
Research paper

Examining the costs and causes of cyber incidents

Sasha Romanosky*

RAND Corporation, 1200 South Hayes St, Arlington, VA 22202, USA

*Corresponding author: E-mail: sromanos@rand.org.

Received 23 January 2016; revised 24 May 2016; accepted 20 June 2016

Abstract

In 2013, the US President signed an executive order designed to help secure the nation's critical infrastructure from cyberattacks. As part of that order, he directed the National Institute for Standards and Technology (NIST) to develop a framework that would become an authoritative source for information security best practices. Because adoption of the framework is voluntary, it faces the challenge of incentivizing firms to follow along. Will frameworks such as that proposed by NIST really induce firms to adopt better security controls? And if not, why? This research seeks to examine the composition and costs of cyber events, and attempts to address whether or not there exist incentives for firms to improve their security practices and reduce the risk of attack. Specifically, we examine a sample of over 12 000 cyber events that include data breaches, security incidents, privacy violations, and phishing crimes. First, we analyze the characteristics of these breaches (such as causes and types of information compromised). We then examine the breach and litigation rate, by industry, and identify the industries that incur the greatest costs from cyber events. We then compare these costs to bad debts and fraud within other industries. The findings suggest that public concerns regarding the increasing rates of breaches and legal actions may be excessive compared to the relatively modest financial impact to firms that suffer these events. Public concerns regarding the increasing rates of breaches and legal actions, conflict, however, with our findings that show a much smaller financial impact to firms that suffer these events. Specifically, we find that the cost of a typical cyber incident in our sample is less than \$200 000 (about the same as the firm's annual IT security budget), and that this represents only 0.4% of their estimated annual revenues.

Key words: cyber insurance; data breach; security incident; privacy violation; data breach litigation.

Introduction

Cyber threats have been described as “among the gravest national security dangers to the United States” (The White House 2015), and “an enormous and an exponentially growing threat” (Comey 2013). In 2013, the President signed an executive order designed to help secure the nation's critical infrastructure from cyber (aka computer-mediated) attacks. As part of that order, he directed NIST to develop a framework that would become an authoritative source for information security best practices (see <http://www.nist.gov/cyberframework/>). Of course, because adoption of the framework is voluntary, it faces the problem of incentives by firms to adopt [“To promote cybersecurity practices and develop these core capabilities, we are working with

critical infrastructure owners and operators to create a Cybersecurity Framework – a set of core practices to develop capabilities to manage cybersecurity risk. . . . While this effort is underway, work on how to incentivize companies to join a Program is also under consideration. While the set of core practices have been known for years, barriers to adoption exist, such as the challenge of clearly identifying the benefits of making certain cybersecurity investments” (Daniel 2013)]. Are firms incentivized to adopt better security controls? And if not, why? This research seeks to examine the composition and costs of cyber events, and attempts to address whether or not there exist incentives for firms to improve their security practices and reduce the risk of attack.

Data breaches, cyberattacks, and privacy violations have become commonplace. Yet despite academic literature and media stories

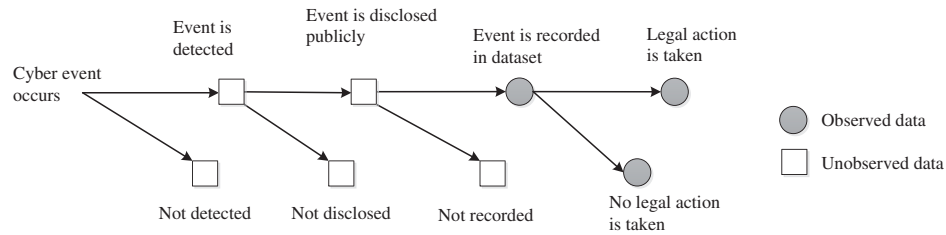


Figure 1. Data generating process.

concerning these events, there exist only a very limited body of research that examines these incidents in order to properly assess the risk and trends (see Edwards et al. 2015; Biener 2015). Therefore, using a unique dataset of over 12 000 cyber incidents recorded over the years 2004 and 2015, we conduct a thorough analysis of these incidents and examine the costs and composition of these events, by industry, and over time.

Throughout this research, we distinguish between four types of cyber events: data breaches (unauthorized disclosure of personal information), security incidents (malicious attacks directed at a company), privacy violations (alleged violation of consumer privacy), and phishing/skimming incidents (individual financial crimes). Of all cyber incidents from our dataset, we find that data breaches are by far the most common, dwarfing rates of all other cyber events. Beyond name and address, we find that credit card numbers and medical information were the most commonly compromised pieces of information. And incidents caused by malicious actions (as opposed to accidental or unintentional activities) have remained relatively constant at around 60% of all incidents. Further, of the almost 1700 resulting legal actions, over 50% continue to be private civil actions brought in federal courts, with only 17% being criminal actions.

In order to better understand the risk by industry, there are a number of potentially relevant metrics, each of which provide a useful, but singularly incomplete, insight. For example, we examine the following metrics for the events in our database: total number of incidents, incident rate, litigation rate, total cost, and cost per event (The incident rate represents the number of cyber incidents within a given industry, divided by the number of firms within that industry. Litigation rate is computed as the number of lawsuits within an industry, divided by the number of cyber events that occur within the industry.). While the Finance and Insurance industry suffers the greatest number of cyber events, Government agencies suffer the highest ‘incident’ rate. Further, Mining and Oil & Gas firms suffer the highest ‘litigation’ rate, while Management firms suffer the highest ‘cost per event’. Overall, when examining each of the metrics together, we find that the Retail, Information, Manufacturing, and Finance and Insurance industries consistently pose the greatest risk, while – contrary to common belief – Health Care and Education services pose some of the lowest risks.

Finally, while we estimate the total costs from cyber events at approximately \$8.5 billion annually. We find that the typical cost of a data breach is less than \$200 000, far lower than the millions of dollars often cited in surveys (e.g. Ponemon 2015). Moreover, we find that cyber incidents cost firms only a 0.4% of their annual revenues, much lower than retail shrinkage (1.3%), online fraud (0.9%), and overall rates of corruption, financial misstatements, and billing fraud (5%).

While in essence, this article represents a descriptive analysis of a single dataset (rather than causal inference), we believe that it does provide relevant and important findings. For example, by comparing observed cyber events with the total number of firms within an

industry, this research provides one of the first true estimates of risk, by industry type. Further, our use of cost data enables us to provide a unique and novel analysis of the scope and magnitude of cyber events, as a function of firm revenues, and other forms of loss, theft, and waste.

This analysis is expected to help inform three main stakeholders: private sector firms, insurance companies, and policy makers. Private sector firms will better understand the risks they face when collecting personal information and operating IT networks that are publicly accessible. Insurance companies can better estimate the risks of their insureds in order to foster a healthy cyber insurance market, and policy makers will better understand the context and impact of these cyber events across industry, and time.

Data

Data source and data generating process

The analysis presented in this article is derived from multiple datasets. First, we acquired a dataset of cyber incidents from Advisen, a US-based, for-profit organization that collects, integrates and resells loss and incident data to the commercial insurance industry regarding many different forms of corporate loss (global supply chain risk, employment practices, directors and officers liability, etc. In total, their loss database exceeds 300 000 observations; <http://www.advisenltd.com> for more information), including cyber events. In order to compile the cyber loss database, Advisen employs a dedicated team of analysts who use a comprehensive set of search strategies in order to find and classify publicly available information regarding cyber events. Specifically, Advisen collects news stories from dozens of local and national online news websites, newsfeeds, and vendor partners. It also sources information from specialized legal information services, as well as multiple online data breach clearinghouses. In addition, it collects information from state and federal governments and agencies both from publicly available websites, as well as employing the freedom of information act (FOIA) and their state analogs. While other publicly available data include 4000–5000 observations, Advisen manages over 15 000 observations, and continues to grow. Therefore, we believe that these efforts have created one of the most comprehensive datasets of cyber events available.

While the search strategies used to collect these data (and any data, for that matter) are thorough, they are not without limitations, as shown in Figure 1.

First, the methods used to collect the incidents are limited to publicly available data sources. For example, of all cyber events, only some will be detected (either by the organization suffering the event, or by a third party who, in turn notifies the firm) (In some cases, it may be law enforcement that first learns of, and notifies the firm. In other cases, it may be a security forensics firm, a security reporter, a credit card processor, or a consumer. The means by which the firm is notified, however, is not relevant for this analysis). An

important factor to consider is that we do not expect all cyber events to be detected equally. For example, most states have adopted laws that require organizations to notify individuals when their information is compromised (Romanosky 2011). However, there are no similar laws for privacy violations or security incidents. This alone suggests a bias in available data toward data breaches, relative to other cyber events.

Conditional on detection, the event may be disclosed to the public. Certainly, firms have an incentive to at least identify attacks, successful or otherwise, against their corporate systems and network if only to stop the attack and reduce any further losses. Our data collection, therefore, is strongly reliant on the effectiveness of information security detection systems, consumer awareness (e.g. in regard to privacy violations), and law enforcement (e.g. in regard to detecting phishing and skimming crimes). In some cases, public disclosure may also occur in the financial statements of publicly traded firms.

Of the publicly disclosed events, some (we hope most) will be captured by the data collectors and included in our data. Our data do not include events that have not been disclosed to the public, nor do they include events which have been missed by an analyst (There are a number of reasons why a breach or cyber event would not be disclosed to the public. First, while most states have breach disclosure laws, many provide reporting exceptions such as a breach that affect only a few individuals, when the stolen information is encrypted, etc.). For example, very small data breaches affecting only a few individuals would likely not be captured by this (or most any) search strategy. That being said, incidents involving many thousands, millions, or tens of millions of individuals would very likely be captured. Cyber events that are detected by firms, but are willfully ignored may also not be included in these data (unless they were eventually discovered by a third party).

However, if indeed the national policy debate and legal doctrine are most influenced by these medium and large incidents, then the search strategies employed in the creation of these data (even if potentially biased toward larger incidents) would very likely produce a representative sample of the population of reported cyber events, and those in which we are most interested, because they would be considered to be the most damaging to firms and consumers. Nevertheless, we recognize that the inferences made within this article apply strictly to the data being evaluated.

Additional data sources used in this article include financial information (e.g. revenue, employees, industry, etc.), industry data collected from the US Census, and numerous industry reports regarding fraud, waste, and abuse.

Types of cyber incidents

Advisen's raw cyber event dataset distinguishes between 11 separate types of cyber incidents – too large for practical analysis. Given that many events shared similar fundamental characteristics, some categories were aggregated based on a combination of approaches: the type of event; industry convention as used in previous academic research and security reports; and conversations with security and privacy experts. Therefore, we have organized the data according to the following categories which are meant to be exhaustive and mutually exclusive (Reasonable people may conceive of alternative taxonomies. However, for the purpose of this research, we employ the categories described within).

Data breach

The unintentional disclosure of personally identifiable information (PII) stemming from loss or theft of digital or printed information.

For example, the theft of laptop or desktop computers containing personal information of employees or customers of a firm, caused either by a hacker or malicious employee. This category also includes the improper disposal or disclosure of personal information (i.e. to a dumpster or website). The unauthorized disclosure of personal information can be used to commit identity theft, and other sorts of tax, medical and financial fraud and theft.

Security incident

An incident involving the compromise or disruption of corporate IT systems (computers or networks) or its intellectual property. For example, a denial of service (DoS) attack, the theft of intellectual property, the malicious infiltration (hack) and subsequent cyber extortion of corporate information, or a disruption of business services.

Privacy violation

The unauthorized collection, use or sharing of personal information. For example, unauthorized collection from cell phones, GPS devices, cookies, web tracking or physical surveillance. Allegations of violations of information protection statutes such as Drivers Privacy Protection Act (DPPA), Video Privacy Protection Act (VPPA), Telephone Consumer Protection Act (TCPA), Children's Online Privacy Protection Act (COPPA), Do-Not-Call, Song-Beverly Act, and the Privacy Act. Also includes unsolicited communication from spam emails, other mass marketing communication (robocalling, texts, emails), or debt collection. The unauthorized collection of personal information can be used to track individuals online with or without their consent, or harass them with unsolicited communication.

The first two categories are generally differentiated from the third in that the first two relate to incidents 'suffered by' the firm (i.e. PII stolen from the firm, or the firm suffering a compromise of business operations because of a hack), while the third category relates to events 'caused by' the firm (e.g. the firm improperly collecting or selling personal information).

Phishing/Skimming

The final category relates to instances of individuals committing particular kinds of computer or electronic crimes directly against other individuals or firms. For example, these crimes would include phishing attacks (wherein criminals seek to harvest account information from users), identity theft (wherein criminals use another person's information for financial gain), or skimming attacks (where criminals install, e.g., a hardware device over ATM machines in order to copy bank account and bank PIN numbers).

Descriptive analysis of cyber events

Figure 2 shows the absolute number of incidents across the four groups with data breaches displayed in the left panel, and all others (privacy violations, security incidents, and phishing/skimming incidents) displayed in the right panel. All figures reflect data over a 10-year period from 2005 to 2014. Note the difference in scale between the left (0–1500) and right (0–250) panels.

These data show how some events have been increasing steadily since 2005, though at a decreasing rate. However, the count of reported data breaches is in some cases an order of magnitude larger than other incidents. For example, data breach reports have seen a four-fold increase from just over 200 in 2005 to over 1200 in 2014, while privacy violations have seen only a modest increase from dozens in 2005, to around 150 by 2014. The rise in data breaches may be, in part, due to some states adopting breach disclosure laws in

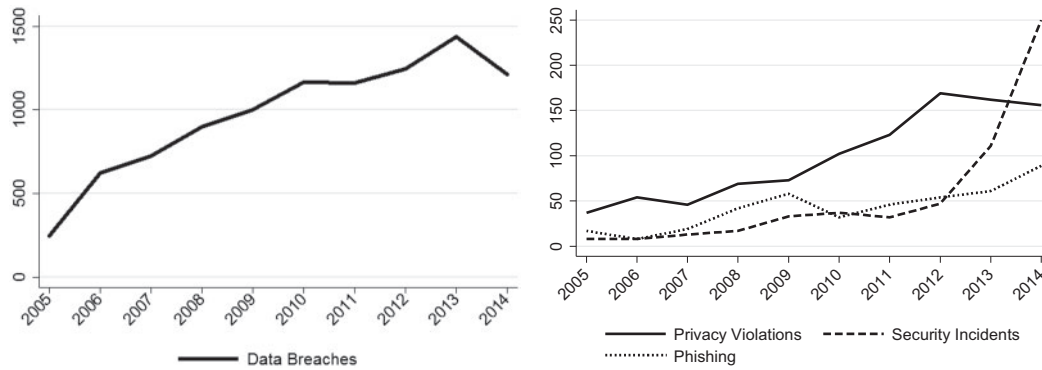


Figure 2. Four types of cyber events.

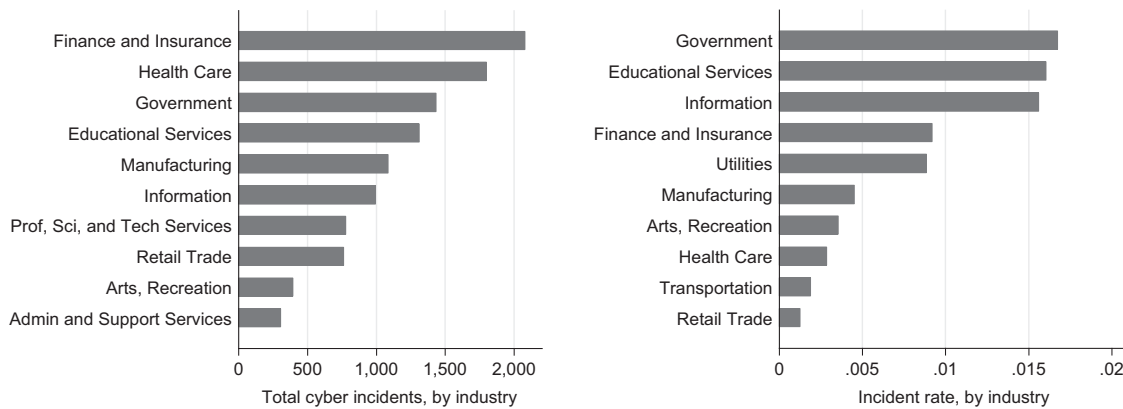


Figure 3. Cyber incidents, and rates, by industry.

later years. However, for most practical purposes, the early adopting states (such as California) could well have provided incentive for most all firms to report, regardless of the state of the individuals affected by a breach. The steady increase in privacy violations is likely due to the national attention that early privacy events and alleged violations occurred (such as Google Street View, Facebook’s rise in popularity, surveillance movement, behavioral advertisement tracking, etc.).

Security incidents, on the other hand, have seen a very sharp rise since 2012 (64) rising to almost 250 by 2014. The underlying cause is unclear given that there has been no substantial policy or industry interventions which would drive material changes in reporting. Absent such reporting changes, one might conclude that the rise is due to an increase in actual events. i.e. that firms are simply suffering more security intrusions.

Next, we examine incidents and incident rates by NAICS industry, as shown in Figure 3 [Note that the NAICS (North American Industry Classification System) coding, while ubiquitous, provides only one segmentation of firms by industry and may not reflect one’s intuition of a firm’s industry. For example, Apple Inc is coded as a Manufacturing company, and health insurance companies are categorized under Finance and Insurance, rather than Health Care. The Standard Industry Classification (SIC) codes are an alternative method for segmenting industries.] (See Appendix for further descriptions of firm types by industry). The left panel identifies the most frequent incidents by industry, while the right panel illustrates the incident ‘rate’ (i.e. number of cyber incidents divided by the total number of firms within that industry). That is, the left panel shows the total count of incidents, while the right panel shows the

percentage of firms within an industry that suffer an incident (The US Census collects many industry-level data, including the total number of firms, establishments, employees, and payroll. The number of public agencies was collected from the dataset: Federal, state, and local government unit: <http://www.census.gov/govs/cog/>, for 2012). See Appendix Table A1 for a full description of all industries, as well as the number of firms per industry.

The results in the left panel show that Finance and Insurance (e.g. insurance carriers, credit intermediaries), Health Care (hospitals, ambulatory care), and Government entities (courts, police, administrative offices, etc.) suffered the highest number of reported breaches of all industries in our dataset, followed by Educational services (schools, universities, and supporting services), Manufacturing (e.g. computer and electronic manufacturing), and Information services (e.g. data processing and hosting services, software vendors, telecommunications companies, internet portals, etc.), respectively. The large number of incidents of Health Care and Finance and Insurance matches other survey results which found that the healthcare and financial services sectors suffered the largest percentage of breaches from their sample (NetDiligence 2014). However, care should be taken to not immediately conclude, simply based on total incidents, that these industries pose the greatest risk of a cyber event.

For example, when we consider incident rate in our dataset (right panel), we find that Government agencies, and firms within the Education and Information services industries are affected at a much higher rate (>1.5%) compared with all other industries. That is, about 15 out of every 1000 firms have suffered a ‘reported’ incident, while firms within the Finance and Insurance, and Utilities industries suffered cyber incidents at a rate of about 9 out of every 1000 firms.

Table 1. Descriptive statistics

Variable	N	Mean	SD	Median	Min	Max
Records compromised ^a	1201	2.39 m	19.2 m	100	1	400 m
Employees ^b	10 929	20 491	123 k	300	1	2.8 m
Revenues (millions)	9360	8031	30 373	64	0	484 b
Type of information compromised^c						
Name (%)	12 574	43.84				
Credit card (%)	12 574	24.9				
Date of birth (%)	12 574	21.73				
Medical (%)	12 574	19.38				
Address (%)	12 574	18.93				
Financial (%)	12 574	17.07				
Email (%)	12 574	14.81				
Drivers license (%)	12 574	11.75				
SSN (%)	12 574	3.07				
Firm type						
	No.	Percent				
Privately held (%)	7072	56.2				
Publicly traded (%)	3350	26.6				
Government (%)	1762	14				
Nonprofit (%)	401	3.2				
Cause						
				Malicious?		
Disclosure/disposal (%)	2841	22.7		N		
Stolen HW (%)	3541	28.29		Y		
Hack/DDoS (%)	2132	17.03		Y		
Insider (%)	1342	10.72		Y		
Unauthorized use or collection (%)	970	7.75		NA		
Lost HW (%)	701	5.6		N		
Espionage, extortion, fraud, bribery (%)	504	4.03		Y		
Phishing (%)	215	1.72		Y		

^aRefers to the number of records containing personal information that were lost or stolen. Note that the mean is computed for all observations with nonzero/missing data.

^bThat is, employees and revenues related to the firm affected by the cyber event.

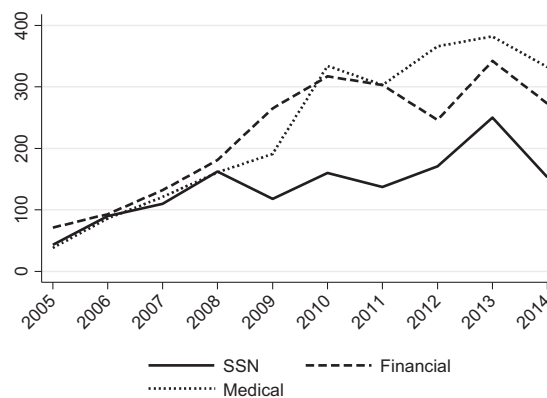
^cRefers to the percent of observations that include the compromised data type.

Note that Health Care and Retail industries, however, suffer extremely low incident rates of around 0.3% or less.

It is certainly true that not all cyber events are detected, and even if detected, they may not be recorded in the dataset (Figure 1). It is also true that there currently exists no reliable way to estimate the number of ‘unknown’ cyber events. However, given that there are approximately 6 million US firms (See data from the US Census at <http://www.census.gov/>), claims by some that “there are two kinds of companies: those who know they have been hacked, and those who don’t yet know they’ve been hacked” are, despite being a catchy media slogan, very likely false because it would suggest that all 6 million firms had, indeed suffered a data breach or cyber incident of some kind [This claim was self-attributed by Dmitri Alperovitch at a public symposium in Washington DC on 10 September 2015 who claimed to be the first person to make this statement. However, the same comment has been attributed to Congressman Mike Rogers, Chairman of the House Intelligence Committee (see <http://dritoday.org/post/Should-Directors-Officers-Be-Concerned-About-Cybersecurity.aspx>) from 2011. In addition, Misha Glenny cited a “friend” who made the same statement to him in 2011 (see http://www.ted.com/talks/misha_glenny_hire_the_hackers/transcript?language=en). Whether the “friend” referred to is Mr. Alperovitch, Congressman Rogers, or someone else, is unclear.]

Additional descriptive statistics are shown in Table 1.

Next we examine the rates of incidents in our dataset according to the kinds of information compromised. Specifically, Figure 4 shows that the rates of compromise of more personal information (and those that could lead to greater consumer harms) is increasing with time,

**Figure 4.** Cyber events by type of information compromised.

and particularly, the number of cyber events that involve medical information has risen most sharply. These trends are particularly concerning given that these data are more difficult (or in the case of medical data, impossible) to change, and therefore individuals suffering a compromise of these data are arguably more at risk of financial, medical, and other forms of fraud and identity theft.

In addition, contrary to other findings suggesting that the rate of malicious incidents (i.e. those committed with intent to cause harm) has been increasing in recent years (relative to accidental ones), we find a relatively stable proportion of malicious events. Specifically, surveys conducted by the Ponemon Institute have shown that these rates

have been increasing from 37% in 2011, to 44% in 2014 and 49% in 2015 (Ponemon 2011, 2014, 2015). By grouping the causes of incidents according to whether the cause was malicious or not (i.e. intentionally committed to cause harm, as shown in Table 1), the rates of malicious incidents from our dataset are presented in Figure 5.

Based on our data, this suggests that the rates of malicious events has remained quite steady at 60% over the past decade. That being said, the slight increase in the most recent years may, indeed, be due to the increase in security incidents and attacks against corporate systems.

Legal actions

Figure 6 illustrates the composition of legal actions is distinguished along three dimensions: civil vs criminal, federal vs state, and private vs public actions. As fully described in Romanosky et al. (2014), private actions typically reflect cases brought by individuals against

firms for the unauthorized disclosure (or use) of their personal information (For example, many arguments in data breach cases include those similar to the ones presented in *Hilary Remijas v Neiman Marcus Group, LLC* (No 14-3122), “(1) lost time and money resolving the fraudulent charges, 2) lost time and money protecting themselves against future identity theft, 3) the financial loss of buying items at Neiman Marcus that they would not have purchased had they known of the store’s careless approach to cybersecurity, and 4) lost control over the value of their personal information.”). Similarly, public actions (whether federal or state) are brought by government agencies (e.g. FTC, state AGs) against companies for their allegedly inferior security practices or careless handling of personal information. On the other hand, the criminal actions are brought by State prosecutors against alleged perpetrators of the crimes and reflect a different category of lawsuit. The type of action is identified along with the number of observations by category, in parentheses. The weight of the line roughly corresponds to the relative number of observations from our dataset.

Of the almost 1700 actions, about 83% are civil, 79% are federal, and 68% are private (combining both federal and state actions). Of just the civil actions, notice that there are about four times as many federal as state actions (1123 vs 271), and of those federal, almost five times as many private suits relative to public enforcements (922 vs 201). Of the 271 state actions, there are over four times as many private actions as there are public (221 vs 50). Criminal actions make up 293 of all legal actions, with the majority of them being prosecuted in federal court.

Next, we examine the frequency of actions by year of filing as shown in Figure 7. The left panel illustrates the count of ‘private’ lawsuits brought in federal and state court, while the right panel shows the count of ‘public’ (civil and criminal) actions brought in federal and state court from 2005 to 2014 [Note that approximately 200 suits were omitted from this analysis because of multiple suits

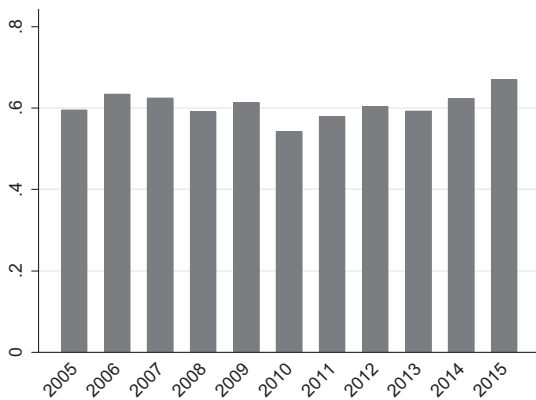


Figure 5. Rates of malicious events.

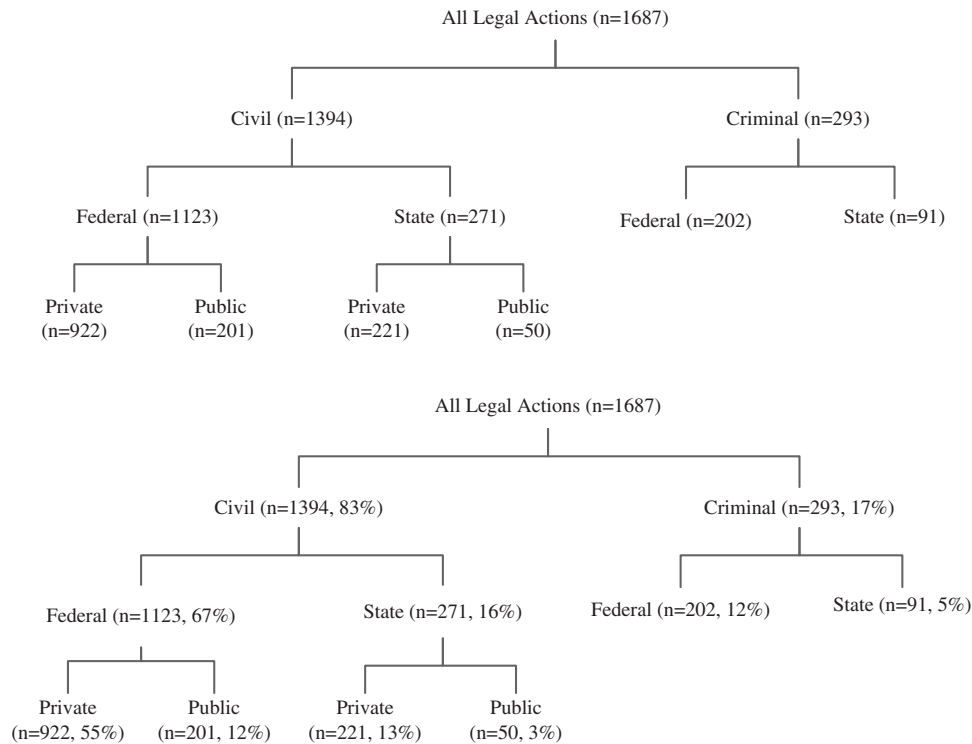


Figure 6. Composition of legal actions.

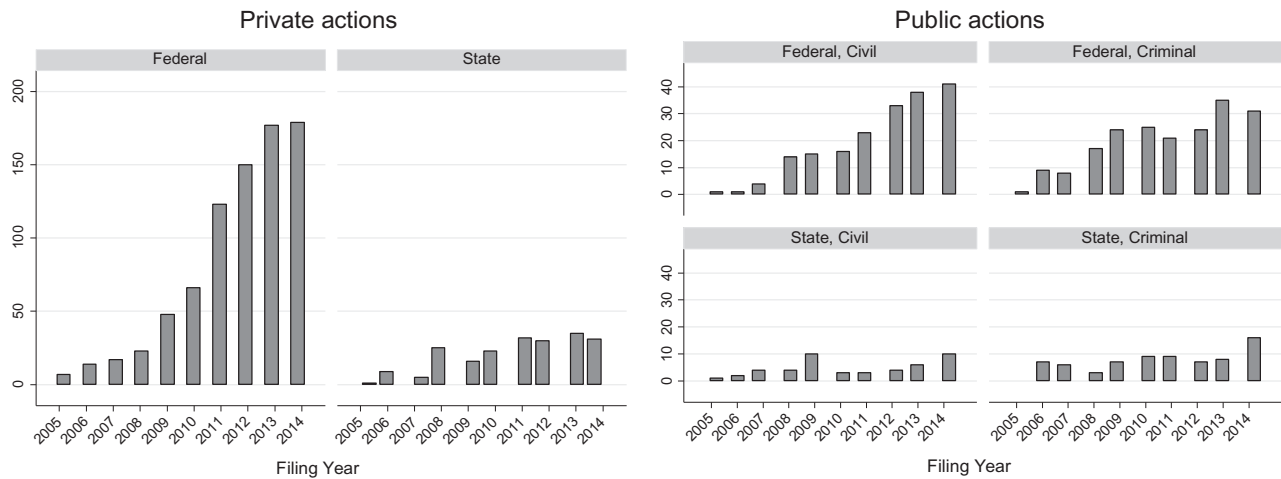


Figure 7. Count of private and public legal actions by year.

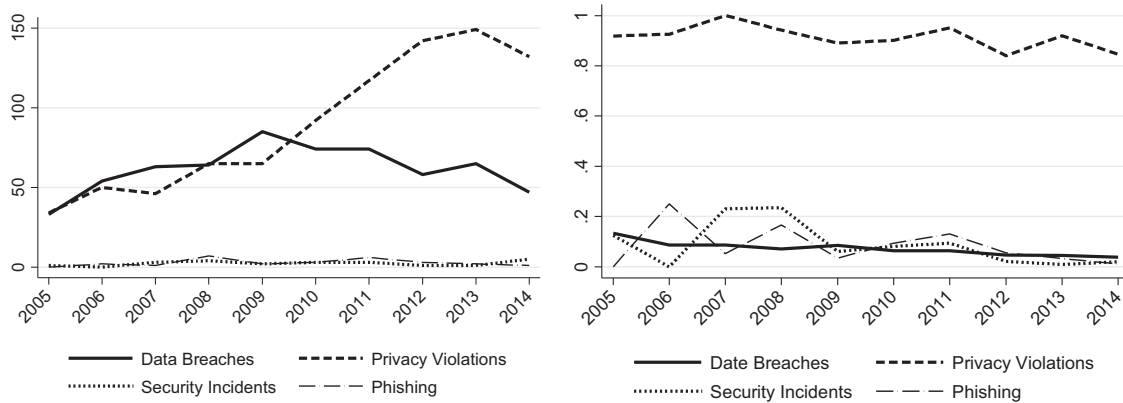


Figure 8. Legal actions and litigation rates, by type of cyber event.

being filed for the same cases, most of which are eventually consolidated into multi district litigation (Romanosky 2014).]. We limit this examination to years beginning in 2005 since this was objectively the first year that systematic reporting and recording of data breaches (and therefore subsequent litigation) began [Data breach disclosure laws first began in 2003, but it was not until 2005 that additional states began adopting similar laws (Romanosky 2011). Further, it was not until 2005 that other nonprofit organizations such as DatalossDB, and the Privacy Right Clearinghouse began systematically recording and reporting these events.].

These histograms help illustrate the change in number of lawsuits filed over time. First, note that the number of public, federal actions has been increasing since 2005 (upper histograms of the right panel), and greatly outnumbers public state actions (lower histograms of the right panel). However, the majority of all legal actions for cyber events are driven by private (civil) actions filed in federal court (left histogram of left panel), and the number of suits has been increasing steadily from a couple of dozen in 2005, to almost 200 by 2014. Interestingly, the causes of action alleged by plaintiffs in these cases include a wide variety of common law and statutory claims, such as unfair business practices, negligence liability, breach of contract, unjust enrichment, privacy torts, the video privacy protection act, drivers privacy protection act, etc. (see Romanosky et al. 2014 for a full description and analysis of these causes of action).

An important qualifier for these data is that state lawsuits are much more difficult to observe and track relative to federal suits

because of the lack of systematic search capabilities of state dockets. And so it is possible that the low number of observed private state suits is largely a function of their absence from our dataset rather than the absence of such cases.

Even though federal civil and criminal actions in our dataset have been increasing since 2005, they are clearly not increasing at the same rate as private actions, nor are they likely to reach such levels. The reason is simply that public agencies are much more resource constrained relative to the number of private litigants, suggesting that public agencies must therefore rely on suits brought by private individuals in order affect change.

Litigation rates

Next we examine the rate of legal actions stemming from the cyber events in our sample. The left panel of Figure 8 shows a count of the number of suits by event type, while the right panel shows the proportion of reported litigated events (aka litigation rate), also by event type. That is, the ratio of lawsuits filed in a given year to the number of events occurring in that year (Given that lawsuits are not always filed within the same year of the event, this formulation of the litigation rate is an approximation).

The left panel, lawsuits, shows that while data breach lawsuits have remained relatively stable at just over 50 per year, lawsuits regarding privacy violations have been increasing dramatically since 2005 and especially since 2009, reaching as many as 150 suits per

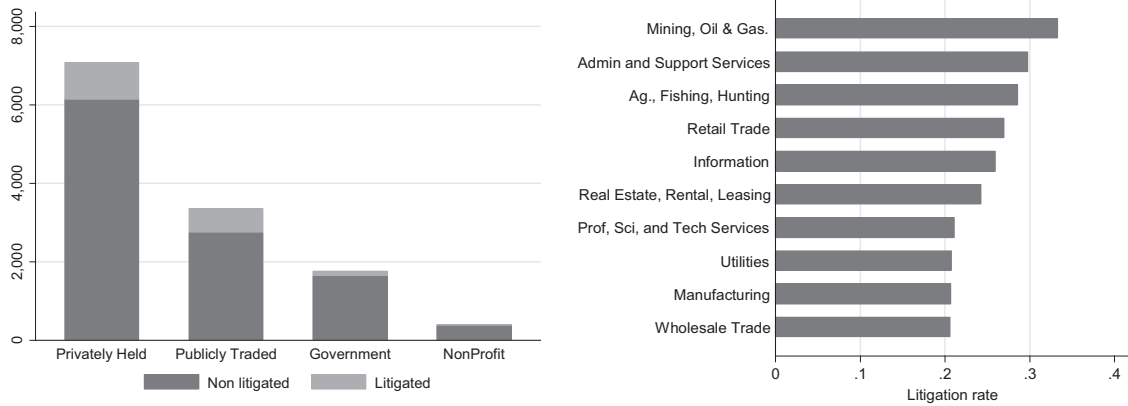


Figure 9. Litigation rate by organization type, and industry.

year in our sample. Of the approximately 900 privacy suits, over half (54%) relate to unsolicited telemarketing and advertising resulting from spam, advertising, phone calls, and faxes. In all, 17% of cases (34% in total) each relate to unauthorized surveillance caused by nonconsensual voice or video recording (eavesdropping or wiretapping), and to alleged harassment from debt collection. About 3% of cases each stem from allegations of cookie tracking, violations of federal and state privacy statutes such as the Video Privacy Protection Act (VPPA), Drivers Privacy Protection Act (DPPA), Child's Online Privacy Protection Act (COPPA), and the California Song-Beverly Act (restricting the collection of a consumer's zipcode during purchase).

Actions for security and phishing incidents, however, have experienced very little increase since 2005 (Note that these values may be slightly underrepresentative of the actual number, since we are only considering unique suits for a given incident.). Despite sharp increase in cyber events in recent years (Figure 2), the litigation rate for all cyber events has been generally decreasing in our sample. For example, the litigation rate for data breaches was around 20% in 2004, but has fallen to about 5% in 2014. This substantiates a similar litigation rate as found by Romanosky et al (2014).

Litigation rates for privacy violations, however, appear dramatically different, suggesting an overall average of around 90%. However, this is likely an overinflated result, driven by artifacts of data collection. The most reasonable explanation is that privacy violations, themselves, are only observable once a lawsuit has been filed (this artifact of privacy incidents represents a considerable hindrance to empirical research in that investigators are unable to observe the total number of privacy events, and therefore distinguish among those that have been litigated), whereas events such as data breaches would likely not suffer from this because of the state breach disclosure laws which provide a legal obligation for firms to report these events to affected individuals (Romanosky et al. 2011).

Legal actions by organization type

While Figure 3 displayed rate of cyber events by industry, Figure 9 provides more information regarding litigation rate by entity type (left panel) and industry (right panel) in our sample. From these figures, we observe that while privately held firms experience the largest total number of breaches (approx. 7000) and lawsuits (over 1000), publicly traded companies face the largest litigation rate (643 lawsuits/3300 breaches = 19.4% litigation rate). It is likely unsurprising to observe that government agencies and nonprofit companies experience relatively small litigation rates (approx. 8% and 9%, respectively), likely because of sovereign immunity (regarding state agencies).

The right panel of Figure 9 shows the top 10 litigation rates, by NAICS industry. Note that while entities in the Government and Educational sectors suffer the highest 'breach' rates (as seen in the right panel of Figure 3), they suffer 'litigation' rates less than 10% (data not shown). Contrary to conventional thought, however, Finance and Insurance, and Health care companies also suffer relatively low (around 10%) litigation rates. On the other hand, Administrative and Support Services (e.g. telemarketing, collection agencies and credit bureaus, etc.) Retail firms, and Information Services (e.g. software vendors, telecommunications companies, internet portals, etc.), in our dataset suffer litigation rates over 25%, while the Mining and Oil & Gas industry (mining, petroleum, gas extraction, drilling, and supporting activities) suffers the highest litigation rate of all other industries (i.e. more than 30% of all cyber events are litigated) [This high rate may well be due to the relatively low number of firms in this sector (approx. 22k) relative to, e.g., the finance industry which has over 234k firms.].

Repeat players

We now examine whether there exist any systematic differences between risk and cost borne by those firms that have never previously experienced a cyber incident, and those that have experienced two or more (what we refer to as a repeat player, a term borrowed from game theory). Specifically, we examine various outcomes conditional on the firm already having suffered a cyber event.

First, the distribution of repeat players is highly skewed. Of the 12 603 cyber incidents, 62% of firms suffered only a single incident, suggesting that 38% of firms in our dataset (almost 4800 firms!) suffered two or more cyber incident. Specifically, as shown in Figure 10, 25% suffered between 2 and 5 incidents, around 4% suffered between 6 and 10, and 8% of all firms suffered more than 10 cyber incidents.

Next, we reexamine the total number of cyber events by industry (per Figure 3, left panel), but now highlight the proportion of those events suffered by repeat players, as shown in Figure 11. We find the industries that bear the greatest percentage of repeat players are Information (56%) and Finance and Insurance (51%). That is, 56% of all cyber incidents that occur to firms in the Information sector have suffered at least one previous incident. These industries are followed by Educational Services (44%), and Manufacturing, Retail, and Government each with around 37%.

In analysis similar to Figure 7 (left panel) and Figure 9 (right panel), we find no substantial change in the percentages of lawsuits filed against repeat players over time, nor any differences across

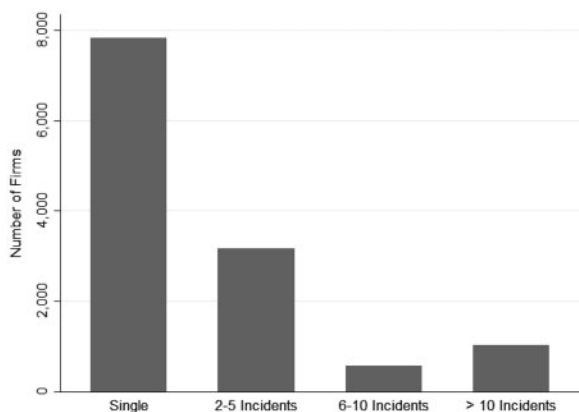


Figure 10. Distribution of repeat players.

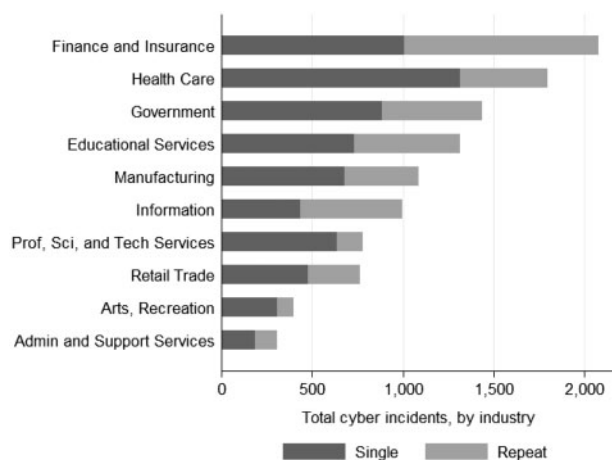


Figure 11. Repeat incidents by industry.

industries in the litigation rate between single and repeat players. That is, firms that suffer multiple cyber events do not appear to be more likely to be sued, relative to firms that suffer a single incident.

Cost of cyber events

The costs incurred by cyber events can largely be differentiated among first and third party losses. First-party losses relate to expenses the firm incurred as a direct result of the incident. For example, in the case of a data breach, this would include the cost of forensic investigation in order to determine the cause, the cost of notifying affected consumers, marketing or public relations campaigns, customer support efforts, and any consumer redress in the form of credit monitoring or identity theft insurance. Third-party losses, on the other hand, relate to costs incurred due to private litigation [e.g. class action lawsuits, judicial rulings, settlements, or *cy præs* awards (in cases of data breaches, money from the defendant is sometimes allocated to assist with identity theft education and awareness, or to fund research in data protection or consumer privacy)], or fines or fees brought by government agencies. In the case of security incidents or phishing/skimming scams, losses may include the dollar value of any financial theft.

There are a number of important qualifiers regarding the cost figures that follow. In some cases, they represent estimated or alleged losses, rather than actual, verifiable costs incurred. They are incomplete in that they likely do not include all costs borne by firms due to these cyber events. For example, they do not account for lost

Table 2. Cost by event type (in millions)

Event type	N	Mean	SD	Median	Min ^a	Max
Data Breach	602	5.87	35.70	0.17	0.00	572
Security Incident	36	9.17	27.02	0.33	0.00	100
Privacy Violation	234	10.14	55.41	1.34	0.00	750
Phishing	49	19.99	105.93	0.15	0.01	710
Total	921	7.84	47.28	0.25	0.00	750

^aValues are presented in millions of dollars and therefore, any zero values are artifacts of rounding functions.

revenue, sales, market valuation (numerous other studies have, of course, examined the effect of data breaches on market valuation and generally find no causal effect). They also do not include intangible or nonfinancial costs such as lost time due to a fired CEO, or any intangible measure of loss of reputation. They also only represent a small percentage of the total observations within this dataset (cost data are available for only 921, or 7.3% of all observations), which is in itself only a subset of all publicly reported breaches, which is a subset of all known events, which is finally only a subset of all actual cyber events (Figure 1). These data do not include the costs borne by consumers (or others) due to identity theft, other sorts of financial, medical, or privacy harms, or out-of-pocket expenses. Finally, they do not include other social costs or externalities borne by other parties because of these events. Therefore, these costs represent a sample (albeit large) of estimated costs incurred by firms due to a cyber event. Summary statistics are shown in Table 2.

Assessing and predicting the costs of data breaches has been a struggle for many years because of the lack of quality data. And naturally, many organizations have an interest in better understanding these costs, e.g., those firms at risk of suffering breaches, insurance carriers, researchers, and social planners. Based on recent survey data, current estimates present the average cost of a data breach at around \$6.5 million (or, \$217 per record; Ponemon 2015). However, given the heavily skewed cost distribution from these data, use of the statistical mean as a measure of the cost of a data breach (or cyber event) is misleading. As shown in Table 2, while the mean loss for a data breach is almost \$6 million, the median loss is only \$170k. Similarly skewed values arise for phishing and security incidents. Privacy violations, however, account for a much larger median loss of \$1.3 million, but is also greatly skewed (Note that cost data are available for a small percentage of all observations and therefore it is possible that these cost data are skewed toward more costly events. If true, this would cause these estimates to be inflated, relative to the true values.).

Similar cost results for data breaches were found using a remarkable dataset of actual cyber insurance claims data which finds median claim payouts of \$144k, and mean payouts of almost \$3 million for large companies (NetDiligence 2014) (These data are remarkable because it is extraordinarily difficult to obtain actual claims information.). In addition, other research that examined a large sample of loss data found estimate of \$1.68 million with a mean of almost \$20 million (Biener 2015). The similarity between estimates is comforting and provides some validation of the cost data used in this analysis. That is, if the insurance claims account for as much of the liability and financial loss suffered by the firm as possible, then one would welcome a strong correlation between those data, and the cost estimates from our dataset.

Next we examine the change in total costs, by event type, over the 10-year period from 2004 to 2014 as shown in Figure 12. Given that the raw data exhibit extreme annual fluctuation, we therefore

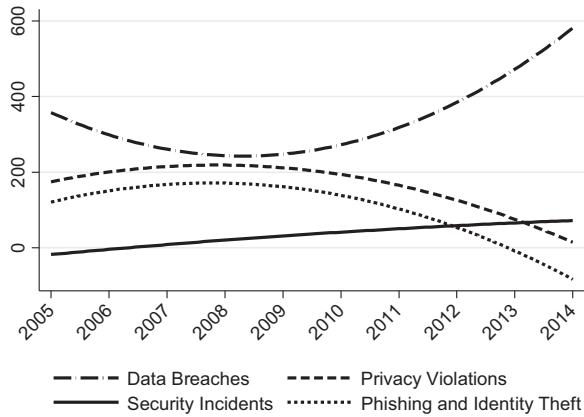


Figure 12. Total cost of cyber events over time.

Table 3. Costs of events for repeat players

		Single player	Repeat player	Stat. different?
Data breaches	Mean	\$4.07 m	\$9.78 m	Yes ($P < 0.052$)
	Median	\$150 k	\$280 k	Yes ($P < 0.005$) ^a
	N	408	194	
Privacy violations	Mean	\$10.6 m	\$8.77 m	No
	Median	\$1.18 m	\$2.51 m	No
	N	168	66	

^aComputed using Pearson chi-squared test of the equality of the medians. Differences are also robust to the Wilcoxon–Mann–Whitney test.

display smoothed lines in order to provide directional insights (Note that it is this smoothing process which produces values less than zero. Clearly negative costs are not observed in our data.).

First, we observe that total costs for privacy violations and phishing attacks have been generally decreasing since 2008, as illustrated by the middle two dashed curves. On the other hand, costs from security incidents have been increasing steadily since 2005, though only at a moderate rate. Total costs for data breaches, however, while declining slightly from 2005 to 2008, increased steadily and dramatically after 2008. While this is in part driven by the increasing number of data breaches, it provides evidence showing how overall costs from these cyber events are, indeed, rising. Moreover, we find that these trends are driven by the large firms in our dataset – i.e., firms within the top 25 percentile of assets. Costs for medium and smaller firms have been decreasing steadily since 2005.

Repeat Players

If we reexamine costs incurred by repeat players, we find that the impact of cyber events varies considerably for this group, relative to firms that incur only a single incident, as shown in Table 3.

The mean cost of a Data breach to repeat players is almost twice that of single-breach firms (\$4.07 million vs \$9.78 million). However, in the case of Privacy violations, mean costs are actually lower for repeat players (\$8.77 million vs \$10.68 million). In addition, the median cost to repeat players is twice that of single players (\$150k vs \$280k), and similarly for privacy violations (\$1.18m vs \$2.51m). Further, while neither the mean nor the median costs of Privacy violations are statistically different (between single and repeat players), they are statistically different for Data breaches (the difference in means is significant at the 10% level, and the difference in medians is significant at the 1% level).

Costs by industry

Next we examine the losses by industry in our dataset, as shown in Figure 13. Note that for brevity, we plot only the top 10 industries. The left panel identifies 10 industries that suffer the greatest losses as a result of cyber events. However, because these data may be driven simply by the number of cyber events within each industry, the right panel divides the loss by the number of events (i.e. it shows the total loss divided by the number of events in our dataset) (Note that we omit six observations that are extreme cost outliers.).

The two panels show that overall, the Information, Manufacturing, and Retail industries suffer the greatest losses relative to any other industry, as well as the greatest loss per event (though in a different order). This additional measure is useful in better understanding in which industry the greatest losses exist, and therefore which industries pose the greatest risk to firms, investors, employees, and potentially consumers.

An important conclusion from this work is that while Information and Retail industries incur the greatest losses, they also seem to incur the greatest likelihood from legal action. Further, despite being the most heavily regulated in terms of information security controls, firms within the finance industry do not appear to be better able to resist cyberattacks or avoid losses in a materially better way, relative to other industries (of course, we do not observe the number or severity of attacks launched against these firms).

Modeling the costs of data breaches and privacy violations

Of particular interest to insurance carriers and, indeed, firms, is to be able to develop useful predictive models concerning the costs of data breaches and other cyber events. However, very little empirical research has been conducted, and the work that does exist provides only basic insights. For example, using Ponemon data, Jacobs (2014) creates the following regression model using number of records compromised,

$$\log(\text{impact}) = 7.68 + 0.76 * \log(\text{records}) \tag{1}$$

Where ‘impact’ refers to the cost of a data breach, and ‘records’ refers to the number of records compromised. The interpretation of this simple regression equation is that as the number of records increases by 10%, the cost of a data breach would increase by 7.6%. While limited, this model is informative and shows an increasing trend between size of a breach and the incurring costs. In related work, Edwards et al. (2015) use data collected from the Privacy Rights Clearinghouse to create a model associating stratified sizes of breaches with the probability of those occurrences. Unfortunately, based on data limitations, they do not model costs.

However, given our rich set of data, we are able to develop a more comprehensive model for cyber incidents that helps better understand the relevant factors driving costs. Consider the following estimating model,

$$\log(\text{cost}_{it}) = \beta_0 + \beta_1 * \log(\text{revenue}_{it}) + \beta_2 * \log(\text{records}_{it}) + \beta_3 * \text{repeat}_{it} + \beta_4 * \text{malicious}_{it} + \beta_5 * \text{lawsuit}_{it} + \alpha * \text{FirmType}_{it} + \lambda_t + \rho_{ind} + \mu_{it}, \tag{2}$$

where *cost* is the total cost of the incident (first and third party losses) incurred by firm *i* in year *t*. *revenue* is the log of the firm’s revenue. *records* is the number of compromised records from the incident, *repeat* is a binary variable coded as 1 if the firm suffered multiple events, and 0 otherwise. *malicious* is a binary variable coded as 1 if the event was caused by malicious intent, *lawsuit* is a binary variable coded as 1 if a legal action resulted. *FirmType* is a vector of

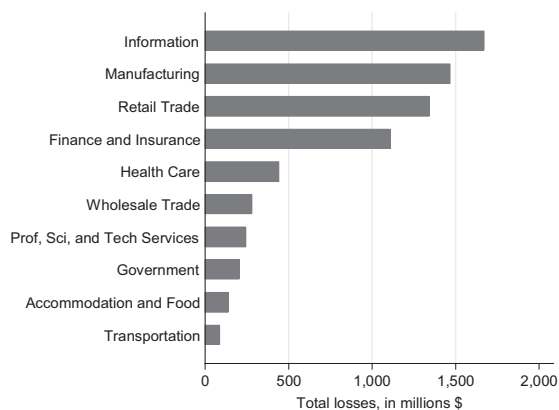


Figure 13. Losses by industry.

binary variables describing whether the affected firm was a government agency, nonprofit, privately held company, or publicly traded (given that these represent a set of completely exhaustive and mutually exclusive categories, by convention we omit one category – nonprofit firms). We also include vectors of year and industry binary variables, represented by λ_t and ρ_{ind} . μ_{it} is the error term, assumed to be uncorrelated with the covariates.

The results from Equation (2) are shown in Table 4. Note that while there exist over 12 000 observations in the full dataset, because of data limitations and missing observations, we estimate Equation (2) for data breaches only. Also, note that this model provides an association between cost and a cyber event, and is not intended to represent causal inference.

These results suggest that revenues are strongly associated with incident cost (We use firm revenues because it provides a reasonable proxy for firm size. Number of employees is a legitimate alternative. However, the correlation between revenues and employees was 0.935 and the regression results presented are robust to the substitution of revenues with employees.). For example, a 10% increase in firm revenues is correlated with a 1.3% increase in the cost of an incident (significant at the 5% level). The number of records compromised was also found to be strongly correlate with incident cost (i.e. a 10% increase in the number of records compromised is associated with a 2.9% increase in cost, significant at the 1% level), approximately one-third the magnitude of the Jacobs (2014) model. This result is unsurprising as one would expect that larger breaches impose greater cost.

Interestingly, despite the unconditional yet statistical difference in cost for repeat players (Table 3), firms that suffered multiple incidents were not found to bear a significant change in cost. Further, while one might expect that malicious events, and those that incur litigation would suffer greater loss, these results do not bear out in this estimation. Overall, however, these results suggest that total losses from a data breach are most strongly correlated with the size of the firm and size of the breach, rather than any particular motives by an attacker, or other outcomes.

Placing costs in context

A critical policy question, and one that this article seeks to resolve is: how much of a problem are cyber events? In order to help address this, we examine the total costs as a function of annual losses recorded in other sectors, and costs as a function of firm revenue.

To begin, we observe cost information on 7.6% of the observations in the data (i.e. 887 out of 11 705 observations between 2005 and 2014), which totals to \$6.5 billion (Given that only partial data

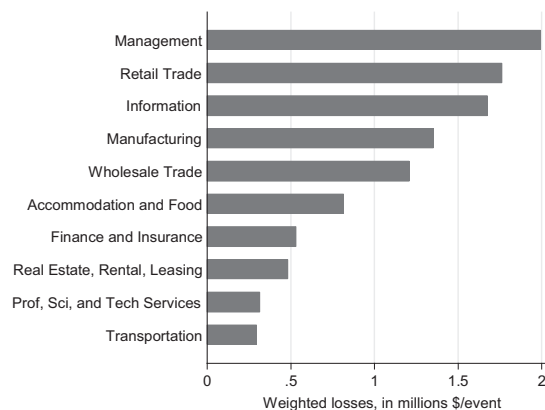


Table 4. Regression results for Equation (2)

Dep var: log(costs)	Data breaches
Log(revenues)	0.133** (0.0592)
Log(records)	0.294*** (0.0386)
Repeat player	-0.352 (0.386)
Malicious	-0.0294 (0.369)
Lawsuit	0.444 (0.351)
Government	-1.339 (1.482)
Private	-1.032 (1.145)
Public	-0.0654 (1.156)
Constant	-3.858* (2.044)
Observations	265
R ²	0.466
Year controls	Yes
Industry controls	Yes

Heteroskedastic robust standard errors in parentheses.
 ***P < 0.01,
 **P < 0.05,
 *P < 0.1.

were available for the 2015 calendar year, this year was dropped for the purpose of this analysis. Therefore, annual cost estimates are based on 11 705 observations instead of 12 603.). If we consider these data to be representative of the full population of (reported) cyber events (this is obviously an approximation, but provides one means to estimate the annual cost), this suggests that the total cost of cyber events (over the 10 years between 2005 and 2014) was approximately \$85 billion or about \$8.5 billion annually (\$85 billion was computed as \$6.5 billion/7.6%. Dividing by 10 years produces an annual estimate of \$8.5 billion in loss. Clearly, this is an estimated value based on the observations for which cost data are available. Further, this is likely an underestimate of the true value given that some breaches and security incidents are not reported, and that it does not include consumer or other social losses.). By comparison, Figure 14 shows cyber losses relative to other losses plotted on a cardinal scale ranging between \$0 and \$250 billion [Sources, cost of

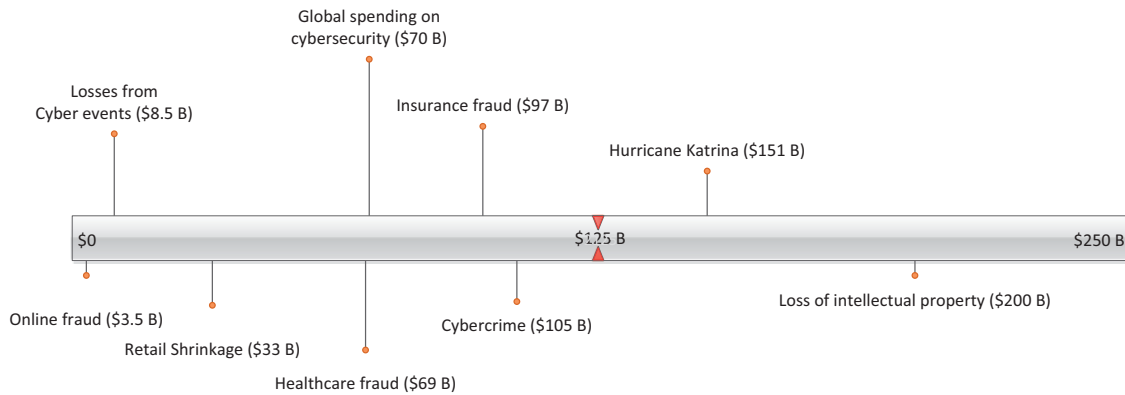


Figure 14. Relative costs and losses (annual, in billions of US dollars) (Indeed, there may be some overlap between some categories (i.e. intellectual property and cyber crime). However, these descriptions are presented as informational estimates only.).

insurance fraud and hurricane Katrina (<https://www.erieinsurance.com/about-us/insurance-fraud/cost>, <https://www.ncdc.noaa.gov/billions/events>), others (CSIS 2014). Note that this is clearly a one-time event, whereas other data points are estimates of annual losses.].

Observe that in relation to the other losses presented, cyber events from this dataset are relatively very small. In addition, with a 2013 US GDP of approximately \$16.8 trillion, these annual cyber losses represent a mere 0.05% of GDP.

A second approach to understanding the overall impact is to consider these costs as a percentage of firm revenues. A similar practice is used by financial lending institutions which face the chronic problem of balancing good versus bad debt (often referred to as bad debt expenses). While they cannot avoid all bad debt, some amount is tolerable, and indeed, efficient (Efficiency comes from the realization that while some effort spent to prevent fraud and abuse is cost-effective, at some point, each additional dollar spent to prevent further waste results in a benefit less than one dollar.). Firms, therefore, track these bad debt expenses and no doubt apply their own threshold for determining a tolerable amount of bad debt. Firms in the retail industry also closely track the loss (shrinkage) that they incur on an annual basis (In a retail clothing store, shrinkage represents lost or stolen merchandise. In the restaurant industry, shrinkage represents the portion of food or drink that is stolen, spoiled, or broken.). Indeed, tracking and reporting this metric is a familiar practice for many firms. For example, hospitals incurred 5.9% in bad debt due to uncompensated care in 2011 (Rodney 2013), while Xerox and Strayer Education incurred less than 1%, and 3.2% of bad debt, respectively (BusinessWire 2015; MarketWatch 2015). Estimates from the restaurant industry suggest much higher losses around 20% (Rowe 2011; Plotkin, undated), and annual shrinkage in the US retail industry was estimated at 1.44% of annual sales (\$33.5 billion; Kays 2010) (based on a survey of 100 retail clothing and supermarket stores).

Therefore, by dividing the total loss (sum of first- and third-party losses) by firm revenues, we obtain a distribution of loss ratios that is heavily skewed with very expensive cyber events drastically driving up mean calculations (Note that all firm revenues are computed as 2015 dollar values.). For example, from our dataset, 66 cyber events exhibited losses less 0.001% of revenue, and 90% of events exhibited losses less than firm revenues. Overall, we find that the median loss was just 0.4% of annual revenue (We removed six extreme outliers from these calculations as they were clearly anomalous events with loss/revenue ratios greater than 100.). That is, based on our data, we find that most cyber events cost firms less than 0.4% of their annual revenues.

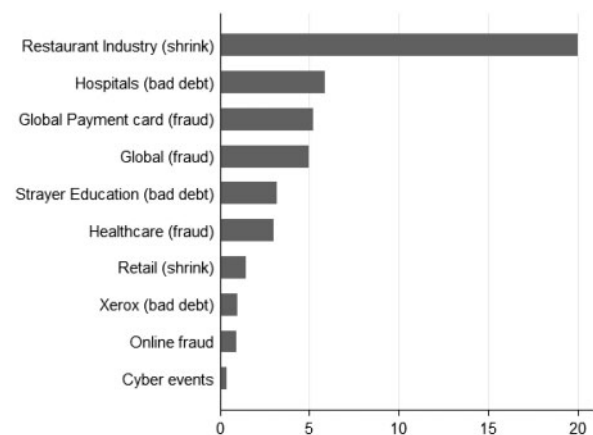


Figure 15. Loss as a percentage of revenues.

We continue this analysis by comparing these losses relative to other sorts of fraud. For example, the total loss from US cyber crime activities was estimated to be \$105 billion annually (CSIS, 2014) (This estimate is based on their estimate of cyber crime being 0.64% of the US GDP. Using a value of \$16.77 trillion, 0.64% is approximately \$107 billion), and US firms were estimated to have lost 0.9% of revenues to online fraud in 2013 (\$3.5 billion; Cybersource 2013) (see also <http://www.cardhub.com/edu/credit-debit-card-fraud-statistics/>). Healthcare fraud was estimated at 3% (\$69 billion) of total US health care expenditures (NHCAFA 2010). Globally, losses due to fraud accounted for approximately \$3.5 trillion, or 5% of annual firm revenues (ACFE 2012) (Based on a survey conducted by the Association of Certified Fraud Examiners between 2010 and 2011 with 1388 firms. Fraud includes corruption, financial misstatements, cash theft, billing and payroll fraud, larceny), while payment card fraud specifically, accounted for \$11.27 billion, or 0.05% of sales volume (Nilson 2013). These data are presented in Figure 15 (Note that these categories may not be mutually exclusive and are therefore presented as information only.).

If it is true that on average, businesses lose 5% of their annual revenue to fraud, and that the cost of a cyber event represents only 0.4% of firm revenues, then one may conclude that these hacks, attacks, and careless behaviors represent a small fraction of the costs that firms face, and therefore only a small portion of the cost of doing business.

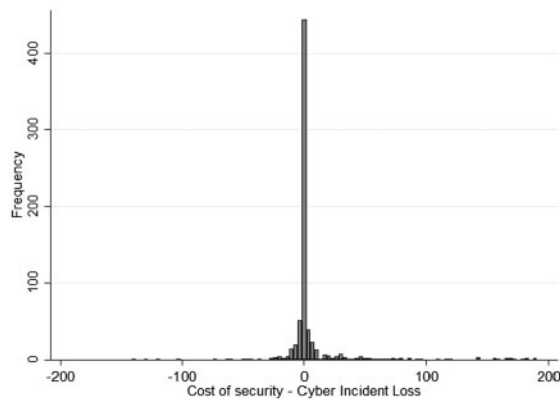


Figure 16. Incident costs – security spending (in millions of \$).

How much should a firm spend on IT Security?

A question that has plagued researchers, executives, and policy makers for decades is regarding the amount of money to invest in protecting a firm's corporate and customer databases, IT systems, and intellectual property. The *Gordon-Loeb model* (2002, 2015) is one attempt to provide such an answer. It suggests that a firm's investment in IT security should not exceed 37% of the losses it expects to incur from a data breach or cyber event (However, note that this estimate of 37% may be considered an artifact of the functional forms used in the model.).

Using data from this research, we can further understand the value of IT security by comparing the losses from these events, with a firm's investment in IT security. A study from 2013 found that an average firm's IT budget was around 5% of its revenues (CIO 2013), and a Gartner survey of 1500 firms in 2010 found that firms spend an average of 5% of their IT budget on information security (Kirk 2010). Therefore, if we assume that IT security spending is, on average, 0.025% of revenues, then we can compare this with actual cyber event losses from our dataset. *Figure 16* presents a histogram showing the difference between cyber incident costs and IT security spending (Note that we omit some extreme outliers for presentation purposes.). The left panel shows observations within $\pm\$200$ million, while the right panel displays the (same data, zoomed in) distribution for values within $\pm\$10$ million.

Values to the left of zero imply that incident costs are greater than IT security costs, while values to the right of zero imply that IT security costs exceed the cost of a breach. A surprising observation is the large mass close to zero. There is no reason to expect, *a priori*, that IT security budgets should so closely match breach costs. Specifically, 77% of incidents cost firms between $\pm\$10$ million of its security budget, and 50% of incidents cost firms between $\pm\$1$ million of its security budget. That is, these data show that half of cyber events cost a firm an amount approximately equal to its annual investment in IT security (i.e. within $\pm\$1$ million of investment).

On one hand, an executive who is skeptical of security investments may believe that unless a firm incurs a breach every year, it is wasting its IT security investment every year it does not suffer a breach. Alternatively, it may imply that a firm can expect to lose the equivalent of its IT security budget each time it suffers a data breach or security incident.

Limitations

The data used in this analysis are acquired from an impressively diverse and comprehensive collection of search strategies employed by Advisen that include scouring local and news sources, searching

legal databases, data breach clearinghouses, and government websites. Indeed, while other leading clearinghouses contain less than 5000 incidents [As of early 2015, the Privacy Rights Clearinghouse, a nonprofit organization dedicated to issues concerning consumer privacy, was cited as having less than 4500 incidents (Edwards et al. 2015)], the full dataset includes over 12 000. However, despite the data likely being the most comprehensive sample of cyber incidents available, they are still based on publicly available data, and therefore suffer from a number of limitations.

First, they would not reflect events which are unobserved by the firm. Similarly, events which are observed but not reported or discovered publicly would also not be included. Further, despite the comprehensive search strategies employed, the observations recorded may still be biased toward larger, more prominent cyber incidents. Therefore, all inferences and conclusions from this analysis relate to known, cyber incidents from these data. However, even if the data were biased toward larger, more severe or prominent cyber events, it is precisely these events which one would expect to be used to inform and drive public policies, firm practices, and regulatory oversight.

Finally, in regard to the cost data, these are clearly based on estimated values for some firm-level categories, and omit other categories such as impact to fired staff or executives, market valuation, etc. Further they do not include consumer losses such as lost time or money, personal anxiety, or emotional distress.

Conclusion

We believe the analysis provided in this article will be relevant in a number of ways to firms, policy makers, consumers, and particularly insurance companies. First, this research has uncovered an interesting paradox. On one hand, aggregate rates of cyber events and litigation both show similar trends – that they are more frequent and therefore potentially more expensive to organizations collecting and using personal information. In addition, the kinds of information being compromised (SSN, medical, and financial), are those that could well lead to more severe and longer lasting forms of consumer identity theft and fraud.

On the other hand, as we examine the actual costs of these events in our dataset (clearly one of the most important outcome measures), we find that they cost most firms less than \$200k, only a fraction of the millions of dollars commonly cited. We also estimate that they represent only 0.4% of firm revenues, far less than other losses due to fraud, theft, corruption, or bad debt (Clearly, however, in some cases, data breaches and other cyberattacks have caused massive losses to firms, as well as some cases of identity theft do cause extreme harms to individuals. Note, however, that these discussions relate to average or median outcomes.).

Therefore, while we show an increase in the number of events and legal actions, our estimates of firm costs do not reflect the same magnitude of consequence, or urgency of attention. An important point can therefore be made concerning optimal investment in security. Given these relatively low costs (i.e. again, not every breach is a "Target"), it may be the case that firms are, indeed, engaging in a privately optimal level of security – that they are properly and efficiently managing cyber risks as they do with other forms of corporate risk. And that for most firms, because their expected losses are relatively low, they subsequently are investing in only a modest amount of data protection.

In addition, other research based on consumer surveys shows that 77% of respondents are very satisfied with firm responses to data breaches, and that only a small percentage (11%) of customers are lost due to attrition (Ablon et al. 2016). Therefore, while the potential

for greater harm and losses appears to be increasing in time, evidence suggests that the actual financial impact to firms is considerably lower than expected. And so, if consumers are indeed mostly satisfied with firm responses from data breaches, and the costs from these events are relatively small, then firms may indeed lack a strong incentive to increase their investment in data security and privacy protection. If so, then voluntary adoption of the NIST Cybersecurity framework may prove very difficult and require additional motivation.

Therefore, where could the incentives originate? It is conceivable that the primary motivation may come from the cyber insurance industry through its use of incentive-based reductions in premiums (or deductibles). Indeed, with over 70 carriers offering cyber insurance policies (based on conversations between the author and Advisen representatives), and an estimated \$2 billion in US premiums (Romanosky 2015), insurance companies may already be driving a *de facto* national cyber security practice across insureds. But while insurance companies do have an incentive to drive security investments, there is, as of yet, no evidence showing that firms are actually improving their posture in response to cyber insurance policies.

Appendix

NAICS descriptions

Table A1. NAIC industry descriptions

Industry (NAICS Code); number of firms	Description and examples
Accommodation and Food Services (72); 495 347 Admin and Support Services (56); 327 214	Hotels, inns, food services, etc. Telemarketing, collection agencies, credit bureaus, travel agencies, armored car services, hazardous waste removal, etc.
Agriculture, Forestry, Fishing and Hunting (11); 21 351 Arts, Recreation and Entertainment (71); 114 969 Construction (23); 640 951 Educational Services (61); 84 503 Finance and Insurance (52); 234 841	For example, farming, orchards, cattle, chicken, logging, farm management, etc. Theatre, dance, sports, museums, casinos, amusement parks, golf, fitness centers. Residential, commercial construction, highway/street bridge construction, etc. Schools, universities, training and trade schools. Commercial banking, savings, credit card issuers, mortgage brokers, investment banking, insurance carriers and brokers.
Government (Public Administration) (92); 90 107	Courts, police, fire protection, administrative offices, national security, international affairs.
Health Care and social assistance (62); 640 ,724	Hospitals, dentists, doctors, medical centers, ambulance services, psychiatry and nursing, day care services.
Information (51); 71 108	For example, software vendors, telecommunications companies, internet portals, news/book publishing, motion picture and music publishing, radio, television, etc.
Management of Companies and Enterprises (55); 26 819	Management of companies and enterprises, offices of holding companies, corporate, subsidiary, and regional managing offices.
Manufacturing (31); 256 363	Food product manufacturing, breweries, wineries, fabrics/clothing, construction materials, newsprint/books, pharmaceuticals, plastics, rubber, iron, steel, computer and computing products, semiconductor, audio/video, truck, car manufacturing, medical equipment.
Mining, Quarrying, and Oil and Gas Extraction (21); 22 149 Other Services (81); 667 176	Mining, petroleum, gas extraction, drilling, and supporting activities. Automotive and computer repair, civic organizations, religious organizations, salons, and other personal services.
Prof, Sci, and Tech Services (54); 772 685	Legal, tax, engineering services, computer programming, management consulting, advertising and public relations companies, direct mail.
Real Estate and Rental and Leasing (53); 270 034 Retail Trade (44); 650 749 Transportation (and warehousing) (48); 168 057 Utilities (22); 5973	Residential and commercial property and equipment leasing, car and truck rental. Automotive, furniture, home centers, food markets, clothing, electronics. Air, train transportation, trucking, taxi, limousine, postal service. Power generation and distribution (solar, hydro, nuclear, wind), water and sewage treatment facilities.
Wholesale Trade (42); 315 031	Furniture, automotive, lumber, commercial and industrial equipment, farm machinery, etc.

Acknowledgements

I would like to thank James Anderson, Samantha Cherney, Matt Crespi, Lawrence Gordon, Jim Graves, Mark Greisiger, Marshall Kuypers, Martin Libicki, Martin Loeb, Trey Herr, Paul Heaton, Jamie Morikawa, Adam Shostack, Aaron Strong, and five anonymous insurance underwriters for their valuable comments and suggestions. This research also benefited from presentations and discussions at the American Association of Law Schools (New York, NY, 2016), Financial Information Systems and Cybersecurity (University of Maryland, 2016), FTC PrivacyCon (Washington DC, 2016), National Academies of Sciences (Washington, DC, 2016), Quantifying Cyber Risk, Center on Law and Security (NYU, New York, NY, 2016), Conference Board of Canada's Cyber Security Centre (Ottawa, 2016), Kogod Cybersecurity Governance Center (American University, Washington DC, 2016), DHS Science and Technology Cyber Security Division R&D Showcase (Washington DC, 2016), Quantifying Cyber Risk, Center on Law and Security (NYU, Brooklyn, NY, 2016), RAND Justice, Infrastructure and Environment board meeting (Santa Monica, CA, 2016), RAND Center for Corporate Ethics and Governance board meeting (New York, NY, 2016), Digital Information Policy Scholars Conference (George Mason University, Arlington, VA, 2016). I would also like to thank RAND's Institute of Civil Justice for its generous support.

Conflict of interest statement. None declared.

References

- Ablon, L., Heaton, P., Lavery, D *et al.* (2016) *Consumer Attitudes Toward Data Breach Notifications and Loss of Personal Information*. Santa Monica, CA: The RAND Corporation.
- Association of Certified Fraud Examiners (ACFE) (2012) Report to the Nations on Occupational Fraud & Abuse.
- Biener, C., Eling, M., Wirfs, J. H. (2015) Insurability of cyber risk: an empirical analysis. *The Geneva Papers*, 40: 131–58.
- BusinessWire (2015) Xerox Reports Second-Quarter 2015 Earnings. <http://www.businesswire.com/news/home/20150724005172/en/Xerox-Reports-Second-Quarter-2015-Earnings#.Vbk4yWfbK01> (1 August 2016, date last accessed).
- Comey, J. (2013) Confirmation Hearing of James Comey. <https://www.fbi.gov/news/podcasts/thisweek/james-comeys-confirmation-hearing.mp3/view> (1 August 2016, date last accessed).
- CSIS (2014) *Net Losses: Estimating the Global Cost of Cybercrime*. Center for Strategic and International Studies.
- Cybersource (2013) 2013 Online Fraud Report, Online Payment Fraud Trends, Merchant Practices, and Benchmarks.
- Edwards, B., Hofmeyr, S., Forrest, S. (2015) Hype and heavy tails: a closer look at data breaches. *Workshop on the Economics of Information Security*.
- Gordon, L.A., Loeb, M.P. (2002) The economics of information security investment. *ACM Transactions on Information and System Security*, 5: 438–57.
- Gordon, L.A., Loeb, M.P., Lucyshyn, W., Zhou, L. (2015) Increasing cybersecurity investments in private sector firms. *Journal of Cybersecurity*, 1: 3–17.
- Jacobs, J. (2014) Analyzing Ponemon Cost of Data Breach. <http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/> (1 August 2016, date last accessed).
- Kays, Joseph (2010) Business Expense. *Fall* 2010.
- MarketWatch (2015) Strayer Education, Inc. Reports Second Quarter Revenues and Earnings; and Summer Term 2015 Enrollments. <http://www.marketwatch.com/story/strayer-education-inc-reports-second-quarter-revenues-and-earnings-and-summer-term-2015-enrollments-2015-07-29> (1 August 2016, date last accessed).
- Moore, Rodney (2013) US hospitals provided \$41.1 billion in uncompensated care in 2011, representing 5.9 percent of annual hospital expenses. <http://www.healthcarefinancenews.com/news/stopping-rise-hospital-bad-debt-0> (1 August 2016, date last accessed)
- National Health Care Anti-Fraud Association (NHCAFA) (2010) Testimony of the National Health Care Anti-Fraud Association, to the House Insurance Committee, House of Representatives, Commonwealth of Pennsylvania.
- NetDiligence (2014) Cyber Claims Study. *NetDiligence*.
- Nilson Report (2013) Global Credit, Debit, and Prepaid Card Fraud Losses Reach \$11.27 Billion in 2012.
- Plotkin, R. (undated) Sticky Fingers – Bartender Theft in the New Economy. <http://www.barprofits.com/pages/newsletter/Vol00202-issue4/page06.php> (1 August 2016, date last accessed).
- Ponemon (2011) 2011 Cost of Data Breach Study: United States.
- Ponemon (2013) 2013 Cost of Data Breach Study: Global Analysis.
- Ponemon (2014) 2014 Cost of Data Breach Study: United States.
- Ponemon (2015) 2015 Cost of Data Breach Study: United States.
- Romanosky, S. (2015) The high cost of hacks: Is cyber insurance to blame for the increasing cost of data breaches? *US News*.
- Romanosky, S., Hoffman, D., Acquisti (2014) A empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, 11: 74–104.
- Romanosky, S., Telang, R., Acquisti, A. (2011) Do data breach disclosure laws reduce identity theft? *Journal of Policy Analysis and Management*, 30: 256–86.
- Rowe, Megan (2011) 7 Ways to Stem Bar Shrinkage, Restaurant-hospitality.com. <http://restaurant-hospitality.com/trends/seven-ways-stem-bar-shrinkage0911> (1 August 2016, date last accessed).
- White House (2015) FACT SHEET: Cyber Threat Intelligence Integration Center. The White House, Office of the Press Secretary. <https://www.whitehouse.gov/the-press-office/2015/02/25/fact-sheet-cyber-threat-intelligence-integration-center> (1 August 2016, date last accessed).