Data Breach

A Common Threat but causes Adverse Damage

Sree Vandana Nadipalli
Computer Science
Portland State University
Portland, Oregon, USA
sreev2@pdx.edu

Swetha Venkatesan
Computer Science
Portland State University
Portland, Oregon, USA
swet@pdx.edu

ABSTRACT

Data breaches are more common than we think they are. "One massive hack after another," this statement would probably best describe what's happening around. Data breaches have become a status quo, an alarming fact, but not surprising, considering how attackers keep finding ways to infiltrate networks and steal information. Although hackers are obvious culprits in uncovering data, oftentimes they had a helping hand from human error resulting in a data breach. In our research we will be providing the origins of Data breach, recent Data Breaches and the effects and damages caused by them, what is stolen in the breach and the necessity to identify these breaches. We are going to point out different analyses and observations that are drawn from different research papers that could dispel the common myths that we have on data breaches. And we will further continue to find methods to identify these breaches. We will be doing an implementation project on identifying one of the potential breaches (Financial). We will be doing this by using credit card transaction data and automating the process of identifying the fraudulent transactions using machine learning algorithms. Our goal will be to create an unbiased model that could detect the fraudulent transactions with high accuracy.

KEYWORDS

Data Breach, Vulnerabilities, Ransomware, SQL injection, Spyware, Phishing, PII, PoS RAM scrapers.

1 Introduction

What exactly is Data Breach? To better understand these breaches, it is important to define the term. The International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 27040 defines a data breach as: "*Compromise of security that leads to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to protected data transmitted, stored, or otherwise processed.*" The U.S. The Department of Justice defines a breach as "*the loss of control, compromise, unauthorized disclosure, unauthorized acquisition, access for an unauthorized purpose, or other unauthorized access, to data, whether physical or electronic*". So, basically it is gaining access to data or information in an unauthorized way. When I say the word, Data Breach, or data being compromised, what picture immediately comes to mind, A hacker sitting in a dark room with a laptop coding something? That could be, but it's not the only thing. To better understand, let us go into history, the origins of Data Breaches.

2 The origin of data breaches

Data breaches didn't start when organizations began to store their data digitally. In fact, data breaches have existed for as long as individuals and companies have maintained records. According to the Office of Inadequate Security website, in 1984 the global credit information corporation known as TRW (now called Experian) was hacked and 90 million records were stolen. Before computing became common, a data breach could be something as simple as viewing an individual's medical details without authorization or finding sensitive documents on a desk instead of in a drawer.

2.1 Laws and Regulations

Laws and regulations like HIPAA, PCI-DSS, GDPR, CCPA, FIPA, the SHIELD Act, and LGPD have created guidelines for organizations handling certain types of sensitive information. While these regulations provide a framework for required safeguards, storage, use, and handling of sensitive information, they don't stop all data breaches from occurring. Because of this, most information about the number of data breaches and their impact focus on the period between 2005 to the present. Largely due to the advancement of technology and the proliferation of electronic data, which have greatly increased the total number of individuals impacted. Today's data breaches often impact millions - even billions - of individuals. Now, Data breaches are complex events. Any business or organization that processes and/or stores sensitive data is a potential breach target. Figure 1 shows the Data Breach incidents that were disclosed between 2005 to April 2015.

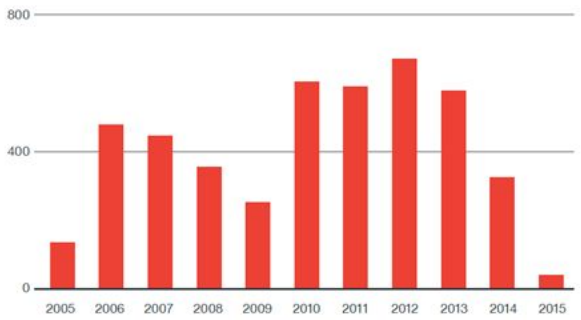

Figure 1: Data Breach incident disclosures from 2005 to April 2015

3 Recent data breaches and statistics

With over 2,000 confirmed data breaches in 2019 and hundreds in 2020. We are listing the recent Data breaches.

3.1 Health Share of Oregon

February 13, 2020: The theft of an employee laptop from GridWorks IC, a third-party vendor of Health Share of Oregon, has exposed the personal and medical information of 654,000 members. The Health Share of Oregon data breach disclosed sensitive data, including names, addresses, phone numbers, dates of birth, Social Security numbers, and Medicaid ID numbers.

3.2 T-Mobile

March 5, 2020: An unknown number of customers' sensitive information was accessed through a T‑Mobile employee email account after a malicious attack of a third-party email vendor. The personal information of T-Mobile customers accessed includes names and addresses, Social Security numbers, financial account information, and government identification numbers, as well as phone numbers, billing and account information, and rate plans and features.

3.3 Zoom

April 14, 2020: The credentials of over 500,000 Zoom teleconferencing accounts were found for sale on the dark web and hacker forums for as little as $.02. Email addresses, passwords, personal meeting URLs, and host keys are said to be collected through a credential stuffing attack.

3.4 J-Crew

March 4, 2020: Hackers successfully accessed online accounts of customers of the apparel retailer, J-Crew, through a credential stuffing attack. Using exposed emails and passwords, the hackers were able to login to an unknown number of J-Crew customer accounts and gain access to stored information including the last four digits of credit card numbers, expiration dates, card types, billing addresses, order numbers, shipping confirmation numbers, and shipment status.

3.5 Paay

April 22, 2020: A card payments processor startup, Paay, left a database containing 2.5 million card transaction records accessible online without a password. The exposed payment transaction belonging to 15 to 20 merchants includes full plaintext credit card number, expiry date, and the amount spent.

3.6 Twitter

June 23, 2020: A security lapse at Twitter caused the account information of the social media company's business users to be left exposed. The number of impacted business accounts has not been disclosed but its business users' email addresses, phone numbers, and the last four digits of their credit card number were impacted.

3.7 General Electric

March 24, 2020: The technology conglomerate, General Electric (GE), disclosed that a third party vendor experienced a data breach, exposing the personally identifiable information of over 280,000 current and former employees. The employee information accessed through Canon Business Process Services included names, addresses, Social Security numbers, driver's license numbers, bank account numbers, passport numbers, and dates of birth.

3.8 Clubillion

July 7, 2020: Popular casino gambling app Clubillion has suffered a data leak, exposing the PII of millions of users around the world according to researchers at vpnMentor. While it was open to searchers, the Clubillion database was recording up to 200 million records a day, including users' IP addresses, email addresses, amounts won, and private messages within the app. The majority of Clubillion's daily users are from the United States.

4 How do data breaches occur?

Data breaches occur when cybercriminals are able to gain unauthorized access to sensitive data. This can be achieved through physical access, or by bypassing security controls remotely. While most data breaches are attributed to cyberattacks or malware, common security threats include insider leaks, identity theft, payment card fraud, loss or theft of physical assets, misconfiguration, and human error. Few of the common ways that data breaches can occur are as follows:

1. ***Vulnerabilities*****,** An exploit takes advantage of software bugs or vulnerabilities to gain unauthorized access to a system or its data.
2. ***Ransomware***, it's a type of malware that denies the access of data till a ransom is paid.
3. ***SQL injection (SQLI):*** A form of cyber attack that exploits a weakness in an SQL database of an insecure website to get the website to give access to information in its database without authorized access.
4. ***Spyware*****:** malware that infects your computer or network to steal personal information, Internet usage, or any other sensitive data it can acquire.
5. ***Phishing*****,** a form of social engineering that aims to manipulate emotions or trick you into revealing sensitive information.
6. ***Insecure password:*** Passwords that are easy to guess, such as dictionary words or common passwords, make it easy for cybercriminals to gain access to sensitive information.
7. ***Physical theft:*** Criminals may steal your computer, smartphone or hard drive to gain access to your sensitive data that is stored unencrypted.

8. ***Third-party vendor breaches:*** Criminals can target third-party business partners or service providers to gain access to large organizations that may have sophisticated cybersecurity standards internally but a poor third-party risk management framework.

5 What Data Breaches Steal?

Hacking or malware were behind 25% of the data breach incidents from 2005 to April 2015, over these years, incidents of payment card data breaches have increased 169%. The healthcare sector was most affected by data breaches, followed by the government and retail sectors. Personally identifiable information (PII) was the most stolen record type and Financial data came in second. Apart from the usual credit card, bank account, and PII dumps—whose prices in the underground have plateaued—there was a prominence of ads selling Uber, PayPal, and poker accounts. type of data that is compromised are

1. ***PII***: Personally identifiable information includes Names, addresses, Social Security numbers, dates of birth, phone numbers, etc.
2. ***Financial data*****:** Banking, insurance, and billing information, etc.
3. ***Health data:*** Hospital and doctors' office records, medical insurance, etc.
4. ***Education data:*** School, college, university, or related records.
5. ***Payment cards:*** Credit, debit, store-branded credit, and prepaid gift cards.
6. ***Credentials*****:** Log-in credentials for eBay, PayPal, Web-based email, online banking, and other accounts.
7. ***Other*** types are: Intellectual property and intelligence about an organization. Unknown types: In many cases, investigators failed to determine what was stolen.

6 Effects of these Breaches

The impact of breach currently is much more damaging when compared to breaches previously. Now a single breach can affect millions, even billions of people and multiple organisations. Evident from the section 2 examples of recent 2020 data breaches. Whenever a breach occurs, not only one individual is affected. Infact every entity surrounding that data will be affected which includes the individual to whom the data belongs, the organisation which stores the data, the third-part vendors or company partners etc. Data breaches hurt individuals by compromising sensitive information.which can often lead to headaches: changing passwords frequently, enacting credit freezes or identity monitoring, and so on. Sometimes these damages can hit us for a long time.

In terms of an organisation the most damaging after-effects of the breach are Financial Losses, Reputational loss, Operational disruption, Intellectual Property loss and legal ramifications. Perhaps the biggest long-term consequence of a data breach is the loss of customer trust.

7 Identifying the Breach

The effects of data breach can be damaging, their impact on victims, their cost, misuse of information and much more. Which makes it an utmost necessity for us to identify it. But How? *We say a breach occurred, when we know that it occurred.* But most of these breaches are unidentified for a long time. And at the time of discovery of these breaches the damages have already happened. There are Reports about data breaches affecting governments, hospitals, universities, financial institutions, retailers, and many other different domains with increasing frequency. But this is merely the tip of the iceberg. But there are the vast majority of incidents that remain unreported and undisclosed.

Data breaches have become an everyday affair, but what is more scary is people are unaware that they are victims of a data breach. There are several factors that result in this. They are as follows, There is an overload of daily news articles on data breaches, Stolen sensitive data is not as tangible as, for example, a stolen mobile phone, The bad consequences of having sensitive data stolen are not instantly felt and There is a lack of understanding of the repercussions of sensitive data theft. And It is impossible to predict if, why, when, where, and how a business or organization will get breached. Breach methods and the data targeted vary across industries and even businesses or organizations within the same industry. Some data breaches are discovered within a matter of hours or days, while others take months or years.

8 Financial Breach

As mentioned in section 5, Financial data is the most common type of data that is stolen from a breach. It can be breach of account details, Payment card details either physically or using digital techniques. For now we are going to focus on breaches of Payment cards and discuss its statistics. Figure2 shows the statistics of Payment card(credit card) data breach incidents recorded from the year 2005 to April 2015. Payment card data breaches exponentially increased from 2010.

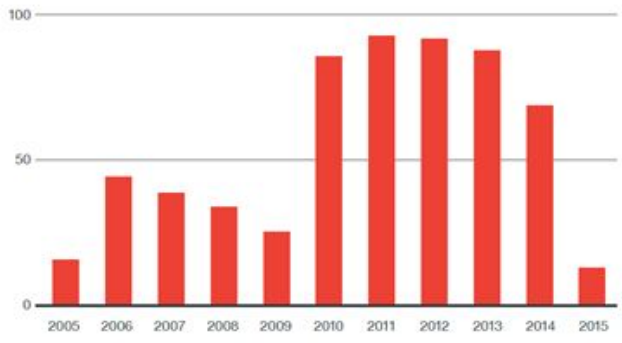

Figure 2: Payment Card Data Breach incident from 2005 to April 2015

Payment card breach, affected not only an individuals holding the card but can affect different organisations or industries Common

techniques for skimming payment cards include, Making a rub of cards, Rigging ATMs or gas pumps with fake panels that steal data, Modifying in-store point-of-sale (PoS) terminals, Using off-the-shelf hardware keyloggers on cash registers. Figure3 depicts the various industries that were affected by the payment card data breach.

Figure 3: Industries affected by payment card data breaches

The techniques which require physical access to the cards or the devices used to process them increases a risk of getting apprehended. Therefore, criminals have resorted to using malicious software like **PoS RAM scrapers** *(PoS - point-of-sale systems used to scan and process credit and debit card transactions*) to steal payment card data, primarily credit card information. A variety of infiltration techniques are used to gain initial entry into and laterally move across the victim's network in order to compromise PoS servers.

Figure4 shows what kind of methods or techniques were used to steal this information. Hacking or malware seems to be the common approach to steal the information.

Figure 4: Payment card data breach methods used

8.1 Our Implementation Project

As we have seen the numerous attacks of data breaches and adverse damages caused by these attacks, The need to identify these breaches increases. We want to take a step towards it. We want to automate the process of identifying these breaches. Financial data being the increasing type of data record that is being stolen. We want to focus our attention in identifying these breaches. We will be implementing a Machine Learning model that could help us automate the process of identifying the breach. For this we will be using credit card transaction information and train a model to identify the fraudulent transactions. Our goal will be to create an unbiased model that could detect the fraudulent transactions with high accuracy.

9 Conclusion

In this paper we have defined what a Data Breach is, the history and its origins, the laws and regulations specifically imposed to prevent data breaches, recent 2020 data breaches and its statistics. We have seen How this breach occurs, what will be stolen in a breach. The adverse effects of a Data breach on an individual and on an organisation, and the necessity to identify them. We later discussed in detail, the second most and the increasing type of data record that is being stolen, which is Financial data record, using the statistics of Payment card fraud, we presented the organisations and industries affected by it and also discussed the common techniques used for Payment card frauds. We further mentioned our steps that we will be taking towards identifying these breaches in an automated way using Machine Learning Algorithms.

We want to conclude the most important takeaway from this is to identify the breach as soon as possible when it happens, to avoid long term damages. Yeah, there is a saying "Prevention is better than cure". But data breaches are something unavoidable, even the most secured networks are hacked. Hackers are finding new sophisticated ways to breach the network and steal the data. In this case we say, making everything ready for treatment is a smart move.

ACKNOWLEDGMENTS

We would like to thank Professor Kristin Tufte, the instructor for this course, for providing advice, ideas, and resources for this project.

REFERENCES

[1] Follow the Data: Dissecting Data Breaches and Debunking Myths Trend Micro Analysis of Privacy Rights Clearinghouse 2005–2015 Data Breach Records Numaan Huq Forward-Looking Threat Research (FTR) Team - A Trend Micro Research Paper https://documents.trendmicro.com/assets/wp/wp-follow-the-data.pdf

[2] Analyzing and Identifying Data Breaches in Underground Forums -YONG FANG , YUSONG GUO, CHENG HUANG , AND LIANG LIU https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8686093

[3] Identity Force A Sontiq Brand https://www.identityforce.com/blog/2020-data-breaches

[4] UpGuard - Breach Statistics https://www.upguard.com/blog/data-breach-statistics#recent

[5] Enterprise data breach: causes, challenges, prevention, and future directions -Long Cheng, Fang Liu and Danfeng (Daphne) Yao https://vtechworks.lib.vt.edu/bitstream/handle/10919/80426/YaoEnterpriseDataBreach2017.pdf?sequence=1&isAllowed=y