Large-Scale Data Breaches:

Keeping Personal Information Protected Online

Emma Brunell and Justin Ross

SUNY Oswego

CSC 333

I. Introduction

In this day and age of big data, there is huge emphasis on protecting one's personal information. The internet has evolved much since its inception; today, nearly everything functions on the web to some degree. From social media websites to online job applications, most individuals have inputted some type of sensitive information online without a second thought.

The culture of using the internet for daily functions has recently become a privacy issue. In 2017, Equifax, a consumer credit reporting agency used by 88 million businesses worldwide, revealed that a data breach released the personal information of 143 million Americans. This personal information included individuals' bank details and social security numbers, as well as a plethora of other sensitive data. In 2018, Facebook also announced that they also experienced a data breach which affected 87 million of their users. The information exposed in this breach was used by the data company, Cambridge Analytica, to attempt to influence the 2016 presidential election by gathering users' personal preferences and displaying false advertising to these targeted individuals.

In sharp contrast to these shocking breaches of personal information, there are also large companies who profit off their users' data by selling it to other organizations. Analytics services monetize user data to provide more targeted advertising and to provide a more personalized Internet browsing experience. One such company, MoviePass, has discussed the possibility of selling their users' locations. In this

scenario, a user would check in to a movie at a theater and watch it. If they had location services on their phone turned on, MoviePass would collect information about where they were before and after their movie. They would be able to partner with other businesses in the user's area to suggest options for where to go out to dinner or go shopping. While data analytics motivations such as this is not nearly as malicious as the Equifax and Facebook data breaches, it continues to raise eyebrows. The average person likely is not aware that their personal information is up for sale, which is a huge privacy risk. However, this is not regarded as serious as targeted attacks on online companies to release data.

The topic to be investigated in this matter deals with the differences between data breaches versus companies who sell user information for profit. Additionally, the specific methods attackers use to cause these breaches to occur will be explored, as well as tactics the average individual can use to protect themselves online from loss of personal information.

II. Data Breaches Versus Data Broker Firms

The main characteristic of a data breach is that the information released was formerly secured. The information is typically bank details, personal health information, and personally identifiable information, in a vulnerable, unencrypted format. Due to this, data breaches pose a large risk in identity theft.

Companies that experience a data breach also tend to see a significant drop in their quarterly profit; breaches affect the economic health of the country where it occurs. According to Ponemon study, data breaches cost healthcare businesses $6.2 billion between 2014 and 2015.

Hackers use the data released to make a profit for themselves. Personally identifiable information can be used to duplicate credit cards. In some cases, the data is sold to others on the deep web. Recently, selling stolen Uber accounts has had a sharp increase in popularity -- they are sold at $1.15 USD each.

Data broker firms collect, analyze, and sell users' personal data to other firms, advertising organizations, and even the government. While consumer marketing is not a new trend, the nature of the data being sold has become increasingly more personal.

The largest data firm in the country is called Axciom. Axciom has collected information on over 200 million Americans. For each individual they have a dossier on, it contains (on average) 1,500 different pieces of data. The danger with companies like Axciom is that they do not disclose exactly what information they collect. These companies know an individual's medical history, psychiatric problems, and even sexual

orientation. In countries where homosexuality is punishable by death, this could cause serious implications. Even if an individual did not explicitly search for, for example, gay bars in their area, these data companies are still able to piece together sexual orientation by data points that are bought and sold. If this individual's portfolio was then sold to the government, they would be in much trouble.

While many people would assume that these data broking companies have high systems security requirements, studies show otherwise. Only 25% of the top 100 data firms use encryption on their website's landing page. Additionally, the session cookie on the landing page is identical to the cookie after logging into the website. Because the cookie is the same throughout the session, a hacker could easily launch a man-in-the-middle attack and hijack a user's account. Additionally, studies show that cross-site scripting attacks and SQL database injections are both common successful attacks on data firms. Because these data brokers have so much data on an individual in one location, a hacker who wanted to gather intelligence on a user would be able to see everything they do and buy online.

Data breaches and brokers who collect information on Internet users seemingly go hand-in-hand. In several years, the information that brokers have on individuals will skyrocket. Due to the lack of transparency between the brokers and the public, they may be gathering more sensitive personally identifiable information on individuals as well. It is getting increasingly more important for the public to realize that their every move is being tracked online.

III. Methods of Data Release

A common method that cyber-criminals often use to infiltrate databases is an SQL injection attack. This type of attack takes advantage of vulnerabilities where a user has to enter sensitive data into a form. This data is then stored inside a database. If the hacker enters code or specific characters into this form, typically where a username and password would go, they are able to gain administrative access to the database.

In 2017, forty-seven percent of data breaches were caused by some form of hacking. Of these hacking attacks, sixty-three percent of them were caused by phishing. When a hacker sends out an email to individuals with the intention of spoofing a website and gaining the individuals' username and password, this is known as phishing.

The second largest cause of data breaches in 2017 was due to company employee negligence and error. This made up nine percent of data breaches. This typically occurs in the healthcare industry. Misdelivery of information, data disposal error, and misconfigurations are common causes of healthcare data breaches. The Ponemon report discovered that less than half of healthcare businesses require their employees to engage in security training. With employee error being one of the largest causes of data breaches, this is extremely concerning; this trend could be averted with some simply security education. Additionally, at smaller businesses that are not typically targeted by cyber-criminals, employee error is the leading factor in data breaches. If an employee accidentally leaves their work laptop at a coffee shop, and it gets stolen, there is the possibility for a data breach to occur.

IV. Keeping Sensitive Data Safe

A. Methods of Protection

One of the main ways companies can prevent security breaches is to keep their security software up to date. In the case of the Equifax data breach, a software update was available to close vulnerabilities, but was not implemented by the company. Companies such as Equifax should regularly check their vendors' websites for any security vulnerabilities that may have been recognized in the current software release, to then be aware of this vulnerability and apply the correct precautions.

Additionally, companies need to monitor all of their data transmissions. Any transmission that is unencrypted should not be allowed if it contains any type of personal data; this includes emails sent between employees of the company. If the company is exchanging unencrypted data on a wireless network, it is especially open to data interception.

As SQL injection attacks are one of the leading causes of data breaches, and one of the easiest for cyber-criminals to carry out, companies should be sure that any queries to their database have specific and typed parameters. Also, companies should limit which websites have complete access to their database on a need-based system. This is known as the principle of least privilege. Many websites don't have a need to be able to delete or insert new items into the database, so they should not be able to do so; as this is how SQL injections occur.

It was found in 2017 that a large number of data breaches that have occured involved compromised user credentials, specifically, their login information. While cyber-criminals may use various tactics to obtain these credentials, the majority of the time an account is attacked, it is accessed through the “main gate”, as it is the path of least resistance. While much of what is involved with keeping user credentials secure is placed onto companies’ security decisions, such as proper encryption and data vaults, individuals also should make sure to make sure their passwords are strong enough that they cannot be subject to a brute force attack.

B. Web of Trust

Due to hacked passwords being a leading cause of personal information being stolen, it is important for individuals to be aware of safe websites and the strongest passwords to keep their accounts secure.

The open source browser addon, Web of Trust, is a website checker that uses machine learning and user reviews to determine the trustworthiness of a site. It shows how safe each link is in a search engine. While knowing if a website might be malicious is extremely important, Web of Trust typically doesn’t have a rating for smaller websites that have less traffic and reviews. For websites like these, if an individual must create an account, they want to be sure their account is as safe as possible. Because passwords are so commonly hacked by brute force, individuals should never use short or easy-to-guess passwords.

The Web of Trust does not have any way to educate users on the strength of their password. The implementation addition to the Web of Trust was to create an educational password cracker for users to test out potential passwords on before creating an account on a potentially untrustworthy website. This cracker uses the brute force method to guess the password.

A python script was written that performs the brute force crack and returns the password and the number of attempts taken. That python script accepts passwords containing every alphanumeric character as well as a set of special characters. The password cracker only accepts up to 9 characters; a password longer than this amount is only more difficult to decode. The python script was run on a local Django server. It needed to be able to parse the HTTP request to get the query containing the password in question, and produce JSON that the Web of Trust extension can understand. The Django server is only for development purposes; if this project were to be professionally implemented, a dedicated server would be used.

The open-source Web Of Trust code was modified to show the password cracker on the “ratings page”. This page can be accessed by clicking the WoT donut in the top right. The source HTML code was modified to display the addition. A custom javascript file was then added to make calls to the Django server and pass the results back to the web page.

The Web of Trust implementation addition teaches users that the length of a password increases its security, as well as a mix of numbers and uppercase

characters. Individuals should never use the same password for every site. While brute force cracking is a way to access accounts, it takes time. If a user has the same password for every account, the hacker doesn't need to brute force their way into the account, making it much easier for them. However, many password crackers are algorithm-based, and will detect patterns better than a brute force method would. Keeping a password random and not following common patterns is a sure way to protect a password.

While data breaches are a large problem for major companies at this point in time, individuals mainly have to trust the website that they are using, and hope that the company used the right precautions to secure their data. On a higher, security-protocol-based-level, there is not much an average individual is able to accomplish with keeping their data better protected. Much of data breaches fall in the hands of the company or employee in charge of the data. However, not falling into phishing scams, installing safe browser add-ons to stop data analytics trackers, making sure login credentials are secure, and only visiting websites that are trustworthy are the best ways that individuals can protect themselves from their personal data being stolen.

V. References

1. How Does a Data Breach Happen? (n.d.). Retrieved June 15, 2018, from https://www.travelers.com/resources/cyber-security/how-does-a-data-breach-happen
2. Snell, E. (2017, January 20). 2016 Healthcare Data Breaches Largely From Employee Error. Retrieved June 16, 2018, from https://healthitsecurity.com/news/2016-healthcare-data-breaches-largely-from-employee-error
3. Almost 90% of Cyber Attacks are Caused by Human Error or Behavior. (2017, May 07). Retrieved June 28, 2018, from https://chiefexecutive.net/almost-90-cyber-attacks-caused-human-error-behavior/
4. 2017 Cost of Cyber Crime Study | Accenture. (n.d.). Retrieved June 24, 2018, from https://www.accenture.com/us-en/insight-cost-of-cybercrime-2017
5. Simon, S. (2017, September 16). What It Might Take To Stop The Data Breaches. Retrieved June 21, 2018 from https://www.npr.org/2017/09/16/551467158/what-it-might-take-to-stop-the-data-breaches
6. How to Protect Against SQL Injection Attacks. (n.d.). Retrieved June 28, 2018, from https://security.berkeley.edu/resources/best-practices-how-articles/system-application-security/how-protect-against-sql-injection

7. WOT Services LLC. (n.d.). Home – WOT (Web of Trust). Retrieved June 3, 2018, from https://www.mywot.com/
8. Marechal, S. (2007). Advances in password cracking. Journal in Computer Virology,4(1), 73-81. doi:https://doi.org/10.1007/s11416-007-0064-y
9. Frequently Asked Questions on Incidents and Spills. (n.d.). Retrieved June 3, 2018, from https://www.archives.gov/isoo/faqs/incidents-spills.html
10. Ponemon Institute's 2017 Cost of Data Breach Study: Global Overview (Rep.). (2017, June). Retrieved June 25, 2018, from Ponemon Institute