Big Data and Privacy: Finding the Balance in Distrust and Progress

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Ziyang Guo
Spring, 2020

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Introduction

Data breaches have become an increasingly serious problem. From 2005 to 2014, the number of personal data leakages has been quintupled. 783 data breaches, with at least 85.61 million total records exposed, were reported in 2014. However, the number of data breaches did not decrease over time; in 2017, 1,579 data breaches were reported (Sobers, 2020). The increase in data cyber hacking was linked to the arrival of new data mining technology and the deficiency in data security (Rechavi et al., 2018). In the past decade, regulatory response continued to disappoint the general public with lacking protection in consumers' data safety. Two historical cases, Anthem Insurance data breach and Facebook user information hack, are analyzed in this paper. In 2015, Anthem Insurance leaked 80 million patient medical records, and this incident became the biggest medical data breach in the American history (Kamoun & Nicho, 2018). Between 2017 and 2018, more than 50 million Facebook users' information, such as passwords and profile information, was leaked and another 40 million users' data was at risk of hacking (Cadwalladr & Graham-Harrison, 2018). Understanding the technical and legislative circumstances is critical to investigate the data safety and to propose potential solutions. This paper illustrates how legal policies failed to protect consumers' privacy and what lessons are learned in the aforementioned cases. Two frameworks, technological momentum and actor network theory, are used to unfold the progression of data breaching events.

Research Question and Methods

This research paper aims to answer the following question: how have the recent data breaches demonstrated the current circumstances in consumer data safety, and what regulations should be put in place to secure consumer privacy? The research is mainly conducted through

studying historical cases, including Anthem Insurance data breach, and Facebook user information leak. Relevant data related to regulations such as HIPAA (Health Insurance Portability and Accountability Act), GDPR (General Data Protection Regulation), and NCSL (National Conference of State Legislature) Data Security Laws are used to present the relationships to data breach cases and analysis of possible future improvement. All information is obtained through reading primary sources and literature reviews. Primary sources include case documents provided by the court, Anthem Inc., and Facebook, Privacy Acts and Codes in different states, and regulations issued by law-making entities aforementioned. Research resources will be presented in two independent case studies that will ultimately point to a unifying conclusion.

Background Information:

In the end of January 2015, Anthem insurance company became aware that its database had been compromised and 80 million customers' records were retrieved by unauthorized attackers (Browning & Tuma, 2015). The company reported this incident to the federal authorities within 72 hours and revealed it to the public within a week. In the company's report, personal information, such as names, birthdays, medical IDs/social security numbers, street addresses, email addresses, and employment information, including income, was accessed and stolen by the hackers (Nelson et al., 2015). After investigation, it was determined that hackers were able to access the database of Anthem insurance company because Anthem did not encrypt its database and user information. In August 2018, Anthem agreed to pay 115 million dollars to settle several civil class-action lawsuits. In November 2018, Anthem also paid 16 million dollars to settle the HIPAA Privacy and Security Rules violations (Armerding, 2019). Anthem

subsequently spent more than 260 million dollars for data security improvements and remedial actions (Morse, 2017).

After the revelation of Anthem's failure to secure its consumers' data, support for stricter privacy law was surging. In 2017, in the midst of Anthem lawsuits, Indiana, where Anthem headquarters was located, amended its Trade Regulation Indiana Code 24-4.9-3-3.5 "Duties of a database owner", requiring all database owners to implement and maintain reasonable procedures and security to protect and safeguard from unlawful use or disclosure any personal information (*Indiana Code 2019—Indiana General Assembly, 2020 Session*, n.d.). Many other states and regions followed suit. Rhode Island passed Identity Theft Protection Act of 2015 shortly after the attack to require all entities, who store, collect, process, maintain, acquire, use own, or license personal information, must implement and maintain a "risk-based information security program that contains reasonable security procedures and practices appropriate to the size and scope of the organization" (*11-49.3-2*, 2015, p. 49).

The second case study is the Facebook login information leak. In September 2018, after the Cambridge Analytica scandal, in which Facebook allowed hundreds of mobile applications to harvest more than 50 million users' personal information, the company was scrambling to regain its users' trust after another security incident that exposed more than 87 million users' private identification and passwords (Cadwalladr & Graham-Harrison, 2018). The breach was caused by a vulnerability, introduced in July 2017, that allowed unauthorized individuals to gain digital credential tokens which can be used to log into other people's accounts. Facebook noticed the unusual activity spike in September 2018 and tried to fix the bug (Rehman, 2019). During the data leaking incident, Facebook failed to notify authorities within 72 hours of becoming aware of

data loss, violating the GDPR requirements on data protection and privacy in the European Union and the European Economic Area (Lovejoy, 2019).

Technological Momentum and Actor Network Theory

There are two frameworks used to interpret the development of both case studies and identify the interactions between the legislative department, companies, and individuals. The first framework is technological momentum. Technological momentum is a phase where social construction and technological determinism both take place (Hughes, 1994). The emergence of data science has followed the trend depicted by technological momentum. Data and computer technologies have brought convenience of storing large amounts of information cheaply and easily. Anyone with a remotely smart electronic device can upload and manage their information digitally at their fingertips. The novel technologies have transformed the ways societies utilize data. With large amounts of information stored on the internet, databases of personal information have become more valuable. Cyber-attacks have therefore become a common occurrence. Big data technology is now often linked to cyber-attacks and a negative connotation. In a 2015 survey from Health Information Technology, 75% of Americans concerned about their health data being used by private companies, and 10% of all Americans decided to withhold all information from their health care providers to avoid private data breaches (Patel et al., 2015). Data science, as technological momentum described, not only transformed the way people deal with information, it also changed how people perceive its related topics. However, technological momentum also has its limitations. Technological momentum focuses on illustrating the idea that technology shapes society, failing to explain the interactions and forces that social development applies to the technology. Technological momentum is unsuitable for innovations that have

never rapidly changed society or have prolonged delays before impact. Many critics believe technological momentum is only appropriate for explaining certain technological phenomena (Kinney, 2012).

Another framework introduced in this paper is actor network theory (ANT). ANT is a constructivist approach to describe a group of unspecified relationships among entities (Latour, 1996). ANT describes all entities as equally important actors which exist in an ever-changing network of relationships. ANT believes that each actor has the ability to affect other actors within the web of relationships. Hence, ANT helps technological momentum explain the backward influence that society has on technology. In this paper, ANT is also particularly relevant in explaining the cause-effect relationship between both case studies and legal regulations. ANT is able to demonstrate how data breaches affect lawmakers to amend the laws and affect people; ANT is also capable of illustrating how some regulations, such as GDPR agreements, are ineffective in protecting consumers' privacy. However, ANT also has its own limitations. One of the biggest criticisms of ANT is that the theory implies that all actors are of equal importance in the network of relationships (Sheldon, 2010). This implication is obviously incorrect. Different factors within the same event often have greatly disparate effects on the final result. For example, in the space shuttle Challenger disaster, there were a myriad of factors that could have contributed to the explosion of the rocket. However, the most critical and outstanding cause was the malfunction of the O-ring under low temperature, and it should not bear the same importance as other trivial factors (*(39) Space Shuttle Challenger Disaster: Ethics Case Study No. 1—YouTube*, 2015).

Results and Discussion

The increasingly rampant data breach incidents, from 157 data breaches in 2005 to 783 in 2014, indicate there are weaknesses and failures in the current system that protects consumer data privacy (Groot, 2018). In the light of two case studies, Anthem Insurance data breach and Facebook user information leak, it is apparent that there are two major factors that contribute to the current data safety status. First, the newly emerged technology, big data, is not well understood and managed due to its infancy. New ways of exploiting the novel technology's weakness appear every day, yet, techniques in defense are still catching up. The rapid shift to big data and the massive impact of data breaching is comprehended as the second phase of technological momentum, a theory of relationship between technology and society proposed by Thomas Hughes (Hughes, 1994). Technological momentum is a time-dependent theory that starts out with social construction and gradually transitions to technological determinism. Social construction describes that society controls and restricts technology and technological determinism states that technology shapes society. The current big data technology status is characterized as in its peak of technological momentum where society both influences and is influenced by technology. Big data technology has forced increasingly more companies to move their valuable data towards cloud managed services due to the sheer size of their data. A survey by Intel Security shows that 73% of companies have already moved or are moving towards their workloads and data to cloud (Armerding, 2018). New big data technology has also introduced many problems the world has never seen before, one of which is data breaching. The second factor is the lack of protective regulations in the data privacy and security field. There are extremely limited amounts of regulations in data privacy protection. Most of the states in America only have notification laws that require companies to report data breach incidents, but

there are no protection laws. Even federal regulators at the U.S. Department of Health and Human Services openly acknowledge, "that most Americans have no grasp of whether their information is protected by the law" (O'Connor, 2018). The urgent need for regulations against data theft is explained by actor-network theory (ANT). In an oversimplified three-way relationship network, companies, consumers, and lawmakers are the main actors that influence and constrain each other. However, with the emergence of big data technology, companies gain more power that eventually breaks this check-and-balance situation. Therefore, in order to restore the equilibrium, it is crucial to increase regulatory power.

Anthem Data Breach

In February 2015, Anthem, American largest health insurer, disclosed a massive cyber-attack that had stolen more than 78.8 million unique user private information such as name, address, birthday, and social security number from the company's database warehouse (McGee, 2017). This 2015 cyber-attack against Anthem was the biggest data breach in the health sector, doubling the total number of records exposed in history (Donovan, 2018). During an internal investigation by security firm Mandiant, which Anthem hired, it was determined the data breach started on February 18, 2014, a year before the revelation of the massive cyber-attack, when a user within one of Anthem's subsidiaries opened a phishing email containing malicious content (McGee, 2017). The email allowed hackers to download virulent files that helped hackers gain remote access to that computer and several systems within the Anthem enterprise. The malicious viruses soon spread among the enterprise and escalated its accessing privileges. By the time the breach was detected, "the attacker utilized at least 50 accounts and compromised at least 90 systems within the Anthem enterprise environment including, eventually, the

company's enterprise data warehouse, a system that stores a large amount of consumer personally identifiable information," noted by the investigating firm (McGee, 2017). The biggest data leaking scandal in American medical history therefore emerged.

According to Anthem spokeswoman Jill Becher, because there was no evidence that any compromised information was sold or used to commit fraud after investigation, no federal or district charges were issued on the case (Pierson, 2017). After two years of litigation, a federal judge, Lucy Koh, approved a 115-million-dollar class action settlement. Under the 115 million dollar settlement, 51 million dollars would go to the victim, 17 million dollars would be spent on credit-monitoring services, 28 million dollars would go to customers who suffered loss, 23 million dollars was earmarked to pay a legal consulting firm, Kurtzman Carson Consultant, who helped settle the case, and the rest 2 million dollars were spent to reimburse attorneys for expenses (Andrews, 2018). The total 115 million dollar settlement mounted to 0.1259% of the 91,341 million dollar total operating revenues of Anthem's earnings in 2018 (Rigg & Becher, 2019). Having the world's biggest data breach scandal, Anthem seemed to sneak away without any punishment. Most money from the settlement did not even go to the victims of this breaching incident. Clearly, no law at this time was able to hold Anthem accountable or responsible for their negligence, and the loss of millions of personal records was just a mere misfortune.

The Anthem data breach incident demonstrates that there is a great vacuum in legislation to penalize companies that cannot effectively protect their consumers' data safety. In other fields and businesses, borrowers are held responsible for what they are lent to. For instance, if a storage management facility lost all the valuable archived documents that belonged to a library in a fire,

the storage facility will most likely be legally obligated to refund the service they failed to provide and compensate for the value that was lost during the accident. However, in the case of Anthem data breach, no laws or regulations in the area of data security could hold Anthem accountable for the damage done. In fact, because there was zero precedence or guideline to follow, Anthem Inc. was able to hire their own legal consultant team to assign themselves a virtually arbitrary settlement. In the amount of the money Anthem Inc. set aside for the incident, less than half, which was about 65 cents per person, was actually going to the victims. Even though it was widely speculated that the cyber-attack was sponsored or even organized by a foreign government, and it was probably true that, "insurers and regulators alone cannot stop foreign government assisted cyber-attacks," said California Department of Insurance Commissioner Dave Jones, the abysmally tiny amount of compensation paid to the victims was still not justified (McGee, 2017). Furthermore, it was shown during the investigation that, because Anthem did not take precaution in protecting user data through encryption, hackers were able to easily extract information from the database warehouse (Abelson & Goldstein, 2015). Essentially, Anthem stored all their users' information in plain text which could be read by any literate person once stolen. Many states, including Indiana where Anthem headquarters is located, failed and are still failing to protect consumers' information by forcing a certain degree of encryption or redaction. In the mere 24 lines of Indiana state law *IC (Indiana Code) 24-4.9-3-3.5 Duties of a Database Owner*, it reads it "requires the database owner to maintain reasonable procedures to protect and safeguard from unlawful use or disclosure personal information" (*Indiana Code 2019—Indiana General Assembly, 2020 Session*, 2009). The word "reasonable" granted companies virtually unrestricted freedom to interpret the level of protective

measures that were needed. Anthem could have hired a Pembroke Welsh Corgi to look after the door of their database warehouse and called it "reasonable." This law was only amended in 2017 after Anthem data breach to include "taking any appropriate corrective action to protect and safeguard from unlawful use or disclosure any personal information", which was still vague in what specific appropriate corrective actions were. Clearly, a lack in legislative power un-incentivizes companies to spend resources to safeguard the private user information that they possess and fails to hold companies in fault to pay for their mistakes.

Facebook Data Leak

The second case study is Facebook user login information leak that allowed third party applications to access users' private photographs without their permission. In December 2018, after the infamous Cambridge Analytica scandal in which Facebook sold their 50 million users information and the largest social media data security breach that expose more than 87 million people's identification, Facebook notified the authority that they were yet again involved in another security failure that allowed up to 1,500 different applications gain access to their users' private photos without permission (Kastrenakes, 2018). The data breach was caused by a bug in Facebook photo API, which was introduced in a software update on September 13, 2018. The company first noticed and fixed the breach on the 25th of the same month. However, Facebook failed to file a report immediately after the revelation of the breach. Facebook waited three months and finally notified the public and affected users in December (Leetaru, 2018). In the United States, Facebook faced no federal charges or privacy law violations; only an application notification was promoted to notify the affected users a data breach might have affected their accounts. In contrast, Facebook faced serious charges for violation of General Data Protection

Regulation (GDPR) laws and potential fines up to 4% of the company's global revenue for the preceding year, which could reach up to 2.23 billion dollars (Price, 2019). Clearly, there is an enormous regulatory discrepancy between the United States and the European Union. In these two considerably equally developed regions, the United States is apparently lagging far behind in regulations to protect consumers' data safety.

In California, where Facebook headquarters is located, Civil Code CIV Division 3 Title 1.18 Obligations Regarding Customer Records mentions when any "personal information was, or is reasonably believed to have been, acquired by an unauthorized person, disclosure shall be made in the most expedient time possible and without unreasonable delay" (*Law section*, 1988). However, ambiguity again makes this notification law difficult to execute. The law states disclosure shall be made without "unreasonable delay." Yet, there is no description in the Civil Code that requires companies to report potential incidents within a specific number of days or hours. It is once again up to the companies to decide what they believe to be a reasonable timeframe to report any data breaching incidents. Over at the opposite side of the Atlantic Ocean, the European Union requires companies to rapidly report data breaches within 72 hours to the general public under GDPR laws (Constine, 2018). Facebook is now under litigation that might cost them billions in dollars. The failure of Facebook reporting their data breach demonstrates that laws and regulations should change and update rapidly and accordingly with the technology. Just a few years ago, big data storage did not exist and no laws or regulations anticipated problems like data breaching would be prevalent in the future. However, throughout the past few years, skyrocketing data breach incidents led the European Union to implement GDPR laws to protect consumers' privacy on May 25th, 2018 (*General Data Protection Regulation (GDPR) –*

Official Legal Text, 2016). The United States was left behind. Consumer data privacy law still does not exist as a federal regulation in America. In addition, states established their data control laws decades ago and have not since updated them. For example, California amended the state laws to include Obligations Regarding Customer Records in 2000 and has not yet made any change to the law in two decades. It is apparent in this case study that the regulatory power fails to keep up with the speed of changing technologies. This transition of lawful attempt to control data safety to technology dominance is yet another prime example of technological momentum. The big data technology has brought tech-companies to the top of the food chain. This newly added actor, big data technology, has broken the equilibrium in the influences of data safety. Society is expected to combat this unbalanced influence with a lacking regulatory actor.

Limitations to this research project include time and resource restraint, and difficulty in finding non-existent regulations. The topic of this project was determined in late 2019 and the research was scheduled to complete in early 2020. The short span of time provided to this project restricts extension in reading materials and therefore narrow the depth of the research. In this project, only two case studies are presented to illustrate the main theme. With more available time, this project will be expanded to incorporate more case studies. Resource deficiency is another factor that influences the depth of this project. Due to the scandalous nature of the investigated events, involved companies tend to present less and limited information to the general public. An interview was successfully set up with the involved companies. Therefore, the decisions and the companies' internal processes during the events could only be speculated. Lastly, the nature of this project is to expose the lack of regulatory response in the surge of data breaching events. It is always more difficult to identify things, in this case, laws and regulations,

that do not exist. Due to the lack of comparison to existing precedence, most of the protective measures are imaginative.

If this project was to be researched in the future, a few improvements should be made. First, the depth of the research scope should be broadened. This research project includes mere two case studies that only cover parts of the data safety field. More cases should be studied to present the audience a more comprehensive view of the current data status and the potential solutions. Second, this project should be improved by incorporating more people in researching and comparing the differences of regulatory codes between different regions of the world to identify the better solutions. Lastly, this project should be improved by securing interviews with companies in question to more closely present the insider aspect. Hence, there are certainly many areas of this research that can and should be improved in the future.

Conclusion

The Anthem and Facebook case studies have shown U.S. regulators are failing to keep up with the fast-moving technologies. Many laws regarding data security and privacy have not been modified since their inception and are clearly outdated. The lack of power in regulation breaks the balance triangle between consumers, companies, and legislation. Companies are now free to abuse the power of data science technology without paying the proper price for violating consumer privacy. More specific instructions, such as number of days or hours a company should report its data breach in and the level of security and penalty should be implemented for data owners, should be considered and added to the regulations.

Work Cited

11-49.3-2. (2015). http://webserver.rilin.state.ri.us/Statutes/TITLE11/11-49.3/11-49.3-2.HTM

(39) Space Shuttle Challenger Disaster: Ethics Case Study No. 1—YouTube. (2015, November 18). https://www.youtube.com/watch?v=QbtY_Wl-hYI

Abelson, R., & Goldstein, M. (2015, February 5). Anthem Hacking Points to Security Vulnerability of Health Care Industry. *The New York Times*. https://www.nytimes.com/2015/02/06/business/experts-suspect-lax-security-left-anthem-vulnerable-to-hackers.html

Andrews, G. (2018, August 20). Anthem data breach judge OKs huge fee award, but not as much as attorneys wanted. *The Indiana Lawyer*. https://www.theindianalawyer.com/articles/47891-anthem-data-breach-judge-oks-huge-fee-award-but-not-as-much-as-attorneys-wanted

Armerding, T. (2018, September 26). *Cloud migration: How and why business is moving to the cloud | Synopsys*. Software Integrity Blog. https://www.synopsys.com/blogs/software-security/cloud-migration-business/

Armerding, T. (2019, February 21). *Throwback Thursday: Whatever happened to Anthem? - Security Boulevard*. https://securityboulevard.com/2019/02/throwback-thursday-whatever-happened-to-anthem/

Browning, J. G., & Tuma, S. (2015). If Your Heart Skips a Beat, It May Have Been Hacked: Cybersecurity Concerns with Implanted Medical Devices. *South Carolina Law Review*, *3*, [i]-678.

Cadwalladr, C., & Graham-Harrison, E. (2018). *Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach*. 5.

Constine, J. (2018, October 1). Facebook breach hit up to 5M EU users, and it faces up to $1.63B in fines. *TechCrunch*. http://social.techcrunch.com/2018/10/01/facebook-breach-europe/

Donovan, F. (2018, August 20). *Judge Gives Final OK to $115M Anthem Data Breach Settlement*. HealthITSecurity. https://healthitsecurity.com/news/judge-gives-final-ok-to-115m-anthem-data-breach-settlement

General Data Protection Regulation (GDPR) – Official Legal Text. (2016). General Data Protection Regulation (GDPR). https://gdpr-info.eu/

Groot, J. (2018, November 12). *The History of Data Breaches* [Text]. Digital Guardian. https://digitalguardian.com/blog/history-data-breaches

Hughes, T. (1994). *Technological Momentum*.

Indiana Code 2019—Indiana General Assembly, 2020 Session. (2009). http://iga.in.gov/legislative/laws/2019/ic/titles/024/#24-4.9-3-3.5

Kamoun, F., & Nicho, M. (2018). A New Perspective on the Swiss Cheese Model Applied to Understanding the Anatomy of Healthcare Data Breaches. *Handbook of Research on Emerging Perspectives on Healthcare Information Systems and Informatics*, 58–81. https://doi.org/10.4018/978-1-5225-5460-8.ch004

Kastrenakes, J. (2018, December 14). *Facebook exposed up to 6.8 million users' private photos to developers in latest leak*. The Verge. https://www.theverge.com/2018/12/14/18140771/facebook-photo-exposure-leak-bug-mill

ions-users-disclosed

Kinney, S. (2012, September 2). *Technological Momentum | Teaching as Dynamic*. https://educatech.wordpress.com/2012/09/02/technological-momentum/

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale Welt*, *47*(4), 369–381. JSTOR.

Law section. (1988). https://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.82

Lovejoy, B. (2019, May 28). *GDPR fines total €56M in first year as Facebook under scrutiny—9to5Mac*. https://9to5mac.com/2019/05/28/gdpr-fines/

McGee, M. (2017, January 10). *A New In-Depth Analysis of Anthem Breach*. https://www.bankinfosecurity.com/new-in-depth-analysis-anthem-breach-a-9627

Morse, S. (2017, January 6). *Anthem cyberattack perpetrated by foreign government, officials say | Healthcare IT News*. https://www.healthcareitnews.com/news/anthem-cyberattack-perpetrated-foreign-government-officials-say

Nelson, G. S., Technologies, T., & Hill, C. (2015). *Practical Implications of Sharing Data: A Primer on Data Privacy, Anonymization, and De-Identification*. 23.

O'Connor, N. (2018, January 30). *Reforming the U.S. Approach to Data Protection and Privacy*. Council on Foreign Relations. https://www.cfr.org/report/reforming-us-approach-data-protection

Patel, V., Hughes, P., Savage, L., & Barker, W. (2015). *Individuals' Perceptions of the Privacy*

and Security of Medical Records. 9.

Rechavi, A., Berenblum, T., & Maimon, D. (2018). *The Secondary Global Market for Hacked Data*. *12*(2), 20.

Rehman, I. (2019). *Facebook-Cambridge Analytica data harvesting: What you need to know*. 12.

Pierson. (2017, June 23). Anthem to pay record $115 million to settle U.S. lawsuits over data breach. *Reuters*. https://www.reuters.com/article/us-anthem-cyber-settlement-idUSKBN19E2ML

Price, E. (2019, August 12). *The EU Might Fine Facebook Billions For GDPR Violations*. Digital Trends. https://www.digitaltrends.com/social-media/facebook-gdpr-decision/

Rigg, C., & Becher, J. (2019, January). *Anthem Reports Fourth Quarter and Full Year 2018 Results Reflecting Strong Core Performance*. Anthem, Inc. https://ir.antheminc.com/news-releases/news-release-details/anthem-reports-fourth-quarter-and-full-year-2018-results

Sheldon. (2010, January 1). *Criticism of Actor-Network Theory*. https://island94.org/2010/01/Criticism-of-Actor-Network-Theory.html

Sobers, R. (2020, January 28). *107 Must-Know Data Breach Statistics for 2020 | Varonis*. https://www.varonis.com/blog/data-breach-statistics/