2020

Cyber-Assets at Risk (CAR): The Cost of Personally Identifiable Information Data Breaches

Omer Ilker Poyraz
Old Dominion University

Sarah Bouazzaoui
Old Dominion University

Omer Keskin
Old Dominion University

Michael McShane
Old Dominion University

Ariel Pinto
Old Dominion University

Follow this and additional works at: https://digitalcommons.odu.edu/emse_fac_pubs

Part of the Corporate Finance Commons, and the Systems Engineering Commons

Original Publication Citation

This Conference Paper is brought to you for free and open access by the Engineering Management & Systems Engineering at ODU Digital Commons. It has been accepted for inclusion in Engineering Management & Systems Engineering Faculty Publications by an authorized administrator of ODU Digital Commons. For more information, please contact digitalcommons@odu.edu.
Cyber-assets at Risk (CAR): The Cost of Personally Identifiable Information Data Breaches
Omer Ilker Poyraz\textsuperscript{1}, Sarah Bouazzaoui\textsuperscript{1}, Omer Keskin\textsuperscript{1}, Michael McShane\textsuperscript{2}, and Ariel Pinto\textsuperscript{1}
\textsuperscript{1}Dept. of Engineering Management & Systems Engineering, Old Dominion University, Norfolk, USA
\textsuperscript{2}Dept. of Finance, Old Dominion University, Norfolk, USA
opoyraz@odu.edu
sboua003@odu.edu
okesk001@odu.edu
mmcshane@odu.edu
cpinto@odu.edu

Abstract: Severe financial consequences of data breaches enforce organizations to reconsider their cybersecurity investment. Although attack frequency and trends seem similar per industry, the impact of a data breach may exponentially increase depending on the type of information and the amount of the stolen data. Also, governments develop and improve laws and regulations to protect the privacy of individuals. Therefore, a failure of data security may yield severe penalties and class-action lawsuits, which can significantly increase the expenses than before. The monetary impact of a data breach is a new field of study that requires more sophisticated research and analysis. There are very few studies that quantify the monetary value of data breaches, which are based on the number of affected people or the number of stolen records. This study proposes a new methodology to quantify the monetary value of the data breaches by categorizing information as personally identifiable information (PII) and sensitive personally identifiable information (SPII). Our findings indicate that the categorization of the stolen information has more relation than solely the number of affected people or the number of stolen records. Also, SPII data breaches have more class-action lawsuits, which yield higher costs than PII data breaches.

Keywords: data breach, cyber risk, the economics of cybersecurity, cyber insurance

1. Introduction

Data breach incidents have been fast becoming a vital instrument in cybersecurity risk assessment. Security of the data plays an essential role in keeping the reputation of the company as well as avoiding financial fees or litigations. A primary concern of data breach is severe financial consequences.

Quantifying the data breach risk into a financial value is a phenomenon that insurers or risk managers still try to solve due to the unavailability of the data and latent costs. Data breach costs may exceed hundreds of million dollars, which can severely jeopardize an organization’s financial health. Therefore, understanding the data breach and its impact on business will guide the decision-makers to invest smarter in cybersecurity as well as considering non-technical options such as cyber-insurance. Also, the insurance industry needs to understand the likelihood and impact of a data breach to underwrite and sell cyber-insurance. This study will bring a new methodology for data breach impact assessment by categorizing the information as sensitive and not-sensitive for decision-makers and insurers.

2. Literature Review

Cyber-crime is not new. The first cyber-crime occurred in 1973 in a New York Bank that a teller embezzled $2 million by using a computer (Wavefront, n.d.). Recent developments in cyber technologies have led to upgrading the cyberattack techniques and methods in cyber-crime. However, the study of a data breach is a new field of cyber-crime that has an increasing monetary impact.

There are plenty of studies to determine the optimal amount of cybersecurity investment, cyber-risk score, or monetary impact of availability compromise (Arora et al., 2004; Bahşi et al., 2018; Cavusoglu et al., 2008; Fielder et al., 2016; Garvey et al., 2013; Gordon and Loeb, 2002; Karabacak and Sogukpinar, 2005; Keskin et al., 2018; Lam, 2015; Shetty et al., 2018; Tatar et al., 2016). Also, there are few studies to understand the trends, frequency, and distribution of the data breach incidents (Edwards et al., 2016; Eling and Loperfido, 2017; Wheatley et al., 2019).

However, there are very few studies to understand the monetary impact of data breaches (Jacobs, 2014; Layton and Watters, 2014; Romanosky, 2016; Tatar, 2019). The relevant studies are summarized below.
Romanosky (2016) used the Advisen dataset to develop a regression model to calculate the monetary impact of a security breach. The proposed model finds revenue, and the number of records has a correlation with the cost. He says that a 10% increase in revenue would raise the impact of 1.3%. Also, according to his model, a 10% increase in the number of records compromised would escalate the cost of 2.9%.

Another regression model is proposed by employing the Cost of Data Breach dataset (Jacobs, 2014). The author claims that a 10% increase in the lost number of records begets a 7.6% increase in the cost of the data breach. Layton and Watters (2014) estimated the tangible costs of data breaches using two case studies, based on a salary guide and rough estimation of the work hours of the people who were dealing with the data breach. Whereas only tangible cost is labor cost, loss of reputation is regarded as an only intangible cost. Tatar (2019) developed a mathematical model regarding dependency and inter-dependency, and confidentiality, integrity, and availability of the assets.

3. Personally Identifiable Information (PII)

This study will focus on personally identifiable information (PII) data breaches. The Department of Homeland Security (DHS) defines the PII as (DHS, 2017): “any information that permits the identity of an individual to be directly or indirectly inferred, including any other information that is linked or linkable to that individual, regardless of whether the individual is a U.S. citizen, legal permanent resident, visitor to the U.S., or employee or contractor to the Department.”

Sensitive PII (SPII) is defined as “personally identifiable information which, if lost, compromised or disclosed without authorization, could result in substantial harm, embarrassment, inconvenience or unfairness to an individual.” (DHS, 2017).

Tables 1 and 2 is provided to elaborate PII and SPII (DHS, 2017; STIP, 2018; WDPI, n.d.):

**Table 1: Examples of PII**

<table>
<thead>
<tr>
<th>PII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
</tr>
<tr>
<td>Account name/ user ID</td>
</tr>
<tr>
<td>Password</td>
</tr>
<tr>
<td>Email</td>
</tr>
<tr>
<td>Address</td>
</tr>
<tr>
<td>Telephone number</td>
</tr>
<tr>
<td>Education credentials/certificates</td>
</tr>
<tr>
<td>Date/place of birth</td>
</tr>
<tr>
<td>Vehicle title number</td>
</tr>
</tbody>
</table>

**Table 2. Examples of SPII**

<table>
<thead>
<tr>
<th>SPII</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social security numbers</td>
</tr>
<tr>
<td>Medical history</td>
</tr>
<tr>
<td>Credit/ debit card numbers</td>
</tr>
<tr>
<td>Driver’s license numbers</td>
</tr>
<tr>
<td>Bank account numbers</td>
</tr>
<tr>
<td>Passport numbers</td>
</tr>
<tr>
<td>Alien registration numbers</td>
</tr>
<tr>
<td>Biometric identifiers</td>
</tr>
<tr>
<td>Taxpayer identification number</td>
</tr>
</tbody>
</table>
4. Why is the information stolen?

A PII data breach may be a result of intentional or unintentional disclosure of data. Finance and healthcare industries are the primary targets of malicious actors. Credit cards and medical information are the most stolen data from organizations (Romanosky, 2016). Massive data breaches in recent years attracted government and media attention to those events such as Home Depot, Target, Anthem, Equifax, Capital One cases. Table 3 shows the five most notorious mega data breaches and their cost to the victim companies.

Table 3: Mega Data Breaches

<table>
<thead>
<tr>
<th>Company</th>
<th>Number of affected people</th>
<th>Cost of Data Breach ($ in millions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anthem</td>
<td>78 million</td>
<td>406.5</td>
</tr>
<tr>
<td>Equifax</td>
<td>147 million</td>
<td>1445</td>
</tr>
<tr>
<td>Home Depot</td>
<td>56 million</td>
<td>340</td>
</tr>
<tr>
<td>Sony PSN</td>
<td>77 million</td>
<td>193</td>
</tr>
<tr>
<td>Yahoo</td>
<td>3 billion</td>
<td>502</td>
</tr>
</tbody>
</table>

Data breach attacks may be carried out by individual hackers, insiders, hacktivists, organized-crime groups, or states. If they steal an intellectual property product or business secret or a blueprint, it gives a significant benefit to the attackers. However, what is the benefit of stealing PII?

Hackers hack for money, fun, challenge, and “to do good in the world” (Tatar and Celik, 2015). Malicious actors search for PII or SPII to make money by duplicating credit cards, printing passports or I.D.s, blackmail, espionage, or profiling purposes for a foreign government (Tatar et al., 2016) (Fingas, 2017; Verizon, 2019). The selling price of the information in the dark web varies depending on the sensitivity of the information; for example, passports or diplomas are more expensive (Stack, 2019). The stolen data may be sold in the dark web by identity thieves as (Stack, 2019):

- Sell the data as a one-off, such as credit card numbers
- Sell bulk data, similar types of information
- Sell bundled data, different types of information are bundled together

5. Why are data breaches significant for governments?

According to the Department of Defense (DoD), a breach of personal information is defined as “the act of accessing by or disclosing information to unauthorized individuals or compromising in a way where the subjects of the information are negatively affected” (DOD Privacy Program, 2014). Information in this context refers to any personal information that can be used to identify or communicate with the subject matter (Johnson, 2007). With technological advancements, this information is easier to collect (Meingast et al., 2006), and the potential misuse of individuals' data in deceptive acts raises an array of unique ethical issues and consequences to both individuals and organizations. Individuals may suffer from embarrassment or blackmail, while organizations may face a lack of public trust, a decrease in reputation, and legal liabilities.

Implementing laws and regulations governing the protection of PII became a necessity at different levels, the federal level, the state level, and international level. This is a list of the most important federal laws and directives related to PII:

- Health Insurance Portability and Accountability Act of 1996 (HIPAA)
- The Gramm-Leach-Bliley Act of 1999 (GLBA)
- Social Security Number Protection Act of 2005
- Identity Theft Prevention Act of 2005
- Privacy Act of 1974
- The E-Government Act of 2002
- The Confidential Information Protection and Statistical Efficiency Act (CIPSEA)
- The Family Educational Rights and Privacy Act (FERPA) – relating to PII protections for student educational records
- The Health Information Technology for Economic and Clinical Health Act (HITECH)
- The Fair Credit Reporting Act (FCRA) – regulating how consumer reporting agencies use credit information
• The Children’s Online Privacy Protection Act (COPPA) – relating to the privacy of children under 1

At the state level, Security breach laws notification related to PII are adopted by all states defining PII and have provisions on the entities that must respect it. The notification law complies with federal laws as well as international laws.

In the next sections, the laws and regulations related to the health and financial sector will be explained. These laws are applied within strict guidelines with the intent to protect the confidentiality and integrity of individuals’ information.

6. Healthcare regulations and laws

In the healthcare sector, there are many regulations that address the collection, use, and disclosure of PII. The Health Insurance Portability and Accountability Act (HIPAA) is the most important one. It was put into effect in 2003 to protect the confidentiality of individuals. This act defines protected health information as any information that can relate to the health condition and the health care of individuals (Al-Fedaghi, 2008). It refers to the information that is transmitted or kept by electronic media or any other medium (under electronic protected health information).

The HIPAA is considered an extension of the privacy act of 1974. The Privacy Act of 1974 protects all information maintained in any system of records, including the healthcare system. Information refers to “any information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph” (Marine corps division instruction 5211.1, 2012). Furthermore, if the service is related to the healthcare and the government service as well, it is posted by the E-Government Act also directed OMB to provide privacy guidance to Federal agencies on many PII related issues such as remote access to PII, and encryption of PII on mobile devices.

In case this act is not kept, criminal and civil sanctions are available. Tables 4 and 5 summarize both (Edemekong et al., 2019)

Table 4: Civil Sanctions

<table>
<thead>
<tr>
<th>Civil</th>
<th>Due to failure to acknowledgment, $100 fine per violation with an annual maximum of $25,000 for those who repeat violation. There is also $50,000 per violation and an annual maximum of $1.5 million.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Due to reasonable cause and not due to willful neglect: There is $1000 charge per violation, an annual maximum of $100,000 for those who repeatedly violate. There is also a $50,000 penalty per violation and an annual maximum of $1.5 million.</td>
</tr>
<tr>
<td></td>
<td>Due to willful neglect, with violation corrected within the required time. There is a $10,000 penalty per violation, an annual maximum of $250,000 for repeat violations. There is a $50,000 penalty per violation with an annual maximum of $1.5 million.</td>
</tr>
<tr>
<td></td>
<td>Due to willful neglect and not corrected. There is a penalty of $50,000 per violation, an annual maximum of $1,000,000, $50,000 per violation, and an annual maximum of $1.5 million.</td>
</tr>
</tbody>
</table>

Table 5. Criminal Sanctions

<table>
<thead>
<tr>
<th>Criminal</th>
<th>For entities that are covered and specified individuals who obtain or disclose individually identifiable health information willfully and knowingly: The penalty is up to $50,000 and imprisonment for up to 1 year.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>For offenses committed under pretenses, the penalty is up to $100,000 with imprisonment for up to 5 years.</td>
</tr>
<tr>
<td></td>
<td>For offenses committed with the intent to sell, transfer, or use individually identifiable</td>
</tr>
</tbody>
</table>
health information for commercial advantage, personal gain, or malicious harm, the penalty is up to $250,000 with imprisonment up to 10 years.

7. Finance regulations and laws

In finance, individual information is governed under the Gramm-Leach-Bliley Act of 1999 (GLBA). The GLBA is the first federal law adopted, it outlines guidelines to protect consumers' data privacy from financial institutions and credit card companies, and it clarifies the responsibilities of financial institutes towards the privacy of their clients' information (Akighbe & Whyte, 2004). If this act is violated, the company will be fined $100,000 for each violation and imprisonment for up to five years. The GLBA complies with the consumer protection act of 1968, specifically the right to safety. Furthermore, the right to financial privacy act is an essential act of 1978. Under the Financial Privacy Act, government officials can only get financial records to an individual if they get written consent or obtain a subpoena.

8. International data breach and regulations

At the international level, specifically in Europe, personal data is protected by a data breach regulation-General Data Protection Regulation (GDPR). This regulation does not specify the inclusion of PII, but understandably PII is part of personal data. The main principal covered is that GDPR is data protection (Goddard, 2017) of E.U. citizens from any organization collecting their data, disregarding the company location (Tankar, 2016).

Failure to comply with this regulation, companies may encounter penalties ranging from “€10 million or 2% of the company’s global annual turnover of the previous financial year to €20 million or 4% of the company’s global annual turnover of the previous financial year, whichever is higher” (Tobin et al, 2017).

9. Limitations

Due to the lack of publicly available data, this study has the limitations below:

- The data breaches include cases where the number of affected people is more than one million
- The study only covers organizations for profit
- The sample size is small due to the lack of available information about the data breach incidents

10. Methodology

10.1 Data Collection

The Privacy Rights Clearinghouse (PRC) data breach database is used to determine the mega data breaches as of November 4, 2018. Only for-profit organizations are considered. As a result, 133 distinct mega data breaches occurred from 2004 to 2018, with the addition of the AOL 2004 data breach. Next, each case is searched for the following parameters:

- Cost type and amount of data breach
- Number and type of stolen data:
  - Total number of PII
  - Total number of SPII

As a result, 31 cases have answers to the parameters above. This dataset mostly focuses on the impact of the data breach on the U.S. and E.U. citizens due to well-defined laws, regulations, and the existence of agencies to monitor cyber-crimes.

10.2 Model

Exploratory data analysis is performed to identify the patterns, factors for similarities in mega data breaches. Next, regression models performed to determine the linear relation of predictors to the response variables.

Simple linear regression is applied to detect the linear relation of the number of people with the cost of the data breach. The first model includes the intercept and the number of people. The 2nd model includes
intercept, PII, and SPII. Cost is the response variable in both regression models. Both models' intercept is statistically non-significant; therefore, the models are run without the intercept in both models. Independent variables are statistically significant in both regression models. The output is illustrated in Table 6.

Table 6: Regression Model Outputs

<table>
<thead>
<tr>
<th></th>
<th>R²</th>
<th>Adjusted R²</th>
<th>1st Coefficient</th>
<th>2nd Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of people</td>
<td>0.367</td>
<td>0.347</td>
<td>1.8245</td>
<td>None</td>
</tr>
<tr>
<td>PII and SPII</td>
<td>0.75</td>
<td>0.733</td>
<td>0.3860</td>
<td>4.6377</td>
</tr>
</tbody>
</table>

1st model regression equation:  
Cost = 1.8245 * (number of people)

2nd model regression equation:  
Cost = 0.3860 * PII + 4.6377 * SPII

10.3 Results

We can explain 36.7% of the change in cost with the number of affected people. If the number of affected people is 1,000,000, the cost may be between $931,000 and $2,718,000 with a 95% confidence interval. However, the second model explains the change in the cost much better than the first model with a 75% R². One million PII and SPII loss costs may range between $3,362,000 and $6,686,000, with a 95% confidence interval. Unlike the earlier regression models based on the number of people or number of records (Jacobs, 2014; Romanosky, 2016), the model with the information categorization -PII and SPII- explains the variation in a cost better.

The data breach cost ranges from $0.7 million to $1.45 billion for mega data breaches. The average cost of a mega data breach is $172,645,200, while the median is $84,000,000. Figures 1-4 represent class-action lawsuits, industries, and the cost of data breaches. There are 19 class-action lawsuits among 31 mega data breaches. Data breaches having class-action lawsuits have more cost than dismissed cases on average. Also, the trend for mega-breaches seems similar per industry, except 2015 is the worst year for medical organizations.

Figure 1. Data Breach per Industry

Figure 2. Cost of the data breach per industry
Figures 5 and 6 show that SPII data breaches have more impact on costs and caused more class-action lawsuits than the PII data breaches. Blue dots display the approved class-action lawsuits for PII and SPII data breaches. They have more cost than the dismissed cases, which are red dots. However, while some data breaches had one type of information breach - either PII or SPII, some of them had both PII and SPII data breaches. It is observed that data breaches have SPII led to more class-action lawsuits and more cost. Class-action lawsuits are more likely to be accepted if there is proof of fraud cases. If there is no proof of fraud cases, judges may dismiss the case until evidence of multiple fraud cases.

Figure 3. Class-action lawsuit per industry

Figure 4. Data breach cost ($ in millions) vs. class-action lawsuits

Figure 5. Relationship between SPII, class-action lawsuits and cost ($ in millions)
Every data breach is not the same. The cases that have SPII loss have more cost comparing to non-SPII data breaches. Although there are different regulations, the type of information determines litigations, fees, or professional/legal expenses. According to our data, the sensitivity and the amount of the data are more critical to determine the cost disregarding the type of industry.

The healthcare regulations mostly focus on the security of medical information, which is only held in the healthcare sector. In contrast, financial information related regulations mostly focus type of information disregarding the industry. For example, all online retailers are responsible for the security of the data, such as credit/debit card numbers, social security numbers, and so on.

The type of cost incurred after a mega data breach may include remediation, investigation, increase in cybersecurity budget, data breach settlement, legal expenses, customer notification, credit monitoring against identity theft, or canceled business deals.

11. Conclusions

The categorization of the information as PII and SPII will provide a more accurate forecasting data breach impact compared to earlier works (Jacobs, 2014; Layton and Watters, 2014; Ponemon, 2019; Romanosky, 2016; Verizon, 2019). Categorization of the information as PII and SPII can explain the variation in cost much better than the number of affected people or the number of stolen records. Also, every type of record does not have equal value as our model coefficients show that SPII's Coefficient is twelve times greater than PII's Coefficient. The U.S. courts approve web-scraping without permission is not illegal (Mehta, 2019). Since many people willingly share their PII, such as name, job, photo, etc. on websites, PII data breaches may have less value compared to SPII data breaches. Therefore, the classification of the information is necessary to forecast the impact more accurately.

The monetary impact of PII data breach compels risk managers, chief information officers, or cyber-insurance firms to reconsider the data breach risk. This study provides a novel approach for data breach impact assessment by categorizing personal information as sensitive and not sensitive. Therefore, future studies may include forecasting the monetary value of intellectual property breaches such as movies, songs, or books.

References


WDPI (n.d.) DPI Personally Identifiable Information.