also the second largest amount of money into protection measures. In Poland consumers are likely to invest into protection measures, but their expenses are the smallest.

5.3 Aggregate cost estimates

We aggregate the overall costs per country using the approach outlined in Section 3.3. To convert estimates from time scales to monetary scales we define $\hat{\alpha}_d$ as the median of gross hourly earnings for each country Eurostat [2010]. Table 7 shows the estimated values for the aggregated cybercrime losses $\ddot{\mathcal{L}}_d$ and protection expenses $\ddot{\mathcal{P}}_d$ over a time period of five years. Both are simply the sum of the monetary losses ($\ddot{\mathrm{M}}_{d,50}$) and the monetary equivalent of time losses ($\hat{\alpha}_d \cdot \widehat{\mathrm{T}}_{d,\mu}$) or the expenses ($\ddot{\mathrm{C}}_{d,50}$) and the time spent ($\hat{\alpha}_d \cdot \widehat{\mathrm{S}}_{d,\mu}$).

Table 7: Aggregate cost estimates per country

Country (d)		Cybercrime losses $\mathcal{L}$ (in €)				Protection costs $\mathcal{P}$ (in €)		
	$\hat{\alpha}_d$	$\ddot{\mathrm{M}}_{d,50}$	$\ddot{\mathrm{O}}_{d,50}$	$\hat{\alpha}_d \cdot \widehat{\mathrm{T}}_{d,\mu}$	$\ddot{\mathcal{L}}_d$	$\ddot{\mathrm{C}}_{d,50}$	$\hat{\alpha}_d \cdot \widehat{\mathrm{S}}_{d,\mu}$	$\ddot{\mathcal{P}}_d$
	(1)	(2)	(3)	(4)	(2)×(4)	(6)	(7)	(6)×(7)
DE	14.90	18.62	10.10	**29.88**	48.50	80.36	**251.55**	**331.91**
EE	4.09	10.16	5.99	4.01	14.17	14.45	27.47	41.93
IT	11.80	10.88	5.58	12.25	23.13	49.51	131.59	181.10
NL	15.36	12.74	7.35	18.77	31.51	75.44	188.42	263.86
PL	4.02	11.90	7.52	4.73	16.63	49.24	47.34	96.58
UK	12.99	**22.77**	**11.12**	27.1	**49.88**	**106.04**	121.69	227.74

Germany (DE), Estonia (EE), Italy (IT), Netherlands (NL), Poland (PL), United K. (UK);
Cybercrime losses of victims ($v = 1\,242$); Protection expenses of full sample ($n = 6\,242$)

In most countries cybercrime rather causes a loss of time than money. Accordingly, the monetary equivalent of time lost by the victims $\hat{\alpha}_d \cdot \widehat{\mathrm{T}}_{d,\mu}$ and spent for protection $\hat{\alpha}_d \cdot \widehat{\mathrm{S}}_{d,\mu}$ is generally larger than the respective monetary costs $\{\ddot{\mathrm{M}}_{d,50}, \ddot{\mathrm{C}}_{d,50}\}$. We find the biggest differences for protection costs in Germany, Italy, and the Netherlands, where the monetary equivalent of time spend on protection is at least 2.5 times bigger than the monetary expenses. Exceptions are monetary cybercrime losses in Estonia and Poland and protection expenses in Poland, which are slightly bigger than the time spent. These results are highly influenced by the choice of $\hat{\alpha}$, in this particular case the low hourly wages in Poland and Estonia.

Protection expenses $\ddot{\mathcal{P}}_d$ are higher than cybercrime losses $\ddot{\mathcal{L}}_d$ in all countries. This holds for monetary expenses and time spent. Estonians roughly spend three times more on protection than they lose to criminals. Citizens in the Netherlands spend more than eight times more. The differences become even larger, if compensation payments are considered. The out-of-pocket losses $\ddot{\mathrm{O}}_{d,50}$ in the Netherlands and Italy are more than ten times smaller than the expenses for protection.

Comparing different countries, we find that the highest cybercrime prevalence in Germany and the UK correlates with the highest cybercrime losses. Looking at the protection expenses, we can see that while Germans spend more time to protect themselves, consumers in the UK rather spend money on protection measures. The smallest cybercrime losses are found in Estonia and Poland. Polish consumers seem to pay for their security with high protection expenses, i. e., they only lose an average of 17 € directly, but spend more than 50 € on protection.

5.4 Indirect costs

In addition to direct losses $\mathcal{L}$ and protection expenses $\mathcal{P}$, cybercrime causes indirect losses $\mathcal{I}$. A large part of $\mathcal{I}$ are opportunity costs created by the reduced uptake of online services by concerned consumers. Anderson et al. [2013] estimate that indirect losses are much larger then $\mathcal{L}$ and $\mathcal{P}$. Their estimates are backed by technology acceptance literature, which finds that individual risk perception hinders technology acceptance and use on online services [Riek et al., 2016]. Featherman et al. [2010], for example, find that reducing perceived privacy risk, through corporate credibility, increases adoption in the context of online bill paying. In a more general approach Riek et al. [2016] show the negative impact of perceived risk of cybercrime on the use of online banking, online shopping, and online social networking, using structural equation modeling for a large pan-European sample.

The avoidance effect might be counter-intuitive given a generally increasing uptake of online services by consumers [e. g., ITU, 2015]. Still we find interesting support for different forms of avoidance through the reactions of the victims of cybercrime.[12] While overall less then 10 % reported to have stopped using online shopping after the incident, more then 20 % reported that they try to avoid it. Furthermore, a remarkable fraction of 65 % stated, that they only purchase from familiar or well-known websites. We find similar results for financial services. While only 9 % of the victims of IDT wrt. to online banking stopped using it, 19 % try to avoid it after the incidents. Furthermore, 29 % of IDT victims wrt. to PayPal closed their account after the incident. The results underline the importance of trust and credibility for online services already found by Featherman et al. [2010] and suggest indirect negative effects of cybercrimeon the online market, by driving customers to the big players. It highlights that avoidance research in the context of online services needs to be more focused to explain the negative impact of cybercrime in a growing online space.

[12] We asked all victims of cybercrime, how they reacted to the incident. Question: "Have you done any of the following, as a consequence of this incident?"

6 Discussion

Driven by the lack of reliable data regarding the economic impacts of consumer-facing cybercrime, we set out to develop a general instrument to conduct consumer surveys which enable robust cost estimation. We collected representative data, including an oversampling of victims, in six European countries for one instance of our instrument. Based on this data set we estimated the costs of cybercrime for the two cost factors, money and time, and two aggregates cost categories, losses and protection expenses. While our data collection took place in Europe, the theoretical and some empirical results can be generalized to other countries.

Limitations. Even though our estimates are based on representative data and oversampling of cybercrime victims, the results are not without limitations. For some types of cybercrime we only find a few incidents for which victims reported a monetary loss. Thus, monetary losses are not broken down by country and measured based on small sample sizes for some types of crime. Moreover, economic constraints on the questionnaire design may introduce bias to our estimates. As we do not collect data on multiple incidents of the same type of crime and screen victims of online shopping fraud with a proxy, we miss a few incidents and likely underestimate the prevalence of cybercrime. Conversely, our aggregated loss estimates likely overstate the losses because we impute the severest incident for unobserved loss amounts. A final important limitation concerns the generalization of results. We can and do not claim to provide exhaustive measurement of all costs of consumer-facing cybercrime because we exclude cybercrimes which are not mainly profit-oriented or part of the cybercriminal infrastructure. Following the cautious remarks in Anderson et al. [2013], we do not calculate a single cost estimate, but use our instrument to compare different cost categories.

Results. Regarding the methodology, our results confirm the benefit of using a two-part model. The model separates the probability of incurring a loss from the distribution of the losses for each victim. It has proven to be particularly helpful to understand victims losses because even many *severest* incidents do not lead to a monetary loss. Our analysis confirms that long tail distributions, in particular the log normal distribution, should be used to model costs of consumer-facing cybercrime. Our theoretical estimates are consistently smaller than the sample mean, supporting earlier proposals that reporting the mean loss over all incidents likely overestimates the costs of cybercrime. The median is a more reliable ad-hoc measure than the mean.

We estimate the costs of seven different types of cybercrime. We find the smallest losses, including money and time, for incidents related to online shopping. The highest initial monetary losses are found for incidents of identity theft related to financial services or online payments. However, the

victims likely receive financial compensation from their provider, reducing the remaining out-of-pocket losses considerably. Interestingly, we do not find evidence that compensated victims lose more time than those who do not receive compensation. While this situation seems acceptable for individual victims, service providers need to socialize the costs by increasing prices for services. This way, all consumers feel the burden of cybercrime losses like an indirect tax. Scams and extortion, which do not include a third party, turned out to be most time consuming. The relatively high prevalence and high monetary loss estimates indicate that scams have the severest impact on citizens. While empirical findings are based on data for Europe, we conjecture that the underlying effects also hold in other parts of the world.

Our aggregate cost estimates show that the main cost of cybercrime is lost *time*. Consumers are more likely to spend time on protection than money and rather lose time after an incident. Accordingly, the monetary equivalents of the time lost almost always exceed the monetary costs. Part of the reason is that monetary costs always go along with some loss of time, e. g., for configuring a purchased security product or investigating a loss. Consequently, clear instructions on effective protection measures and the provision of help and efficient processes to report incidents can reduce a large part of the costs.

We find that consumers behave generally protective because the aggregated protection costs are always bigger than the losses of the victims; in most countries more than fives times, even before compensation payments. The difference is further amplified by the fact that we estimate the losses based on the severest incident of each respondent. While one explanation is that consumers are risk averse, the difference can also be explained by the impacts of the cybercriminal infrastructure. Our data shows that malware infections are more prevalent than all other types of cybercrime combined. Even if the major part of infections does not lead to a more serious crime, consumers incur losses which they try to avoid by using preventive protection measures.

Outlook. A straightforward avenue for future research is to scale the survey up across countries and over time. If longer questionnaires are affordable, additional types of cybercrime can be added. Another possible direction is to ask for each incident independently in the case of multiple victimization. This removes the need for the severest case heuristic. Another particular suggestion, which follows from our discussion of indirect costs, is a study of the different facets of avoidance as a consequence of victimization.

While a comprehensive and longitudinal series of studies promises interesting insights, we also want to highlight the *costs of measuring the costs of cybercrime*, which might be another niche for empirical research. The data collection for this study has costed a high six-digit euro amount, which could only be financed in the context of an international effort. Moreover, we need to account for the time spent by more than 6000 respondents.

Acknowledgments

The authors thank Elena Lucica for her help during the creation of the survey instrument as well as Marie Vasek and Stefan Laube for their comments on the draft of the study. The paper draws on research performed as part of the E-CRIME project funded by the European Union's 7th Framework Programme under grant agreement number 607775.

References

ITU. Measuring the Information Society 2015. Technical report, International Telecommunication Union, Geneva, 2015. URL `www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2015.aspx`.

Julie J Ryan and Theresa I Jefferson. The use, misuse, and abuse of statistics in information security research. In *Proceedings of the 2003 ASEM National Conference*, 2003.

Brian Cashell, William D Jackson, Mark Jickling, and Baird Webel. The economic impact of cyber-attacks. Congressional Research Service, Library of Congress, 2004.

Dinei Florêncio and Cormac Herley. Sex, lies and cyber-crime surveys. In Bruce Schneier, editor, *Economics of Information Security and Privacy III*, pages 35–53. Springer, New York, 2013.

Paul Hyman. Cybercrime: It's serious, but exactly how serious? *Communications of the ACM*, 56(3):18–20, 2013.

Ross Anderson, Chris Barton, Rainer Böhme, Richard Clayton, Michel J. G. Eeten, Michael Levi, Tyler Moore, and Stefan Savage. Measuring the cost of cybercrime. In Rainer Böhme, editor, *Economics of Information Security and Privacy*, pages 265–300. Springer Berlin, Heidelberg, 2013.

Budi Arief, Mohd Azeem Bin Adzmi, and Thomas Gross. Understanding cybercrime from its stakeholders' perspectives: Part 1–attackers. *IEEE Security & Privacy*, (1):71–76, 2015.

Orin S Kerr. Cybercrime's scope: Interpreting 'access' and 'authorization' in computer misuse statutes. *NYU Law Review*, 78(5):1596–1668, 2003.

Huseyin Cavusoglu, Birendra Mishra, and Srinivasan Raghunathan. The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. *International Journal of Electronic Commerce*, 9(1):70–104, 2004.

Mike McGuire and Samantha Dowling. Cyber crime: A review of the evidence. Technical report, UK Home Office, 2013.

European Commission. Special Eurobarometer 423 Cyber security. Wave EB82.2, 2015.

Erika Harrell. Victims of identity theft, 2014. Technical report, Bureau of Justice Statistics (BJS) and US Deptarment of Justice and Office of Justice Programs of the United States of America, 2015.

Johannes Rieckmann and Martina Kraus. Tatort internet: Kriminalität verursacht bürgern schäden in milliardenhöhe. *DIW Wochenbericht*, 82: 295–301, 2015.

Leyla Bilge, Sevil Sen, Davide Balzarotti, Engin Kirda, and Christopher Kruegel. Exposure: A passive DNS analysis service to detect and report malicious domains. *ACM Transactions on Information System Security*, 16(4):14:1–14:28, 2014.

Chris Kanich, Christian Kreibich, Kirill Levchenko, Brandon Enright, Geoffrey M Voelker, Vern Paxson, and Stefan Savage. Spamalytics: An empirical analysis of spam marketing conversion. In *Proceedings of the 15th ACM Conference on Computer and Communications Security*, pages 3–14, 2008.

Cormac Herley and Dinei Florêncio. A profitless endeavor: Phishing as tragedy of the commons. In *Proceedings of the 2008 Workshop on New Security Paradigms*, pages 59–70, New York, NY, USA, 2008.

Kirill Levchenko, Andreas Pitsillidis, Neha Chachra, Brandon Enright, Márk Félegyházi, Chris Grier, Tristan Halvorson, Chris Kanich, Christian Kreibich, Liu, Damon McCoy, Nicholas Weaver, Vern Paxson, Geoffrey M Voelker, and Stefan Savage. Click trajectories: End-to-end analysis of the spam value chain. In *IEEE Symposium on Security and Privacy (SP)*, pages 431–446, 2011.

Damon McCoy, Andreas Pitsillidis, Jordan Grant, Nicholas Weaver, Christian Kreibich, Brian Krebs, Geoffrey Voelker, Stefan Savage, and Kirill Levchenko. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *USENIX Security 12*, pages 1–16, 2012.

Jason Franklin, Adrian Perrig, Vern Paxson, and Stefan Savage. 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*, CCS '07, pages 375–388, New York, NY, USA, 2007. ACM.

Kurt Thomas, Damon McCoy, Chris Grier, Alek Kolcz, and Vern Paxson. Trafficking fraudulent accounts: The role of the underground market in twitter spam and abuse. In *USENIX Security 13*, pages 195–210, 2013.

Computer Security Institute (CSI) and Federal Bureau of Investigation (FBI). Computer Crime and Security Survey. htpp://www.gocsi.com, 1996.

Center for Strategic and International Studies (CSIS). Net losses: estimating the global cost of cybercrime: Economic impact of cybercrime II. Technical report, McAfee, 2014. URL www.mcafee.com/mx/resources/reports/rp-economic-impact-cybercrime2.pdf.

Deloitte. Irish information security and cybercrime survey. Technical report, Deloitte, 2013. URL www2.deloitte.com/content/dam/Deloitte/ie/Documents/Risk/cybercrime_survey_risk_2013_deloitte_ireland.pdf.

Krista Jansson and Great Britain Home Office. *British Crime Survey – Measuring crime for 25 years*. Home Office, 2007.

Detica, Office of Cyber Security, and Information Assurance. The cost of cyber crime. Technical report, 2011. URL www.cabinetoffice.gov.uk/resource-library/cost-of-cyber-crime.

Antoine Frachot, Olivier Moudoulaud, and Thierry Roncalli. Loss distribution approach in practice. pages 527–554. Risk Books, London, 2004.

Kabir Dutta and Jason Perry. A tale of tails: an empirical analysis of loss distribution models for estimating operational risk capital. *FRB of Boston Working Paper*, 2006.

PwC. Global State of Information Security survey. Technical report, PricewaterhouseCoopers, 2015. URL www.pwc.com/gx/en/issues/cyber-security/information-security-survey.

Kaspersky Lab. Global IT Security Risks Survey. Technical report, 2015. URL www.media.kaspersky.com/en/business-security/it-security-risks-survey-2015.pdf.

Ponemon Institute. 2015 Cost of Cyber Crime Study: Global. Technical report, 2015. URL www8.hp.com/us/en/software-solutions/ponemon-cyber-security-report/.

Federation of small businesses. Cyber security and fraud: The impact on small businesses, 2013. URL www.fsb.org.uk/policy/assets/fsb_cyber_security_and_fraud_paper_final.pdf.

Erika Harrell. Victims of identity theft, 2012. Technical report, Bureau of Justice Statistics (BJS) and US Deptarment of Justice and Office of Justice Programs of the United States of America, 2012.

European Commission. Special Eurobarometer 390 Cyber security. Wave EB77.2, 2012.

European Commission. Special Eurobarometer 404 Cyber security. Wave EB79.4, 2013.

Julio Hernandez-Castro and Eerke Boiten. Cybercrime prevalence and impact in the UK. *Computer Fraud & Security*, 2014(2):5–8, 2014.

David Modic and Ross Anderson. It's all over but the crying: The emotional and financial impact of internet fraud. *IEEE Security & Privacy*, 13(5): 99–103, 2015.

Yongyi Min and Alan Agresti. Modeling nonnegative data with clumping at zero: a survey. *Journal of Iranian Statistical Society*, 1(1):7–33, 2002.

John Aitchison. On the distribution of a positive random variable having a discrete probability mass at the origin. *Journal of the American Statistical Association*, 50(271):901–908, 1955.

John G Cragg. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica: Journal of the Econometric Society*, pages 829–844, 1971.

Naihua Duan, Willard G Manning, Carl N Morris, and Joseph P Newhouse. A comparison of alternative models for the demand for medical care. *Journal of business & economic statistics*, 1(2):115–126, 1983.

Zhou Xiao-Hua and Wanzhu Tu. Comparison of several independent population means when their samples contain log-normal and possibly zero observations. *Biometrics*, 55(2):645–651, 1999.

Trendmicro. New crypto-ransomware locky uses malicious word macros, Feb. 2016. URL `www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/new-crypto-ransomware-locky-uses-word-macros`.

Eurostat. Structure of Earnings Survey (SES). Technical report, Eurostat, the statistical office of the European Union, 2010.

Markus Riek, Rainer Böhme, and Tyler Moore. Measuring the influence of perceived cybercrime risk on online service avoidance. *IEEE Transactions on Dependable and Secure Computing*, 13(2):261–273, 2016.

Mauricio S Featherman, Anthony D Miyazaki, and David E Sprott. Reducing online privacy risk to facilitate e-service adoption: the influence of perceived ease of use and corporate credibility. *Journal of Service Marketing*, 24(3):219–229, 2010.

APPENDIX

6.1 Description of variables

Table 8 shows a set of examples of variables we use to model the costs of consumer-facing cybercrime. It provides a short description for each variable. $\hat{\alpha}$ is the only external data source not collected in the survey. It is defined as the median of gross hourly earnings for each country and measured using a the Eurostat indicator `earn_ses_hourly` [Eurostat, 2010]. The remaining variables are measured using the data from the survey. Note that the list of variables in Table 8 is not exhaustive, but provides an example for each type.

Table 8: Description of variables

Variable	Description	Definition
c	Type of cybercrime	
d	Country	
v	Number of cybercrime victims	
n	Number of all respondents	
$\hat{\alpha}$	Time conversion factor: Gross hourly earnings	
Conditions		
$\widehat{\mathrm{P}}_{c,\mu}$	Empirical prob. of victimization	
$\widehat{\mathrm{Q}}_{c,m,\mu}$	Empirical prob. of monetary loss per type of crime	Eq.: 10
$\widehat{\mathrm{Q}}_{d,c,\mu}$	Empirical prob. of expenses per country	
Conditional estimates		
$\widehat{\mathrm{Z}}_{c,m,\mu}$	Empirical mean of monetary losses per type of crime	
$\widehat{\mathrm{Z}}_{c,m,50}$	Empirical median of monetary losses per type of crime	
$\vdots$		
$\widetilde{\mathrm{Z}}_{c,m,\mu}$	Theoretical mean of monetary losses per type of crime	
$\widetilde{\mathrm{Z}}_{c,m,50}$	Theoretical median of monetary losses per type of crime	
$\vdots$		
Unconditional estimates (combined indicators)		
$\ddot{\mathrm{M}}_{c,\mu}$	Initial monetary loss per type of crime: Expected value loss	Eq.: 11
$\ddot{\mathrm{M}}_{c,*}$	Initial monetary loss per type of crime: Adjusted median loss	Eq.: 12
$\ddot{\mathrm{M}}_{c,50}$	Initial monetary loss per type of crime: Harmonized loss	Eq.: 13
$\widehat{\mathrm{T}}_{c,\mu}$	Time losses per type of crime: scaled mean	
$\vdots$		
$\ddot{\mathrm{C}}_{d,50}$	Protection expenses per country: Harmonized loss indicator	
$\vdots$		
$\widehat{\mathrm{S}}_{d,\mu}$	Time spent per country: scaled mean	
Aggregated estimates		
$\ddot{\mathcal{P}}_d$	Aggregated protection costs per country	Eq.: 6
$\ddot{\mathcal{L}}_c$	Aggregated cybercrime losses per type of crime	Eq.: 8
$\ddot{\mathcal{L}}_d$	Aggregated cybercrime losses per country	Eq.: 9

6.2 Proxy for online shopping fraud

Initial descriptive statistics suggest that the wording for online shopping fraud (see Table 2) could have been too general, as the victimization rate for this type of cybercrime is significantly higher compared to all other types of cybercrime and other surveys. A possible explanation for this could be the fact that respondents who have experienced inconveniences when shopping online, which were not necessarily caused by fraud (e.g. products not being of the quality they had expected, or delivery problems), were also classified as cybercrimevictims. To mitigate the problem we developed a proxy logic to approximate a more realistic extent of online shopping fraud victimization. We use detailed questions on the criminal case to identify respondents that have been victims of online shopping fraud among the ones that were originally identified. The following logic identifies a respondent to be a victim of online shopping fraud:

1. The respondent is only victim of online shopping fraud or the severest incident reported by the respondent is online shopping fraud.
2. The respondent lost money due to the incident.
3. The respondent was *not* able to recover all losses.

Using the proxy variable significantly reduces the number of cases of online shopping fraud from 2 052 respondents who answered *Yes* to the original question to 551 respondents who meet all three conditions of the proxy logic. Table 9 shows the relative effects per country.

Table 9: Online shopping fraud victims per country. Original and proxy definition.

	DE	EE	IT	NL	PL	UK
Original	36.90 %	30.49 %	17.28 %	29.46 %	36.02 %	43.34 %
EB 423	13 %	13 %	11 %	16 %	19 %	16 %
Proxy	8.36 %	9.10 %	4.98 %	10.23 %	9.67 %	9.01 %

United Kingdom (UK), Netherlands (NL), Estonia (EE), Germany (DE), Poland (PL), Italy (IT)
Eurobarometer on Cyber Security 2014 (EB 423) [European Commission, 2015];

The proxy logic and the number of resulting cases provide confidence that the selected respondents are indeed victims of online shopping fraud. As our victimization rate for online shopping fraud is still smaller than in the comparable Eurobarometer survey for all six countries [European Commission, 2015], we believe that our proxy is still on the conservative side and rather underestimates the real victimization rate.

6.3 Demographics of the sample

Table 10 shows the demographics of the full sample and the victim subgroup for each of the surveyed countries.

Table 10: Demographics

Variable	DE		EE		IT		NL		PL		UK	
	v	n	v	n	v	n	v	n	v	n	v	n
Gender												
Male	0.53	0.58	0.47	0.58	0.53	0.57	0.51	0.47	0.46	0.48	0.5	0.49
Female	0.47	0.42	0.53	0.42	0.47	0.43	0.49	0.53	0.54	0.52	0.5	0.51
Age												
18-20	0.05	0.06	0.07	0.04	0.06	0.09	0.05	0.05	0.07	0.04	0.07	0.06
21-30	0.2	0.18	0.23	0.3	0.21	0.24	0.18	0.25	0.28	0.33	0.2	0.2
31-40	0.19	0.25	0.19	0.24	0.23	0.16	0.17	0.22	0.26	0.33	0.17	0.19
41-50	0.23	0.21	0.22	0.24	0.25	0.22	0.21	0.22	0.19	0.15	0.18	0.19
51-60	0.19	0.21	0.17	0.14	0.16	0.18	0.18	0.12	0.14	0.1	0.18	0.18
61-70	0.07	0.06	0.09	0.03	0.08	0.1	0.13	0.09	0.05	0.03	0.12	0.09
70+	0.05	0.04	0.03	0.01	0.01	0.02	0.08	0.04	0.01	0.01	0.06	0.05
Area of living												
Big city	0.19	0.2	0.54	0.59	0.16	0.16	0.27	0.32	0.28	0.27	0.13	0.12
Suburbs	0.17	0.14	0.05	0.08	0.08	0.07	0.18	0.18	0.11	0.1	0.19	0.22
Town	0.34	0.36	0.22	0.19	0.35	0.34	0.24	0.25	0.31	0.39	0.39	0.38
Village	0.25	0.23	0.16	0.12	0.37	0.36	0.27	0.22	0.23	0.2	0.2	0.18
Countrys.	0.04	0.06	0.03	0.01	0.04	0.06	0.03	0.02	0.07	0.04	0.06	0.08

6.4 Impact of victimization

Table 11 illustrates the effects of the victimization for each type of cybercrime. The majority of the respondents lost time after the incident, while a much smaller proportion lost money. Others also reported personal and professional problems as well as other impacts.

Table 11: Impact of cybercrime victimization

Cybercrime	Lost time	Lost money	Personal problems	Professional problems	Other problems
IDT wrt. o. banking	95 %	33 %	23 %	7 %	9 %
IDT wrt. bank cards	96 %	35 %	25 %	5 %	17 %
IDT wrt. PayPal	95 %	24 %	26 %	5 %	17 %
IDT wrt. o. shopping	96 %	17 %	14 %	6 %	14 %
Online shopping fraud	88 %	91 %	7 %	2 %	7 %
Extortion	93 %	13 %	14 %	9 %	10 %
Scams	92 %	45 %	18 %	5 %	13 %

Identity theft (IDT); Based on the severest cases ($v = 1\,242$)

6.5 Cybercrime victims with monetary losses

Table 12 shows the victims (only severest incident) with monetary losses broken down by type of cybercrime and country. Accordingly, Table 12 contains the reported point estimates that were used to estimate the parameters of the loss distributions.

Table 12: Cybercrime incidents with monetary losses by country and type

		Incidents with financial losses						
Cybercrime	$\hat{Q}_{c,m}$	UK	NL	EE	DE	PL	IT	Total
IDT wrt. o. banking	1.6 %	1	6	1	2	2	10	22
IDT wrt. bank cards	2.6 %	6	18	18	12	2	13	69
IDT wrt. PayPal	1.2 %	2	2	1	2	0	5	12
IDT wrt. o. shopping	2.2 %	3	1	5	1	2	5	17
Online shopping fraud	8.6 %	73	95	48	102	92	78	488
Extortion	2.1 %	3	2	1	1	4	3	14
Scams	3.2 %	18	18	14	11	13	16	90
Total		106	142	88	131	115	130	712

United Kingdom (UK), Netherlands (NL), Estonia (EE), Germany (DE), Poland (PL), Italy (IT)

6.6 Distribution of initial monetary losses

Table 13 shows the parameter estimate for the initial loss distributions $g_{c,\hat{\theta}}$ along with the number of point estimates n and the relative goodness-of-fit indicators AIC and BIC for each type of cybercrime. Figure 7 – Figure 12 each show the empirical and theoretical distribution of expenses for protection measures for one country. Note, that for the rate parameter in the gamma distribution $\hat{\theta}_2$ left bounds are fixed to 0.005, to avoid unsuccessful termination of the maximum likelihood estimation. Accordingly, these bounds are rounded to 0 in Table 13.

Table 13: Parameter estimates ($\hat{\theta}$) for initial cybercrime losses

		Log normal			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
IDT wrt. OS	17	4.94 (.434)	1.53 (.307)	172	174
IDT wrt. OB	22	6.14 (.485)	1.85 (.343)	242	244
IDT wrt. BC	69	5.80 (.269)	1.81 (.190)	708	712
IDT wrt. PayPal	12	6.19 (.668)	2.10 (.473)	169	170
Extortion	14	4.31 (.574)	1.84 (.406)	135	136
Scams	90	5.29 (.207)	1.66 (.146)	928	933
OS fraud	488	3.98 (.600)	1.34 (.420)	5685	5694
		Gamma			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
IDT wrt. OS	17	1.14 (.390)	0 (.2)	191	192
IDT wrt. OB	22	2.82 (.871)	0 (.2)	436	438
IDT wrt. BC	69	2.12 (.379)	0 (.1)	961	965
IDT wrt. PayPal	12	2.92 (1.940)	0 (.2)	288	289
Extortion	14	0.63 (.220)	0 (.1)	132	133
Scams	90	1.45 (.219)	0 (.1)	1358	1363
OS fraud	488	0.53 (.260)	0 ()	5981	5990
		Weibull			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
IDT wrt. OS	17	0.64 (.136)	310.39 (146.177)	175	176
IDT wrt. OB	22	0.56 (.109)	1170.8 (576.457)	244	246
IDT wrt. BC	69	0.62 (.710)	796.27 (201.141)	709	713
IDT wrt. PayPal	12	0.58 (.146)	1317.02 (762.114)	168	169
Extortion	14	0.75 (.189)	167.82 (73.22)	132	133
Scams	90	0.55 (.460)	468.42 (113.924)	947	952
OS fraud	488	0.65 (.190)	107.34 (7.817)	5855	5863
		Normal			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
IDT wrt. OS	17	196.77 (692.43)	139.13 (201.12)	201	203
IDT wrt. OB	22	1087.52 (4133.86)	767.39 (287.75)	288	290
IDT wrt. BC	69	275.96 (1853.25)	195.19 (810.1)	810	814
IDT wrt. PayPal	12	1017.27 (3199.6)	719.98 (191.31)	191	192
Extortion	14	72.35 (232.39)	51.17 (145.68)	146	147
Scams	90	494.3 (3943.81)	348.32 (1246.53)	1247	1252
OS fraud	488	29.24 (653.23)	20.68 (7891.26)	7891	7900

Based on cybercrime victims with point estimates of monetary losses ($v' = 712$)

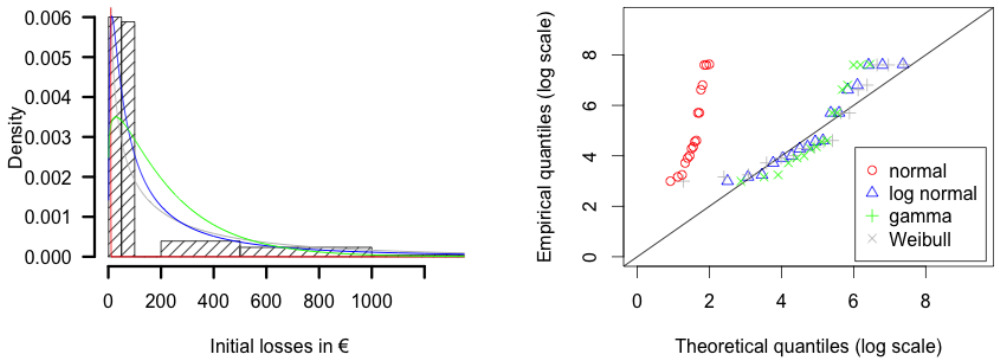

Figure 7: Initial losses from **IDT wrt. online shopping**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

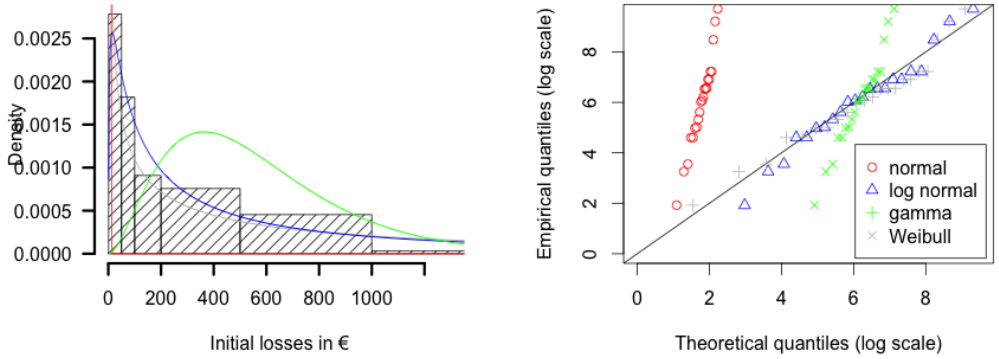

Figure 8: Initial losses from **IDT wrt. online banking**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

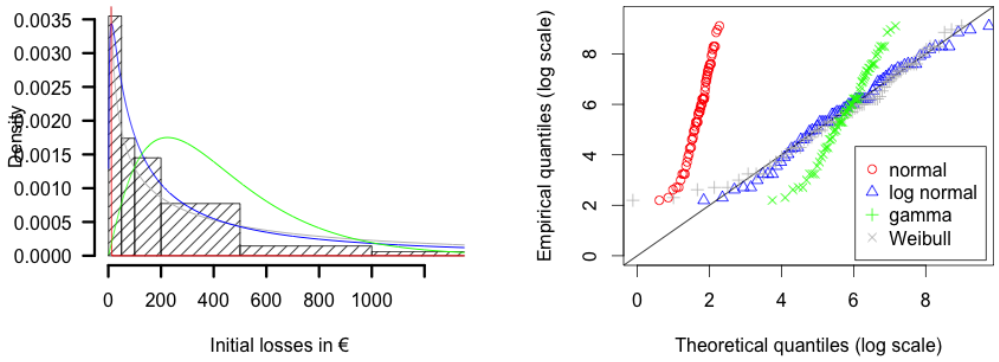

Figure 9: Initial losses from **IDT wrt. bank cards**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

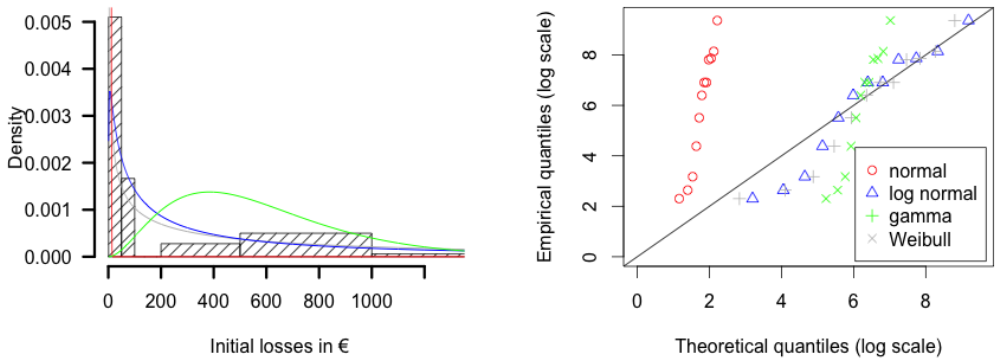

Figure 10: Initial losses from **IDT wrt. PayPal**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

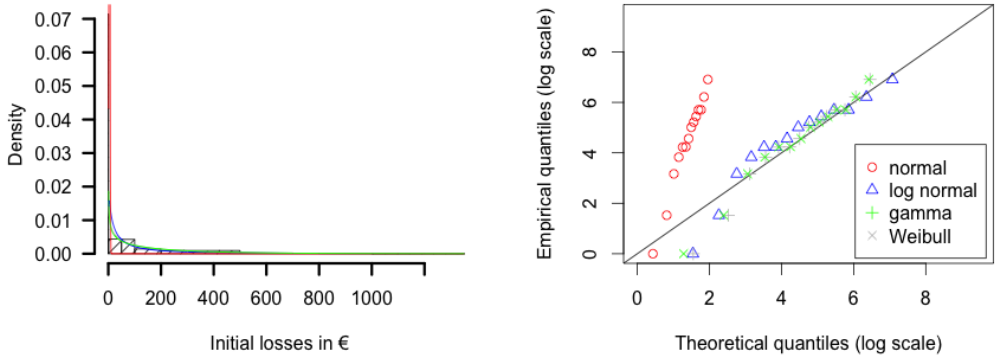

Figure 11: Initial losses from **extortion**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

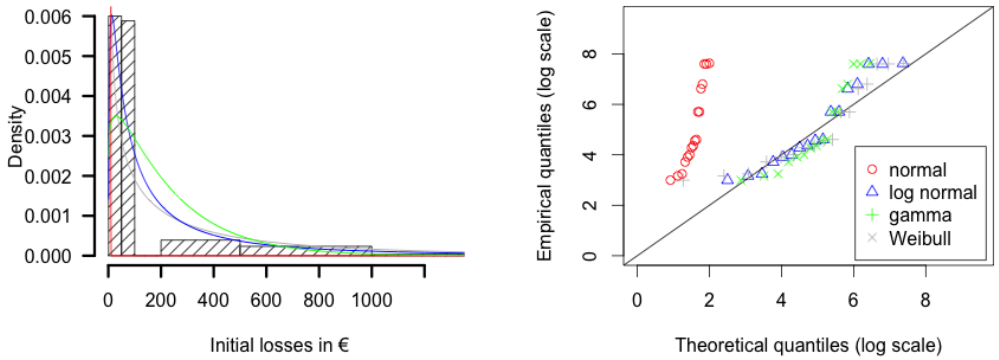

Figure 12: Initial losses from **online shopping fraud**; Left: Histogram and candidate loss distributions, Right: Q–Q plot of candidate loss distributions on log scale

6.7 Distribution of out-of-pocket losses

Table 13 shows the parameter estimate for the distributions $g_{c,\hat{\theta}}$ of the out-of-pocket losses along with the number of point estimates n and the relative goodness-of-fit indicators AIC and BIC for each type of cybercrime.

Table 14: Parameter estimates ($\hat{\theta}$) for out-of-pocket cybercrime losses

		Lognormal			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
OS fraud	481	3.73 (.620)	1.38 (.440)	5408	5416
IDT wrt. OS	10	4.53 (.601)	1.63 (.426)	98	99
IDT wrt. OB	11	5.81 (.734)	1.96 (.519)	117	117
IDT wrt. BC	36	5.14 (.370)	1.8 (.261)	343	346
IDT wrt. PayPal	6	5.31 (1.162)	2.49 (.822)	74	74
Extortion	13	4.16 (.567)	1.78 (.401)	125	126
Scams	82	5.22 (.217)	1.66 (.154)	833	837
		Gamma			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
OS fraud	481	0.54 (.280)	0 ()	5666	5674
IDT wrt. OS	10	0.89 (.393)	0 (.3)	106	106
IDT wrt. OB	11	2.15 (.965)	0 (.2)	195	196
IDT wrt. BC	36	1.31 (.323)	0 (.1)	409	412
IDT wrt. PayPal	6	1.48 (.828)	0 (.3)	95	95
Extortion	13	0.69 (.258)	0 (.2)	122	123
Scams	82	1.38 (.218)	0 (.1)	1197	1202
		Weibull			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
OS fraud	481	0.66 (.200)	84.59 (6.150)	5549	5558
IDT wrt. OS	10	0.61 (.169)	216.62 (139.264)	100	100
IDT wrt. OB	11	0.54 (.148)	886.31 (657.330)	117	118
IDT wrt. BC	36	0.61 (.910)	407.68 (146.314)	344	347
IDT wrt. PayPal	6	0.55 (.214)	626.1 (555.700)	73	73
Extortion	13	0.81 (.215)	138.55 (56.827)	122	123
Scams	82	0.55 (.490)	433.43 (110.704)	850	855
		Normal			
Cybercrime	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
OS fraud	481	24.47 (543.77)	17.3 (7624.12)	7624	7632
IDT wrt. OS	10	202.93 (548.54)	143.62 (116.90)	117	118
IDT wrt. OB	11	1250.5 (3337.24)	883.01 (139.72)	140	141
IDT wrt. BC	36	316.05 (1537.95)	223.44 (419.80)	419	422
IDT wrt. PayPal	6	503 (1079.94)	355.6 (81.49)	81	81
Extortion	13	48.16 (151.40)	34.07 (130.58)	131	132
Scams	82	512.42 (3910.11)	362.33 (1128.85)	1129	1134

Based on cybercrime victims with point estimates of monetary losses ($v'' = 639$)

6.8 Distribution of protection expenses

Table 15 shows the parameter estimates for the expenses for protection measures $g_{\hat{\theta}}$ along with the number of point estimates n and the relative goodness-of-fit indicators AIC and BIC for each country. Figure 13 – Figure 18 each show the empirical and theoretical distribution of expenses for protection measures for one country.

Table 15: Parameter estimates for the protection expenses in each country

		Lognormal			
Country	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
DE	597	3.98 (.600)	1.34 (.420)	7457	7466
EE	173	4.94 (.434)	1.53 (.307)	2020	2027
IT	452	6.14 (.485)	1.85 (.343)	5502	5510
NL	476	5.8 (.269)	1.81 (.190)	6033	6041
PL	628	6.19 (.668)	2.1 (.473)	7311	7320
UK	609	4.31 (.574)	1.84 (.406)	7862	7871
		Gamma			
Country	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
DE	597	0.53 (.260)	0 ()	7572	7581
EE	173	1.14 (.390)	0 (.2)	2006	2012
IT	452	2.82 (.871)	0 (.2)	5580	5589
NL	476	2.12 (.379)	0 (.1)	6016	6024
PL	628	2.92 (1.94)	0 (.2)	7258	7266
UK	609	0.63 (.220)	0 (.1)	7928	7937
		Weibull			
Country	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
DE	597	0.65 (.190)	107.34 (7.817)	7617	7625
EE	173	0.64 (.136)	310.39 (146.177)	2009	2015
IT	452	0.56 (.109)	1170.8 (576.457)	5587	5595
NL	476	0.62 (.710)	796.27 (201.141)	6028	6037
PL	628	0.58 (.146)	1317.02 (762.114)	7275	7284
UK	609	0.75 (.189)	167.82 (73.220)	7960	7969
		Normal			
Country	n	$\hat{\theta_1}(sd)$	$\hat{\theta_2}(sd)$	AIC	BIC
DE	597	29.24 (653.23)	20.68 (7891.26)	8561	8570
EE	173	196.77 (692.43)	139.13 (201.12)	2196	2203
IT	452	1087.52 (4133.86)	767.39 (287.75)	6370	6378
NL	476	275.96 (1853.25)	195.19 (810.10)	6297	6305
PL	628	1017.27 (3199.6)	719.98 (191.31)	7915	7924
UK	609	72.35 (232.39)	51.17 (145.68)	8565	8574

United Kingdom (UK), Netherlands (NL), Estonia (EE), Germany (DE), Poland (PL), Italy (IT)
Based on the all consumers who spend money on protection software ($s' = 2\,935$)

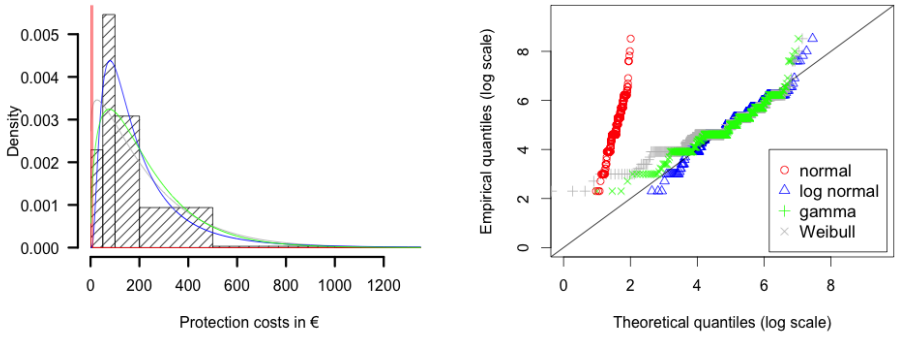

Figure 13: Protection expenses for **German** consumers; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

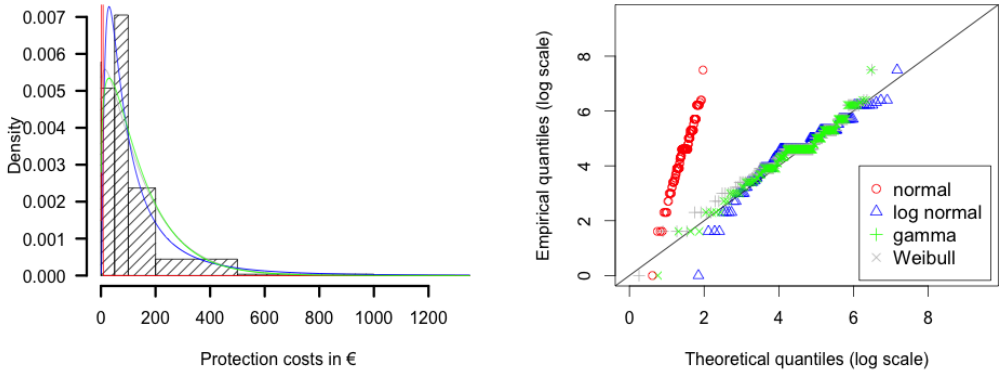

Figure 14: Protection expenses for **Estonian** consumers; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

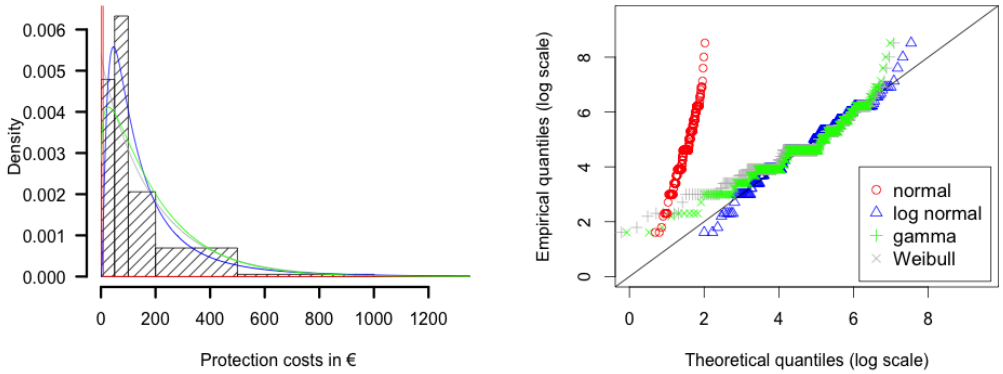

Figure 15: Protection expenses for **Italian** consumers; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

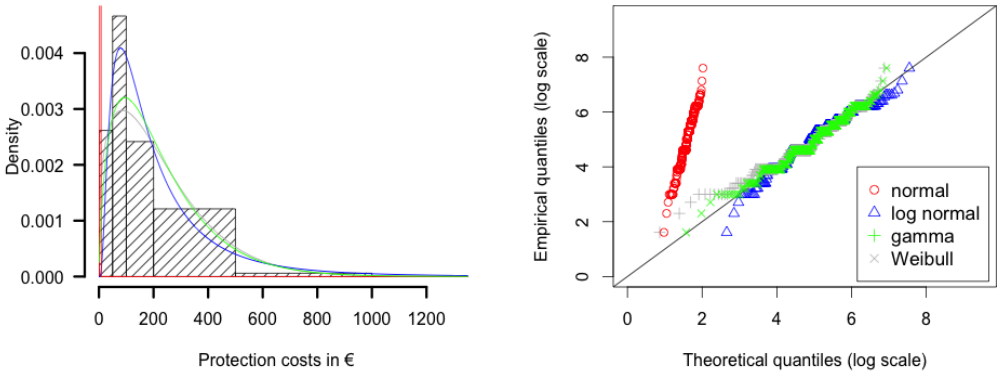

Figure 16: Protection expenses for **Dutch** consumers; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

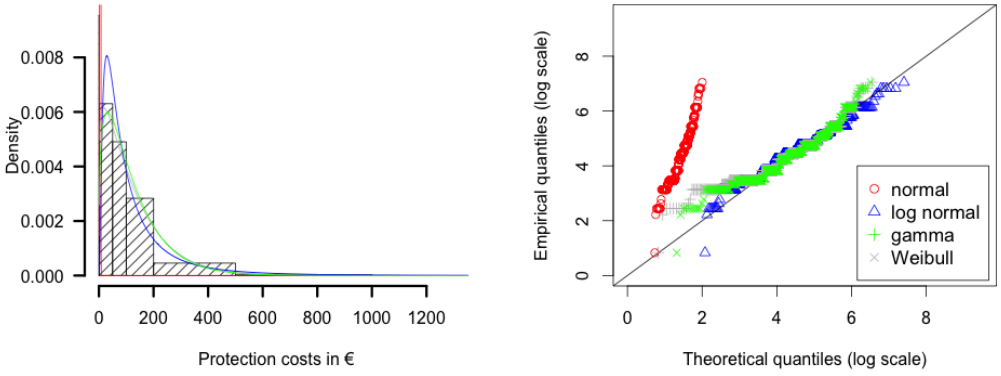

Figure 17: Protection expenses for **Polish** consumers; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

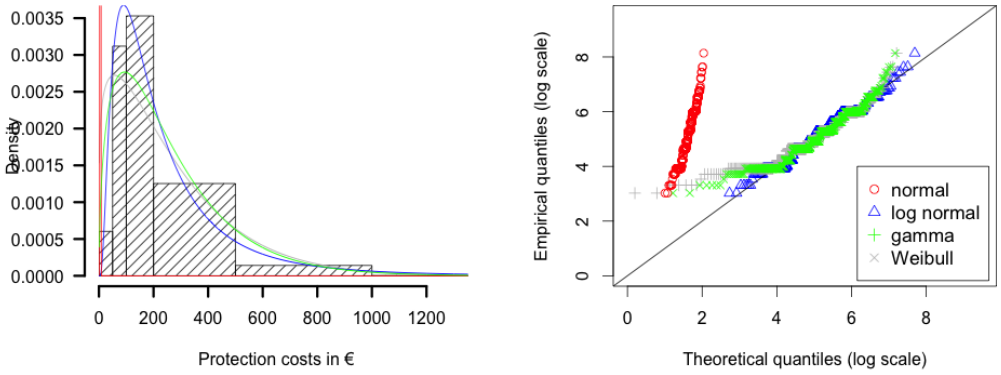

Figure 18: Protection expenses for consumers in the **UK**; Left: Histogram and candidate cost distributions, Right: Q–Q plot of candidate cost distributions on log scale

6.9 Reactions of cybercrime victims

Table 16 shows all reactions, reported by cybercrime victims after their severest incident concerning avoidance of online services. The reactions can be roughly distinguished into avoidance of payment services and online shopping. Note, that victims were able to report multiple reactions.

Table 16: Reported reactions of victims after the severest incident

	PayPal & online banking (OB)			Online shopping (OS)		
Cybercrime	Closed PayPal	Avoid OB	Stop to use OB	Avoid OS	Trusted Shops	Stop to use OS
IDT wrt. OB	9.30 %	18.60 %	9.30 %	11.36 %	51.16 %	9.52 %
IDT wrt. BC	7.58 %	18.94 %	9.16 %	24.43 %	65.65 %	10.69 %
IDT wrt. PayPal	29.27 %	14.63 %	7.14 %	19.51 %	63.41 %	4.88 %
IDT wrt. OS	17.39 %	18.84 %	8.57 %	23.19 %	69.57 %	11.43 %
OS fraud	4.74 %	12.75 %	5.83 %	22.26 %	59.56 %	7.65 %
Extortion	10.53 %	17.33 %	9.21 %	22.37 %	71.05 %	5.33 %
Scams	10.56 %	20.42 %	9.15 %	24.65 %	71.83 %	9.22 %
Average	12.77 %	17.36 %	8.34 %	21.11 %	64.61 %	8.39 %

Based on the severest incident for all victims ($v = 1\,242$); multiple answers possible