Difficulties in Determining Data Breach Impacts

John W. COFFEY
Department of Computer Science
The University of West Florida
Pensacola, FL. 32514, USA
jcoffey@uwf.edu

ABSTRACT

Internet searches for information regarding the impacts of data breaches yield a large amount of data. While significant effort is made to determine the costs associated with data breaches, important questions can be asked regarding the accuracy of such reporting. This article examines some of the drivers of difficulty in determining the impacts of data breaches, both from the perspective of the organization that is breached, and, possibly more importantly, from the viewpoint of individuals whose data is breached. This article will make the case that, from the start of the process in forensic analysis, difficulties can be detected. Separately, dis-incentives to report more than required by law is another impediment. In the United States, no uniform reporting requirement exists. Ultimate impacts on the individuals whose data is breached are often delayed, based upon incomplete self-reports, and difficult to summarize. This article will make the case that all these factors negatively impact the quality of data breach reporting.

1. INTRODUCTION

A broad view of all the issues surrounding data breaches presents a complex, interconnected picture. Multiple constituencies are impacted by data breaches including the business, governmental organization, or non-governmental organization that has data beached, and individuals who are impacted by breaches of these organizations. Despite many efforts by a variety of businesses, governmental organizations and NGOs to do so, it remains difficult to assess the overall impacts of data breaches to individuals and organizations.

The ultimate sufferers of impacts stemming from data breaches are the individuals whose data was breached. Breached organizations immediately go into damage control mode following detection of a breach. Organizations have difficulties assessing the extent and impacts of breaches and they have strong dis-incentives to disclose more about the breach than mandated by widely varying laws. For this reason, the individuals who are ultimately affected by having their data breached routinely have far less than optimal information regarding potential impacts.

The remainder of this article will make the case for why it is difficult both for organizations and individuals to know the ultimate impacts of data breaches. This discussion will be couched in the context of cybersecurity and digital forensics as so-called “wicked problems.” A variety of factors will be explored including those that start at the source of the problem: difficult aspects of determining the extent of damage done by a breach. Also, highly variable state-level reporting requirements in the United States that are more geared to impacts on the firm contrast with consumer-oriented reporting requirements mandated in the European Union.

Some cost estimates for the firm presented by IBM [1] are contrasted with costs to the individual [2]. Limitations of such studies are discussed and further understandings of difficulties in quantifying impacts are presented. The paper will conclude with a brief mention of the field of anti-forensics, which enable attackers to obfuscate the nature of their intrusions, a consideration of protective actions individuals can take to protect themselves proactively, and some conclusions.

2. THE CYBERSECURITY RISK AND REPORTING PROBLEM

This section contains a discussion of so-called “wicked problems” in order to create a context from which to view such problems. It also contains a discussion of how complex and convoluted the forensics and reporting environment for data breaches is currently.

2.1 Wicked Problems

The difficulties in preventing and assessing data breaches exhibit many of the characteristics of so-called “wicked problems” [3]. The overall landscape is inherently complex and multi-faceted. For instance, data breaches can occur because of successful hack attacks of many different types, through unintended disclosure through regular or email communications, or even from someone inadvertently throwing away old hardcopy records with sensitive information.

The basic nature of the problem changes over time. Early phishing attacks started with emails asking for help getting millions out of a foreign country for only a few thousand up-front dollars. Over the decades since the start of the Internet, phishing attacks have evolved into highly sophisticated endeavors involving social engineering – convincing targets that the attackers are legitimate contacts making legitimate requests for information.

The problem is never solved definitively. Year-in and year-out security features in systems continue to evolve and so do the means employed to try to break into systems or to convince users to do something stupid. Wicked problems have no stopping rule – they are ongoing in a changing environment. Organizations that are successfully attacked have to devote resources to the response that would much more productively be used for something else.

With wicked problems, various parties to the problem have different and potentially competing incentives with regard to their resolution. Firms want to minimize damage to their reputations and in the case of businesses, to their bottom lines. Damage minimization involves minimizing negative publicity. Minimizing reporting is a basic means of damage control. On the other hand, individuals want to know that their sensitive information is safe and they want to know as much as possible about the nature of an attack and its potential implications. Consequently, the interests of organizations and those of individuals are fundamentally opposed.

2.2. The Complex Picture of Data Disclosure Risk

The overall scope of the data breach risk landscape is extensive. The huge volume of sensitive data in the world includes many different types of information stored in many electronic or hardcopy formats. A data breach may disclose personally identifiable information, financial records, healthcare records, payment card data, education data, information about peoples' credentials, or a variety of other types of data. It is often the case that an unknown amount of data is disclosed.

The means through which the data breach occurs is similarly various. The data may be exfiltrated through a successful hack attack or the installation of malware, through payment card fraud, disclosures made by insiders with or without malicious intent, by loss or theft of hardcopy data, or through unknown causes. The various types of data that might be disclosed interacts with the many different ways that it might be disclosed to create a truly complex landscape.

3. DATA BREACH REPORTING IN THE UNITED STATES

3.1 Overview

The National Council of State Legislatures (NCSL) [4] publishes state reporting guidelines for data breaches in the United States. Reporting requirements are specified at the state rather than federal level. It is only as of 2018 that all 50 states have reporting requirements, with Alabama and South Dakota finally adding reporting requirements in 2018.

Reporting laws typically specify a taxonomy of organizations and reporting requirements by organization type, specific definitions of what constitutes personal information, what type of event meets criteria as a reportable data breach, the type of notice that must be given, and any exemptions from reporting. A typical reporting policy is long, technical, and contains significant legalese.

3.2 Privacy Rights Clearinghouse Reporting

The Privacy Rights Clearinghouse [5] was founded in 1992 at the Center for Public Interest Law at the University of San Diego School of Law. PRC seeks to identify and bring attention to critical privacy-related issues. PRC was the first consumer organization in the nation to raise awareness of the concept of identity theft and provide assistance to victims. PRC has worked for passage of several landmark laws including:

- data breach notice law
- security freeze law

PRC participates in state and federal public policy task forces pertaining to federal privacy legislation and administrative agency proceedings.

PRC has one of the most comprehensive databases on data breaches, reporting 11,580,000,000 records breached in 9,094 documented events since 2005. In many cases, the number of records breached is estimated; often the number is unknown. Of the 984 cases of unintentional disclosure, 279 cases (28%) involved an unknown number of records.

For instance, a significant percentage of cases involving medical records reveal that an unknown number of records were breached. In many cases, no particulars regarding how the breach occurred are available. In some cases, the total number of records breached is reported but without details regarding the amount of unique sensitive information. Sometimes the narratives provide ranges such as “between 5,600 and 23,000 patients were affected.”

4. COSTS FROM THE ORGANIZATION'S PERSPECTIVE

IBM and Ponemon Institute have released their 2017 "Cost of a Data Breach" Study [1]. The study was based upon 419 organizations across 13 countries or regional samples. The report was based upon interviews with more than 2,200 individuals who are knowledgeable about the data breach incidents in their organizations.

The Ponemon report provides a variety of measures including the number of customer records lost or stolen in the breach and the percentage of the customer base that was lost following the data breach. Additionally, the report describes the amount the organization spent upon discovery of the breach for forensics and investigations, and activities conducted in the aftermath of discovery, such as the notification of victims and legal fees.

The researchers determined a variety of alarming statistics including that a typical breach entailed a total cost of more than $3.6 million, and that there was almost a 30% chance that once breached, an organization would experience another breach within two years. Both total cost and probability of another breach have increased consistently in recent years.

IBM enumerated several limitations of the study. A representative but non-statistical sample of global entities was analyzed. Therefore, statistical inferences, margins of error, and confidence intervals could not be applied to the data. Limitations also included the fact that the current findings were based on a small (n = 419) sample of companies. Since non-response bias could not be tested, it is possible that companies that did not participate had substantially different data breach costs than the ones in the study.

The report authors also suggested that the study might be biased toward companies with more mature privacy or information security programs. Ponemon Institute is known for careful work, but these limitations are indicative of inherent problems in trying to perform such analyses.

Abril [6] enumerates the largest corporate data breaches (with the number of user records breached) of 2018. She notes that very large corporations with massive quantities of customer data were involved in large breaches in 2018, and remain vulnerable. The following is the list she culled:

- Marriott International (500 million users)
- Twitter (330 million users)
- My Fitness Pal (150 million users)
- Facebook (147 million users in multiple breaches)
- Firebase (100 million users)
- Quora (100 million users)
- MyHeritage (92 million users)
- Uber (57 million users)
- Ticket Fly, owned by Eventbrite (27 million users)
- Google+ (500,000 users)
- British Airways (380,000 users)

The Marriot International breach is illustrative. According to reports from the Federal Trade Commission [7] customers' names, addresses, phone numbers, email addresses, passport numbers, birthdays, gender, and payment card numbers were exfiltrated. External attackers were able to exploit vulnerabilities in the reservation system the company had acquired in 2014, and the breach was ongoing for years. Complicating the assessment of damage was the fact that, although the credit card numbers were encrypted, it is not known if the information needed to decrypt them was also stolen. Other breaches from the year all have their own twists on deficiencies in operations and problematic aspects in determining just how much damage was done.

5. COSTS FROM THE INDIVIDUALS' PERSPECTIVE

Systematic attempts to determine costs of data breaches in the United States will generally lead to the conclusion that, unlike Europe with the GDPR [8] in place, costs to the firm are more studied than costs to the individual whose data was breached. The following section enumerates some available data on costs to the individual.

5.1 Consumer Impacts: Identity Theft Statistics

The following is from the Consumer Sentinel Network (CSN) [2] that is maintained by the Federal Trade Commission (FTC). Its purpose is to track consumer fraud and identity theft complaints that have been filed with federal, state and local law enforcement agencies and private organizations. According to CSN, in 2017, 6.64 percent of consumers became victims of identity fraud, or about 1 in 15 people. That equals 16.7 million victims, an increase of 1 million from 2016. Over 1 million children in the U.S. were victims of identity theft in 2017, costing families $540 million in out-of-pocket expenses. Identity theft is one of the most common consequences of data breaches, as 31.7 percent of breach victims experienced ID theft. It is estimated that there is a new victim of identity theft every 2 seconds. It takes most victims of identity theft 3 months to discover that their identities have been used improperly, but it is estimated that 16 percent don't find out for 3 years.

Total known costs to consumers were estimated to be $905 million. The median amount consumers paid in these cases was $429. 00. The fraud category, imposter scams accounted for $328 million in losses. In 2017, 14 percent of all complaints were related to identity theft that was a direct consequence of unintended data disclosures or exfiltration of sensitive personal information. Identity theft complaints increased almost 70 percent from 2013 to 2015. However, they fell about 24 percent from 2015 to 2017. Credit card fraud was the most reported incident to the CSN, with 133,000 reports.

5.2 Consumer Impacts: The Equifax Breach

Equifax is one of three major Credit Reporting Agencies in the United States with data on virtually every adult in the U.S. An application vulnerability in one of their websites exposed data on what was initially estimated to be 143 million consumers [9]. The breach was estimated to have occurred in mid-May, 2017, but it was only discovered on July 29, 2017. In October 2017, Equifax raised its estimate of the number of impacted consumers to 145.5 million.

On March 1, 2018, the company raised the number by another 2.4 million, bringing the tally to 147.9 million. The United States Federal Trade Commission disclosed that Equifax has agreed to a global settlement with the FTC, the Consumer Financial Protection Bureau, and 50 U.S. states and territories. The settlement included up to $425 million for individuals who incurred damages as a result of the breach [10].

6. ANTI-FORENSICS

Anti-forensics has been defined as "any attempt to compromise the availability or usefulness of evidence to the forensics process" [11]. Digital anti-forensics are the specialized field of forensics applied to computer systems, and they have their own unique characteristics. There are four basic categories of digital anti-forensics approaches:

- data hiding
- artefact wiping
- trail obfuscation
- attacks against both the forensic process and forensic tools.

Data hiding involves making data difficult to find. Data obfuscation and encryption limits the ability of forensics personnel to identify salient information. For example, a program named *Slacker* breaks up a file and places each piece of that file into the slack space of other files. Artefact wiping might involve overwriting information that is critical to an investigation, deleting log files, or writing random characters on a disk until it is full. Overwriting metadata can make construction of breach timelines difficult or impossible. Data falsification is a typical form of trail obfuscation. MACE (Modified-Accessed-Created-Entry) modification is a common obfuscation technique employed by attackers. The program *TimeStomp* is typical of programs that can be used for MACE attacks.

Malicious insiders make it much easier to carry out anti-forensics efforts. Malicious insiders can run malicious programs from CDs, bootable jump drives and virtual machines. Doing so can enable attackers to execute programs without permanently installing them on the target machine. Digital forensics specialists tend both to utilize similar computer forensics software tools (CFTs) and to employ similar workflows. Because of predictability in the tools and processes digital forensics specialists use, a range of attacks against both tools and process are possible.

While some attackers want their exploits to be known, more commonly, attackers would rather get into systems to extort money through ransomware attacks, or to stay inside without detection. Anti-forensics tools and procedures enable them to pursue these goals, and ultimately make it that much more difficult to determine the extent and implications of successful attacks.

7. INDIVIDUAL PROTECTIONS

Given that the factors enumerated here mediate against individuals knowing the full impact of having their sensitive data breached, arguably, the best means of protection are those that can be taken proactively. Snider [12] enumerated a variety of means of self-defense. Users must use unique passwords for different accounts. Password managers make keeping track of passwords relatively easier. Also, two-factor authentication, while requiring an additional step before access is allowed to the user's accounts is helpful, entailing a "defensive in depth" approach.

It is always necessary to scrutinize email carefully. It is estimated that fully 7 out of 10 cyberattacks (71 percent) start with a phishing email. Evidence indicates that people are still willing to click on either links or attachments in emails without a lot of thought. It is also necessary to keep software up to date.

Applying software updates on devices as they become available means one has done all that is possible to prevent known attacks. This policy does nothing to help with zero-day attacks that exploit vulnerabilities that are currently not known by the software developers.

Credit freezes and other measures make it more difficult for identity thieves to open accounts with a stolen identity. Since the Equifax breach, credit freeze/unfreeze transactions are free. Finally, limiting the personal information that is given out is important. Individuals should not give out emails without a good reason. "Tracker Blocker" software can be installed to protect against malware and ransomware delivered through online advertisements.

8. CONCLUSIONS

Data breach severity determination and reporting meet the criteria of a "wicked problem." The problem is multi-faceted and changing over time, it presents no definitive stopping state or rule, and different constituencies have different incentives and goals relative to the problem. A variety of factors work against the goal of clear understanding of the impacts of data breaches. Anti-forensic measures potentially make accurate, comprehensive assessment of a breach difficult. Even with an accurate assessment of the damage that has been done, firms have strong incentives to disclose as little as possible about data breaches, as there are significant financial costs and perhaps even more importantly, the loss of trust of customers. Reporting laws are a patchwork in the United States. Compliance is challenging given the significant amount of legalise in reporting laws. Impacts on consumers are often delayed and difficult to detect until significant damage has been done. In the face of all these difficulties, proactive measures are critical to afford the best possible protections to individuals.

9. REFERENCES

[1] Ponemon Institute. 2017**. Ponemon Institute Cost of a Data Breach Study.** Online, Available: https://securityintelligence.com/media/2017-ponemon-institute-cost-of-a-data-breach-study/

[2] FTC. **Consumer Sentinel Network.** Online, Available: https://www.ftc.gov/enforcement/consumer-sentinel-network.

[3] Camillus, J.C. **Strategy as a Wicked Problem**. Harvard Business Review. 86: 98–101.

[4] NCSL. **Security Breach Notification Laws.** Online, Available: https://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

[5] PRC. **Data Breaches.** Online, Available: https://privacyrights.org/categories/data-breaches

[6] Abril, D. **These Were the Worst Data Breaches and Vulnerabilities of 2018.** Online, Available: http://fortune.com/2019/01/04/worst-data-breaches-of-2018/

[7] Gressin, S. **The Marriot Data Breach.** Online, Available: https://www.consumer.ftc.gov/blog/2018/12/marriott-data-breach

[8] GDPR. **GDPR Portal.** Online, Available: https://www.eugdpr.org/

[9] Ragan, S. **Equifax says website vulnerability exposed 143 million US consumers.** Online, Available: https://www.csoonline.com/article/3223229/security/equifax-says-website-vulnerability-exposed-143-million-us-consumers.html

[10] FTC. **Equifax Data Breach Settlement.** Online, Available: https://www.ftc.gov/enforcement/cases-proceedings/refunds/equifax-data-breach-settlement

[11] Garfinkel, S. **Anti-Forensics: Techniques, Detection and Countermeasures.** The 2nd International Conference on i-Warfare and Security (ICIW), Naval Postgraduate School, Monterey, CA, March 8-9, 2007. pp 77-84.

[12] Snider, M. **Your data was probably stolen in cyberattack in 2018 – and you should care.** Online, Available: https://www.usatoday.com/story/money/2018/12/28/data-breaches-2018-billions-hit-growing-number-cyberattacks/2413411002/