future internet

MDPI

Article

Impact, Compliance, and Countermeasures in Relation to Data Breaches in Publicly Traded U.S. Companies

Gabriel Arquelau Pimenta Rodrigues †, André Luiz Marques Serrano †, Guilherme Fay Vergara †, Robson de Oliveira Albuquerque †,* and Georges Daniel Amvame Nze †

Professional Post-Graduate Program in Electrical Engineering (PPEE), Department of Electrical Engineering (ENE), University of Brasília (UnB), Brasília 70910-900, Brazil; gabriel.arquelau@redes.unb.br (G.A.P.R.); andrelms@unb.br (A.L.M.S.); guilherme.vergara@redes.unb.br (G.F.V.); georges@unb.br (G.D.A.N.)
* Correspondence: robson@redes.unb.br
† These authors contributed equally to this work.

Abstract: A data breach is the unauthorized disclosure of sensitive personal data, and it impacts millions of individuals annually in the United States, as reported by Privacy Rights Clearinghouse. These breaches jeopardize the physical safety of the individuals whose data are exposed and result in substantial economic losses for the affected companies. To diminish the frequency and severity of data breaches in the future, it is imperative to research their causes and explore preventive measures. In pursuit of this goal, this study considers a dataset of data breach incidents affecting companies listed on the New York Stock Exchange and NASDAQ. This dataset has been augmented with additional information regarding the targeted company. This paper employs statistical visualizations of the data to clarify these incidents and assess their consequences on the affected companies and individuals whose data were compromised. We then propose mitigation controls based on established frameworks such as the NIST Cybersecurity Framework. Additionally, this paper reviews the compliance scenario by examining the relevant laws and regulations applicable to each case, including SOX, HIPAA, GLBA, and PCI-DSS, and evaluates the impacts of data breaches on stock market prices. We also review guidelines for appropriately responding to data leaks in the U.S., for compliance achievement and cost reduction. By conducting this analysis, this work aims to contribute to a comprehensive understanding of data breaches and empower organizations to safeguard against them proactively, improving the technical quality of their basic services. To our knowledge, this is the first paper to address compliance with data protection regulations, security controls as countermeasures, financial impacts on stock prices, and incident response strategies. Although the discussion is focused on publicly traded companies in the United States, it may also apply to public and private companies worldwide.

Keywords: compliance; countermeasure; cybersecurity; data breach; privacy

check for updates

Citation: Rodrigues, G.A.P.; Serrano, A.L.M.; Vergara, G.F.; Albuquerque, R.d.O.; Nze, G.D.A. Impact, Compliance, and Countermeasures in Relation to Data Breaches in Publicly Traded U.S. Companies. *Future Internet* **2024**, *16*, 201. https://doi.org/10.3390/fi16060201

Academic Editor: Gianluigi Ferrari

Received: 25 April 2024
Revised: 21 May 2024
Accepted: 25 May 2024
Published: 5 June 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Data leakage severely threatens the operations of enterprises, including private corporations and government agencies. The loss of sensitive information can result in substantial reputational damage and financial losses and jeopardize the long-term stability of an organization. Commonly leaked data include employee and customer information, intellectual property, and confidential medical records. Recently, numerous high-profile data breaches have cost companies billions of dollars [1]. This trend has been further amplified by the rapid growth of data in the digital age, making data leaks more frequent than ever, particularly with the rise of Industry 4.0. As an illustration of the scale of cybersecurity threats, in January 2024, 26 billion records were exposed, marking it as the largest known data breach, termed the Mother of All Breaches (MOAB). This breach comprised 12 TB of data leaked from platforms such as LinkedIn, Twitter, Weibo, and Tencent [2]. Thus, preventing the

unauthorized disclosure of sensitive information has become a critical security concern for enterprises.

Confidential information breach represents a data breach, and, as defined by the Code of Federal Regulations, it entails loss, theft, or unauthorized access to data containing sensitive personal information, whether in electronic or printed format, posing a potential threat to the confidentiality and integrity of the data. As stated by Privacy Rights Clearinghouse (PRC) (privacyrights.org/data-breaches accessed on 7 April 2024), in 2021 there were 2266 reported data breaches in the United States alone, leading to the compromise of a total of 93,824,801 records. For each of these records, the Personal Identifiable Information (PII) exposed may have represented a physical-safety or financial concern for the data subjects.

Moreover, companies may suffer financial losses as a consequence of direct costs (e.g., sales disruption, stock price drop, extortion payments) and indirect costs (e.g., reputational damage, reduced credit rating, system downtime) [3]. Indeed, data protection is such a relevant concern that numerous laws and regulations have been enacted worldwide to address it. Furthermore, depending on the data type a company stores, it may be compelled to comply with specific regulations, as outlined by Yimam and Fernandez [4].

Hence, to improve customer privacy, prevent financial losses, and ensure compliance with relevant regulations, companies must make strategic investments in cybersecurity and deploy security measures that adeptly mitigate the risks linked to data breaches. It is, however, critical to do so in a planned and managed manner to optimize the allocation of resources.

This optimization must establish a relationship between specific metrics within the dataset and key cybersecurity features, including governance, risk and compliance, as well as impacts and vulnerability mitigation [5]. Additionally, further analysis of the dataset is conducted to contribute to the understanding of the interrelationship between the risk of data breach and the attributes of the affected company.

In the age of big data, data assumes a key role in a company's core operations, as the effective analysis of vast data volumes offers a substantial competitive edge to corporations. However, this advantage is closely associated with an elevated risk of loss or theft of sensitive and valuable data. Thus, safeguarding against data leakage presents a challenge for businesses, thereby raising significant security concerns. The process of storing, using, sharing, and analyzing escalating data quantities leads to an upsurge in potential vectors for data leakage. These vectors encompass a wide spectrum of avenues, such as cloud file sharing, email, web pages, instant messaging, FTP (File Transfer Protocol), removable media and storage, database and file system vulnerabilities, camera incursions, laptop theft, lost or stolen backups, and vulnerabilities associated with social networks.

This way, to promote the reasoning regarding a data breach within this work, the dataset supplied by Rosati and Lynn [6] was analyzed in terms of various statistical aspects. As related in Section 3, this dataset was filtered to comprehend solely data breaches inflicted on companies listed on the New York Stock Exchange (NYSE) and NASDAQ. This refinement allowed for a more in-depth exploration of the financial implications of these incidents, even though estimating such losses is a complex task [7].

1.1. Contributions and Limitations of the Work

In this work, we assessed the statistics of data exposures that have affected publicly listed U.S. companies and we used the observed patterns and study cases as the foundation of a discussion regarding the following: (i) compliance with data protection regulations in both the geographical and sectoral scopes; (ii) security, technical, and administrative controls applicable to protecting the data while in use, in transit, and at rest for different attack vectors and throughout their entire life cycle: create, store, use, share, archive, and destroy; (iii) guidelines for adequately responding to this kind of incident; (iv) the consequences of an exfiltration for both the data subject and the data owner. To our knowledge, this is the first study to comprehend these four areas.

For this analysis, we have also improved the dataset proposed by [6] with the industrial sector of the target companies. As a consequence, the paper evaluates the frequency and impact of data breaches on different economic sectors, providing information regarding the prevalence of these incidents and advancing the discourse on prioritizing resources allocation. This is presented in Section 4. In Section 5, we review the literature on effective security measures to safeguard sensitive information against the studied attack vectors, providing a roadmap for organizations to enhance their resilience against data breaches. The paper also discusses compliance aspects, reviewing the role of data protection regulations in protecting organizations against data breaches.

Although it did not interfere with the compliance, mitigation, and response discussion, the dataset used was restricted to companies in the U.S., and the statistics presented may not represent the data breach trends worldwide. Additionally, the study's focus on publicly listed companies may not fully represent the broader corporate landscape, as privately held firms may face distinct cybersecurity challenges. Another limitation of this work was the potential under-reporting of data breaches, as not all incidents may have been publicly disclosed.

1.2. Structure of the Work

The structure of the remaining sections in this paper is as follows: Section 2 reviews relevant literature pertaining to data breaches; Section 3 concisely presents the dataset used along with the new attribute proposed; exploratory data analysis is presented in Section 4; Section 5 correlates the observed statistics of the dataset with various cybersecurity factors; Section 6 concludes the paper and outlines potential directions for future research.

2. Related Works

In the literature review, an examination of prior research that has investigated the assessment of data breaches was undertaken, aligning with certain aspects of this study. Reinforcing the importance of the study of data breaches in the United States, Sood and Cor [8] estimated that 82.84% of Americans have had their data breached at some time, with each person being victimized by three breaches on average.

To attain a better perception regarding data breaches, studies have used exploratory data analysis to reveal meaningful observations about this type of security incident. In that regard, Hammouchi et al. [9] used the PCR dataset, in the time range between 2005 and 2018, to assess several data breach metrics, such as inter-arrival time and trend analysis. They found that breaches ensuing hacking activity are the most significant regarding leaked data and financial impact. They also concluded that the most targeted economic sectors were medical and manufacturing/technology/communications. This paper bases its discussion on a similar analysis, although restricted to publicly traded companies and with additional investigations, such as geographical statistics.

Similarly, Raghupathi et al. [10] used a dataset gathered from the U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR) to visualize patterns and trends in data breaches related to medical and health data. They also noticed the noteworthy contribution of hacking-type breaches in companies in this branch. Text mining methods have also been applied in the HHS OCR dataset, using co-occurrences of keywords to identify characteristics, vulnerabilities, impacts, and responses to insider threats in the healthcare industry [11]. Specifically to companies in the medical and health sector, data protection policies and controls implemented must be compliant with the Health Insurance Portability and Accountability Act (HIPAA) and the Health Information Technology for Economic and Clinical Health Act (HITECH).

Due to the sensitiveness of medical data, several authors have studied the risks pertaining to healthcare privacy globally, such as in India [12], in China [13], and in general [14]. In this work, we discuss the compliance and regulatory aspects of data protection applicable in the United States for different types of data, including medical,

and in effect in different U.S. States. Recent studies have demonstrated that compliance with these standards assists in cybersecurity risk management [15].

With respect to risk management and assessment, Algarni et al. [16] stated that cybersecurity risks had not yet been formally addressed in regard to data breaches and proposed a model for estimating the likelihood and cost of a data breach, which are key components for a quantitative risk assessment. They remarked that not using encryption and not involving the Business Continuity Management team in the incident response raise both the cost and the probability of a data breach. Additionally, they observed the increased probability and cost of exposures due to malicious attacks compared to human error and system glitches and that the healthcare/pharmaceutical and technology/software industries endure greater costs. It has also been concluded that the longer the retention time of sensitive data, the higher the cost of data leakage. Bayesian frameworks have also been used in cyber security data breach risk management, with a suggestion for insurers to use this model to price their products [17].

To anticipate data breaches, Barati and Yankson [18] employed Poisson and negative binomial models in the Privacy Rights Clearinghouse dataset. Their findings suggest that their proposed system demonstrates the ability to predict data breach incidents with minimal deviation from the actual numbers.

After a breach has occurred, an inquiry must be conducted to contain it, identify the root cause, and mitigate the incident. Considering these response steps, Masuch et al. [19] studied the impact on companies' stock prices after different approaches adopted by them: either justify the breach, in which case the severity of the incident is minimized, or apologize for it, admitting guilt. The results have shown that a justification typically does not impact the stock market, whereas an apology negatively influences it. Other authors have also studied the reaction of a company's stock prices after it breaches data [1,20,21]. It has been observed that companies subjected to a data breach tend to increase cybersecurity risk factor disclosures, a mandatory notification outlined by the US Securities and Exchange Commission (SEC), especially in the case of larger breaches [22]. We also present guidelines for effectively responding to these incidents and briefly show possible financial consequences on affected companies and the challenges regarding studying stock price reaction to cybersecurity breaches.

Studying the customer's reaction after having their data leaked through a company is also relevant. Lulandala [23] studied the change of behavior of Facebook users approaching advertisements after a data leak. The authors have remarked that these security incidents reduce customers' trust and lower ad engagement.

Molitor et al. [24] conducted research to identify the key terms and legal implications of data breaches through Machine Learning (ML) and text mining in litigation cases. The results showed that the litigants were concerned about significant topics such as identity theft, hackers, negligence, insurance, phone devices, credit cards, and privacy.

Schlackl et al. [25] reviewed relevant academic papers about data breaches, emphasizing their antecedents and consequences, summarizing the literature on what influences a data breach and the subsequent repercussions. Schlackl et al. [25] also indicated which fields are more studied than others. Equivalently, Patterson et al. [26] systematically reviewed the literature on data breaches and identified future research, pointing out that organizations have not fully maximized the potential benefits of learning from incidents and have not conducted thorough evaluations to determine the effectiveness of their learning processes. Ref. [27] also reviewed the literature on data breaches, focusing on risk management and briefly reviewing data breach causes, prevention, containment, and recovery.

Addressing data exfiltration countermeasures, Ullah et al. [28] reviewed 108 papers, mapping the countermeasures to different attack vectors from malicious external threats, and they concluded that there is no emphasis in research on the data in rest and in transit states and on investigative countermeasures. In this work, we promote discussion on all data states and on the improvement of accountability and digital forensics, encompassing both internal and external threats.

Aslam et al. [29] discussed countermeasures and compliance of data breaches, mainly centered on Smart Cities and the Internet of Things (IoT), presenting a taxonomy of prevention and the consequences of these incidents, seen in Figure 1. This work addresses all prevention taxonomy elements, while focusing on the consequences for asset prices in the impact part. The impact on customers is discussed in specific study cases and in the works reviewed. Also, despite the emphasis of this paper on publicly listed U.S. companies, its contributions and the promoted discussion apply to public and private organizations worldwide.

Figure 1. Taxonomy of data breaches, in which the red boxes represent the fields explored in this work. Adapted from [29].

Data Breach Datasets

Other authors have contributed to the field by publishing datasets focused on data breaches. For instance, Neto et al. [30] created a dataset encompassing global data breach incidents that transpired between 2018 and 2019. Among their findings, they observed a significantly higher frequency of data breach incidents in Europe, with approximately 160,000 reported incidents during that period, in contrast to the United States, where there were approximately 10,000 such incidents.

Considering the sensitiveness of medical data breaches, Ronquillo et al. [31] published a dataset comprised of healthcare data breaches in the United States, observing that, in this sector, hacking activities were responsible for roughly 25% of the incidents but compromised nearly 85% of the records. This indicates that hack data breaches caused an excellent average of records leaked per incident in this dataset.

Park [32] provides a dataset on data breaches that occurred between 2012 and 2016 in California, including information such as whether the company was sued, the duration of time the free credit monitor service was provided to the affected customers, the economy sector of the breached company, the attack vector of the incident, and the size of the breach.

Data breaches and ransomware attacks that occurred in Australia between 2004 and early 2020 were provided by Tsen et al. [33], along with information on technical and administrative countermeasures employed by the affected organizations, such as the use of encryption, established security policies, and improper network segmentation. Also, in Australia, Biddle et al. [34] conducted a survey to understand better the Australian public's attitudes towards data governance, including the level of concern about data breaches. Additionally, Ikegami and Kikuchi [35] proposed a probabilistic model that estimated a

given company's risk of a data breach. They referenced two datasets associated with data breaches in Japan from 2005 to 2018.

The United Kingdom's Information Commissioner's Office publishes information about data breach cases quarterly (https://ico.org.uk/about-the-ico/our-information/complaints-and-concerns-data-sets/self-reported-personal-data-breach-cases/ accessed on 7 April 2024). Likewise, the United States Government's open data website supplies a dataset covering data breach incidents that affect at least 500 Washington State residents (https://catalog.data.gov/dataset/data-breach-notifications-affecting-washington-residents accessed on 7 April 2024).

These datasets, however, either fall outside the geographical scope of this paper or are overly constrained for our research purposes. At the time of this writing and to the best of our knowledge, 12 other works have referenced the work of Rosati and Lynn [6]. These studies are briefly presented in Table 1.

The dataset offered by Privacy Rights Clearinghouse, while more comprehensive in terms of the types of companies breached and the time frame covered compared to Rosati and Lynn [6], was deemed excessively broad in scope for the focus of this paper, which centered on publicly listed companies. Furthermore, the new fields incorporated by Rosati and Lynn [6] into the dataset greatly enhanced our comprehension of the risks associated with these incidents.

This paper adopted the dataset published by Rosati and Lynn [6]. This choice was primarily based on the dataset's concentration on data breaches in publicly traded U.S. companies over an extended time frame and the new fields added by the authors, which aligned with the scope and objectives of our research.

Table 1. Publications that cited the work of [6].

Reference	Description
[36]	Assesses the effects of cybersecurity investments within the realm of data breaches.
[37]	Explores the connection between cybersecurity performance influencing factors that lead to human errors and the subsequent occurrence of data breaches.
[38]	Investigates the impact of non-financial disclosure laws obligations to breached companies.
[39]	Investigates the responses of organizations to data breach incidents.
[40]	Presents statistical distributions of malicious and negligent data breaches.
[41]	Assesses the implications of implementing an inverted firm strategy with the use of Application Programming Interfaces (APIs), highlighting the elevated risk of data breaches associated with the use of public APIs.
[42]	Proposes a cybersecurity risk-quantification-and-classification framework designed for application to real-world data breaches.
[43]	Examines the compliance strategies that IT business leaders within nonprofit organizations should adopt to reduce the risk of cyber threats that could potentially lead to data breaches.
[44]	Reviews predictive algorithms and analyzes trends of breach incidents.
[45]	Performs a brief exploratory data analysis of the dataset.
[46]	Studies the effectiveness of K-Means Featurization (KMF) in addressing the challenges presented by complex datasets.
[47]	Examines how earnings disclosures influence cyberattackers' behavior.

3. The Dataset

The dataset used in this work, published by Rosati and Lynn [6], was originally retrieved from the Privacy Rights Clearinghouse repository. The authors subsequently filtered the PRC dataset to encompass data breach events that impacted companies listed on the NYSE or NASDAQ exclusively. Therefore, the work of Rosati and Lynn [6] published and described the dataset, which this study analyses and discusses.

According to PRC, the data breaches were primarily gathered from U.S. Attorneys General and the Department of Health and Human Services, and they do not constitute an exhaustive list of all breaches, reflecting only the data breaches that have been reported

and made publicly available in the United States. The list is even less exhaustive in the case of this work, as it comprehends publicly listed U.S. companies exclusively. Nevertheless, the dataset was deemed suitable for the objectives of this study, which were to estimate breach sizes and explore various security facets related to data breaches within NYSE- and NASDAQ-listed companies.

The dataset is presented in a tabular format, in which each column contains specific information about these events. In total, the table comprises 506 rows, each representing a data breach, and 15 columns. To enhance the analysis conducted, the dataset was enhanced with information regarding the economic sector in which each company operates, using Yahoo! Finance API. This augmentation allowed for a more comprehensive examination of the factors pertinent to data breaches. However, for this paper, only nine of the original columns were used, with the addition of the economy sector column. A description of these 10 dataset fields is presented in Table 2.

Table 2. Description of the dataset fields studied in this work. The industry_sector field was created in this work. Adapted from [6].

Field	Description	Possible Values
event_ID	Unique event identifier.	[1, 2, 3, . . . , 506]
ticker	Ticker of the targeted organization.	Examples: AAPL, CAKE
event_date	Date of occurrence of the data breach.	Example: 21/06/2014
confound_dum	Whether the affected company made any other announcement in the 7 days prior to the breach announcement.	0: no announcement
		1: announcement made
confound_type	Type of announcement, if any.	Earnings: earnings Investigation: regulatory investigation IPO: Initial Public Offering M&A: merger or acquisition Restatement: restatement of previously issued statement Statement: release of financial results Other: other type of announcement
breach_size	Number of records breached.	Example: 930,000
breach_type	Attack vector.	CARD: fraud involving payment cards HACK: hacked by a malicious party INSD: malicious insider PHYS: paper documents that are lost, discarded, or stolen PORT: portable device lost, discarded, or stolen STAT: stationary computer loss, inappropriately accessed, discarded, or stolen DISC: unintentional disclosure UNKN: unknown cause
event_state	State in the U.S. where the incident took place.	Example: New York
hq_state	The location, potentially outside of the USA, where the headquarters of the impacted company is situated	Examples: Texas, Tokyo
industry_sector	Economy sector in which the affected company operates.	Examples: Financial, Energy, Healthcare

4. Exploratory Data Analysis

This section presents several remarks concerning the dataset provided by Rosati and Lynn [6].

4.1. Geographical View

As outlined in Section 3, the dataset pertains to data breaches reported in the United States, and, therefore, the geographical analysis is restricted to this country.

As noted in Figure 2, the States that experienced the highest number of reported breaches were California, with 79 breaches, and New York, with 75. Conversely, during the period covered by the dataset, there were no reported breaches in the following States: Hawaii, Mississippi, Montana, New Mexico, North Dakota, South Dakota, West Virginia, and Wyoming.

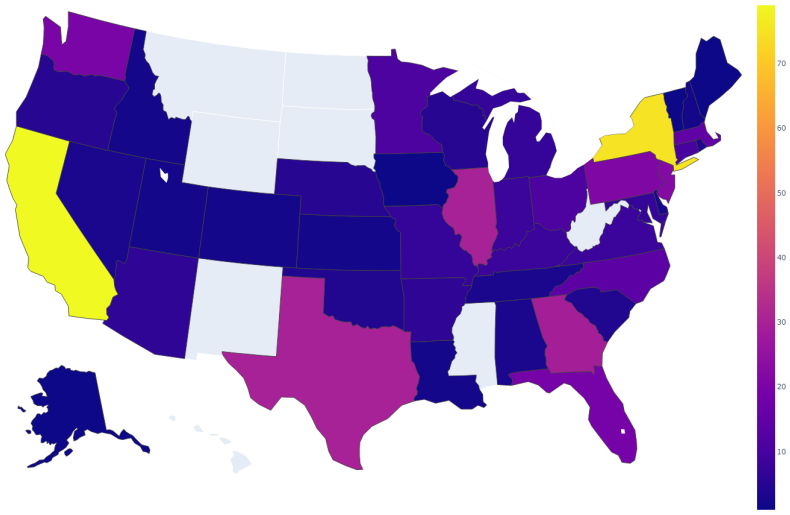


Figure 2. Geographical distribution of count of reported breaches by occurrence State.

Among the States where breaches occurred, Alaska, Iowa, Maine, and Vermont registered the fewest incidents, with only one occurrence each. The Maine exposure transpired in 2012 and targeted New York State Electric & Gas (NYSEG) and Rochester Gas and Electric (RG&E), subsidiaries of Iberdrola USA. This breach resulted in the disclosure of 5100 records containing Social Security numbers, dates of birth, and financial institution account numbers [48]. In response to the breach, the company later offered a credit-monitoring-assistance membership (https://oag.ca.gov/ecrime/databreach/reports/sb24-22146 accessed on 7 April 2024).

Consistent with the choropleth map shown in Figure 2, the distribution of breach types among the 10 most affected States, as depicted in Figure 3, was predominantly led by California and New York. This chart also shows similar breach-type proportions in the States. It is remarked, however, that while hacking breaches represented a significant share of exposures in these States, New Jersey had only one reported case of this type.

This case was the Heartland Payment Systems data breach that took place in 2009. This attack was orchestrated by international hackers and involved a U.S. Secret Service informant who returned to his criminal life as a hacker and exploited an SQL injection vulnerability [49]. The company waited a year to announce the disclosure of the credit card information of 130 million customers, causing a drop in its stock price by almost 80% [50].

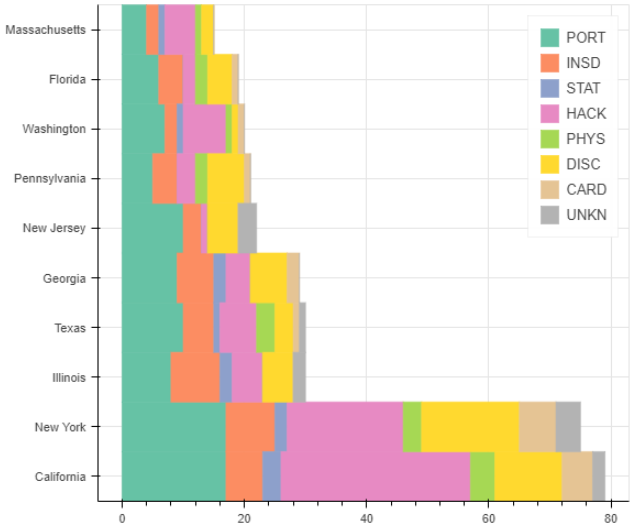


Figure 3. Distribution of breach type per occurrence State.

4.2. Companies

The dataset comprises 506 data breaches distributed among 274 unique companies. Figure 4 represents the 10 companies that experienced the highest number of violations.

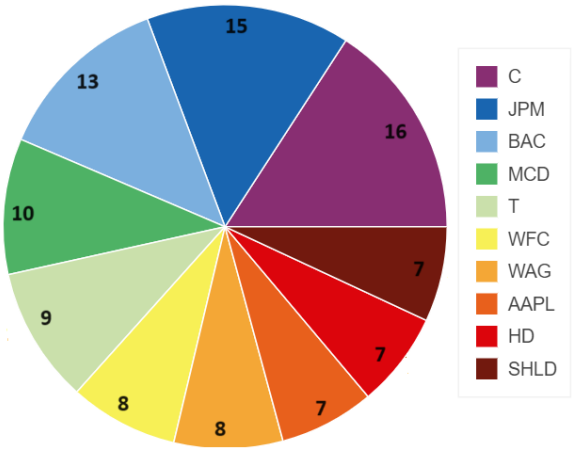


Figure 4. Count of breaches in the 10 most-breached companies (ticker).

A translation between the company ticker and its name is provided in Table 3, which also indicates the company sector. Among the 10 companies that experienced the most breaches, 4 operated within the financial sector. As outlined in Section 4.5, this is the most targeted economic sector.

With the aim of identifying areas of increased vulnerability within these companies, Figure 5 illustrates the various incident types for each corporation. Consequently, it is worth noting, for example, that while Sears Holding had a somewhat even distribution of types, all security breaches at Apple were attributed to hacking activities, while a significant proportion of McDonald's disclosures resulted from malicious insider actions. Furthermore, debit card and credit card fraud incidents were only observed in financial companies, and stationary computer loss, inappropriately accessed, discarded, or stolen, successfully breached Walgreens only.

Table 3. Dictionary of most-breached-companies names from their tickers.

Ticker	Name	Sector
C	Citigroup Inc.	Financial
JPM	JPMorgan Chase & Co.	Financial
BAC	Bank of America Corp.	Financial
MCD	McDonald's	Consumer Cyclical
T	AT&T Inc.	Communication
WFC	Wells Fargo & Co.	Financial
WAG	Walgreens	-
AAPL	Apple	Technology
HD	Home Depot	Consumer Cyclical
SHLD	Sears Holding Corp.	-

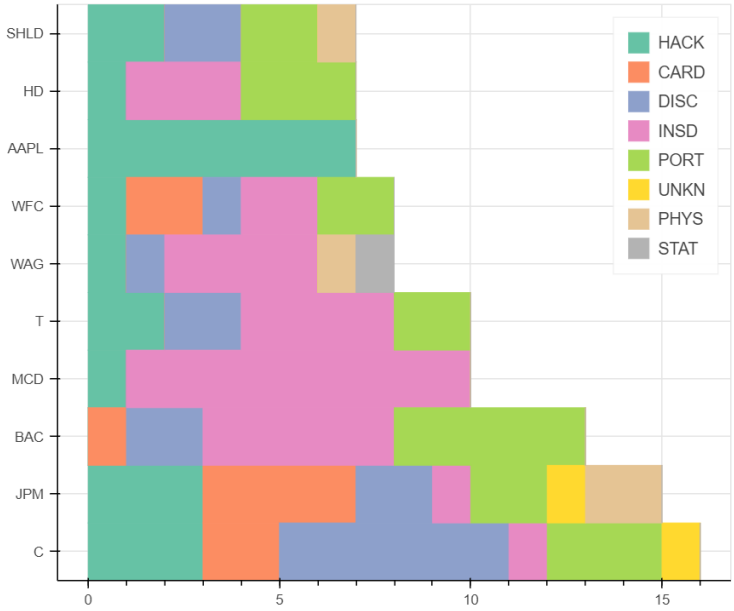


Figure 5. Breach types in the most-breached companies.

4.3. Breach Sizes

The total sum of records breached in the dataset, which considered exclusively publicly traded companies, resulted in roughly 1.07410×10^9, which was approximately 3 times the population of the country in 2013, at around 3.16128×10^8 people [51]. This may have been due to the breach of data of deceased people and to the same person's data being leaked multiple times.

The 10 most significant breaches, in terms of records leaked, in the dataset and their corresponding types are depicted in Figure 6. A dictionary between each of these affected companies' ticker and their name is provided in Table 4.

In the case of the LinkedIn incident, which was the most voluminous data exposure examined within this study, it was reported that unsalted SHA-1 hashed passwords were leaked. However, limited information is available regarding the specific method by which the data were stolen [7]. Poornachandran et al. [52] were able to successfully crack approximately 2.5% of these passwords.

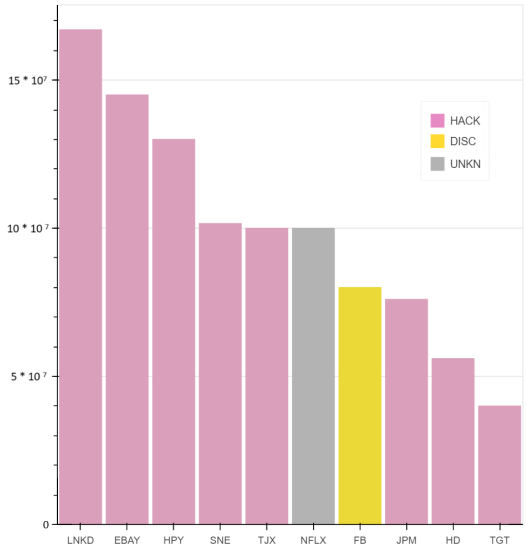


Figure 6. The 10 biggest breaches in the dataset, and their types.

Table 4. Dictionary of companies with the biggest breaches, relating to their names, tickers, and breach date.

Ticker	Name	Date	State
LNKD	LinkedIn.com	29 May 2012	CA
EBAY	eBay & Co.	21 May 2014	CA
HPY	Heartland Payment	20 January 2009	NJ
SNE	Sony	26 April 2011	NY
TJX	TJ stores	17 January 2007	MA
NFLX	Netflix & Co.	01 January 2010	CA
FB	Facebook	17 July 2008	CA
JPM	JPMorgan Chase & Co.	27 August 2014	NY
HD	Home Depot	02 September 2014	GA
TGT	Target Corp	13 December 2013	MN

In a noteworthy incident in 2014 that occurred in California, a spear-phishing campaign effectively compromised approximately 145 million records stored by eBay [53]. As outlined by the author, this breach exposed sensitive data, including customer names, email addresses, physical addresses, phone numbers, and birth dates, all in unencrypted plaintext, resulting in an estimated cost of $300 million to the company.

Minkus and Ross [54] demonstrated the severity of this breach, highlighting that, at that time, it enabled the adversary to retrieve the complete purchase history for a known username. As a consequence, it facilitated the identification of buyers of sensitive items, such as firearms and pregnancy and HIV tests.

Other cases present in Figure 6 that affected the Heartland Payment Systems, Home Depot, and Target cases are discussed in Sections 4.1 and 5.2.1.

Figure 7 illustrates the cumulative sum of breach sizes for each announcement type and the corresponding case counts. This analysis allows for a better understanding of how the timing and nature of announcements relate to the scale of data breaches. Rodrigues et al. [45], however, suggest that previous announcements do not significantly impact the amount of data leaked.

While only 185 out of the 506 breaches were preceded by an announcement, it is noteworthy that the most significant announcement type was 'Earnings'. This observation could indicate a potential financial motive by the attackers. Additionally, it is worth highlighting that despite a lower case count for 'Statement' announcements, the breach

sizes associated with this category were more substantial compared to those related to 'Investigation'.

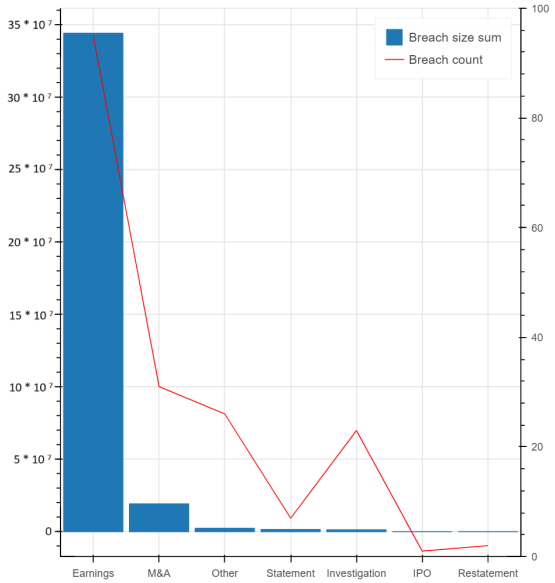


Figure 7. Accumulated number of records breached per announcement type.

4.4. Breach Types

As observed in Figure 8, the most prevalent breach types are PORT (139 events) and HACK (118), collectively accounting for 50.79% of the incidents within the dataset. This understanding reinforces the significance of these breach types in the overall landscape of data breaches. Cases with unknown causes (UNKN) may indicate either an ineffective forensic investigation or a lack of transparency.

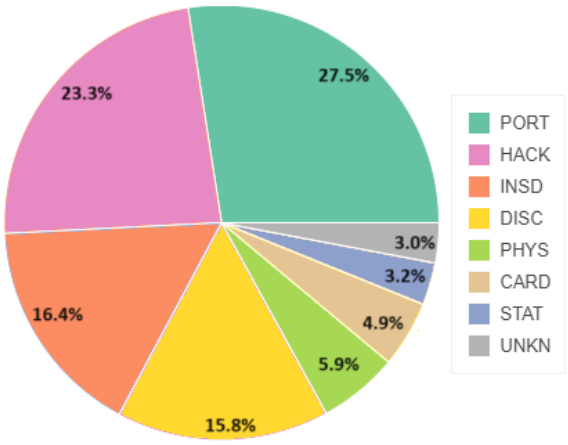
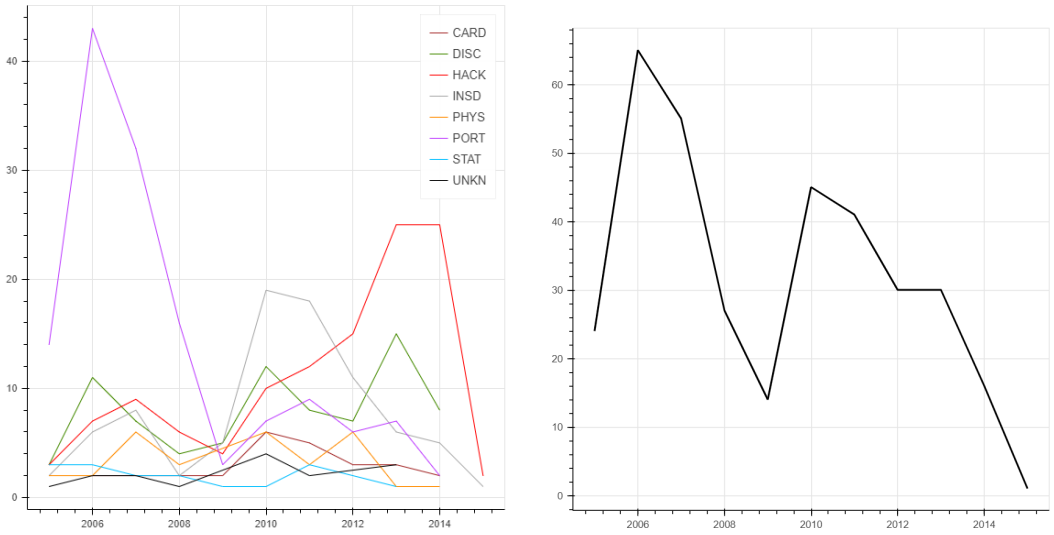


Figure 8. Count of breaches by type.

Determining whether the problem of data breaches is worsening and identifying the prevailing trends are critical concerns. It is important to emphasize that these trends are not always immediately apparent. Thus, there is a need for rigorous data analyses to ascertain whether any discernible trends exist. Moreover, when possible, such studies can help make predictions of the trajectory of data breaches.

To improve the understanding of the Tactics, Techniques, and Procedures (TTPs) employed by attackers in data breaches and their evolving trends over time, Figure 9a provides a visualization of the total breaches reported by year for each breach type. When correlating Figures 8 and 9a, it becomes evident that while PORT breaches are the most frequent category of disclosures, HACK attacks have grown significantly since 2012.



(**a**) By type (**b**) By human factor causes

Figure 9. Count of breaches per year.

The disclosure of data as a consequence of PORT TTPs was more prominent during the early years of the dataset, with 2006 standing out, especially with 43 breaches in this category. This number was nearly four times greater than the second-most-relevant type, DISC, which had 11 breaches. These trends offer relevant information about attackers' shifting tactics and priorities over time.

The statement that humans are often considered the weakest link in cybersecurity is well-documented. Building upon this, Hammouchi et al. [9] concluded that breaches originating from human factors were on the decline, potentially due to increased awareness among personnel. A similar analysis was carried out on companies listed on the NYSE and NASDAQ to validate this assertion in this scenario.

In this context, we focused on breach types INSD, PHYS, PORT, STAT, and DISC, which are associated with human factors. Figure 9b displays the cumulative count of breaches related to these types over the years. This analysis led us to conclude that human involvement in data breaches is diminishing in publicly listed companies. However, it is important to note that this reduction in cases may also be influenced by the decline in PORT cases, as previously discussed, which were particularly numerous in 2006.

In Section 5.2, we further explore potential causes for the declining trend in human-related data breaches. Additionally, we present a more comprehensive discussion of mitigation strategies specific to each breach type. This analysis provides a deeper understanding of the factors influencing the reduction in such breaches and reviews best practices on how organizations can effectively address these vulnerabilities.

4.5. Company Sector

By combining data breach type with the sector in which a company operates, it is possible to verify whether a predisposition of a TTP exists when targeting a specific economic sector. Although no preference for attack vector is evidenced by Figure 10, which shows a similar distribution of types in the company sectors, some other observations are made. For instance, it is noted that the healthcare industry was not targeted by any hacking activity in the time frame covered by the dataset, nor were companies in the industrial sectors leaked via paper documents (PHYS). These observations provide valuable context regarding the distribution of breach types within different economic sectors and may help inform security strategies for these industries.

Furthermore, a more robust relationship is found between fraud involving debit/credit cards (CARD), which is the second-least-frequently-known breach type (Figure 8), and financial companies. This correlation aligns with the nature of financial operations, demonstrating a concentration of this specific exposure type in the financial sector. Differently,

attacks targeting stationary devices (STAT), the least-frequently-known breach type, were primarily present in the four most-breached sectors: namely, Financial Services, Consumer Cyclical, Industrials, and Technology.

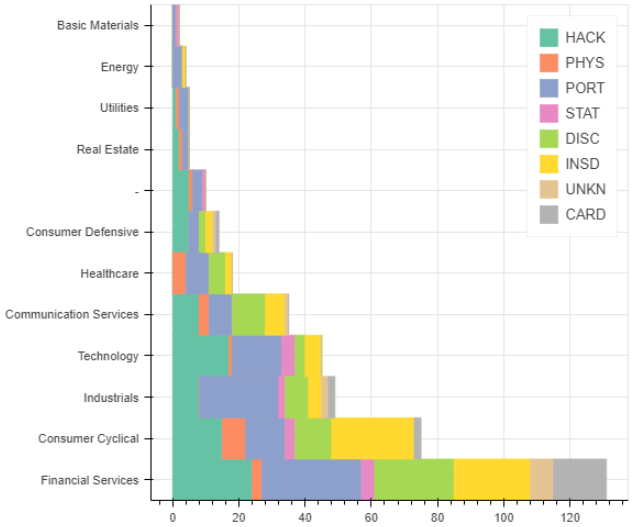


Figure 10. Distribution of breach type per company sector.

These findings contrast with those of Hammouchi et al. [9], who, in their analysis of the PCR dataset from 2005 to 2019, noted that the most-targeted companies were in the sectors of healthcare and manufacturing/technology/communications, categorized as "BSO" by PRC, due to the sensitiveness of the data they hold. However, Figure 10 shows that when restricting this analysis to the publicly listed companies in the dataset this scenario changes, and that financial companies, which also hold sensitive data, are the most breached.

Figure 11 illustrates that the substantial number of breaches in this sector, shown in Figure 10, was not concentrated in a specific period but was relatively constant throughout the time contemplated in the dataset. This provides valuable information about the persistence of security challenges faced by companies in the financial sector. In Figure 10, the string "-" represents companies to which the Yahoo! Finance API could not identify the pertaining sector.

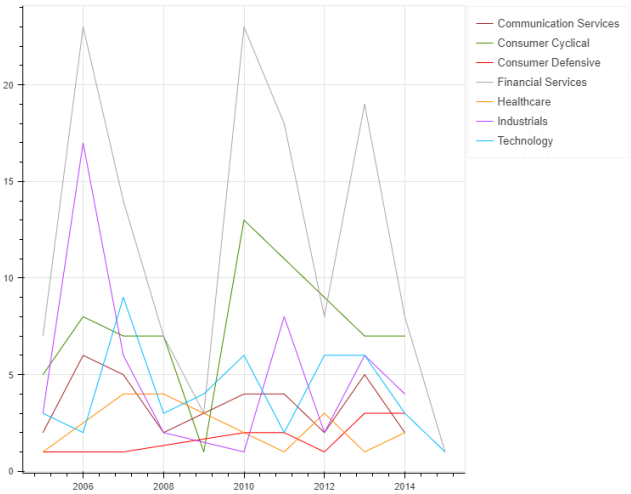


Figure 11. Count of breaches by company sector per year.

It is important to emphasize that the decrease in the number of cases in 2015 was due to the incompleteness of the data for that year, which only went up to March. On the other

hand, the breach cases for the 'Industrials' companies were more frequent in 2006 and have demonstrated a descending trend since then.

An association between the company sector and data protection standards is pertinent. This connection arises from the close relationship between a company's industry and the type of data it stores, influencing the data owner's responsibilities and obligations. This relationship and an overview of compliance will be further explored in Section 5.1.

4.6. Stock Market

Data breaches can have significant implications for the stock market, often leading to fluctuations in the share prices of affected companies. When a data breach occurs, investor confidence may fade due to concerns about the company's ability to protect sensitive information, resulting in a decline in share value.

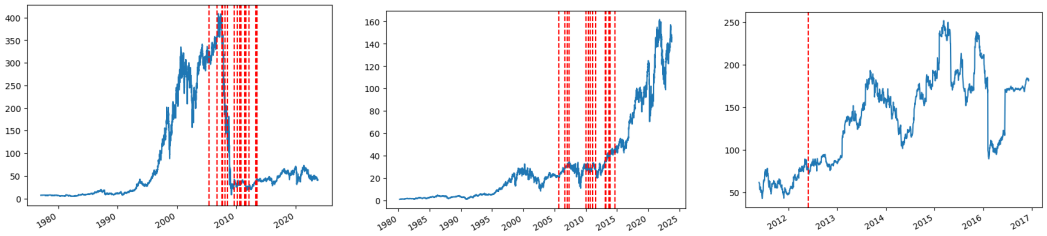
As examples, we mention the historical stock price trajectories of three noteworthy companies within the dataset: Citigroup Inc., which endured the highest number of breaches; JPMorgan Chase, the second-most-affected company; and LinkedIn.com, which suffered the most voluminous breach in the dataset. They are represented in Figure 12, and it is possible to observe different market reactions to these data exposures, allowing for the observation of varying stock price responses to these data exposures.

While, in some cases, a data breach can indeed contribute to a stock price drop, other factors can also be at play. A company's stock price decline can result from various factors, including data breaches, financial crises, regulatory modifications, and broader macroeconomic dynamics. Because of that, it is essential to emphasize that the connection between data breaches and stock price depreciation is not straightforward. In certain instances, a data breach can trigger a decline in share value by diminishing investor confidence in the company. This may be due to concerns that the company is incapable of adequately safeguarding the sensitive information of its customers and employees, or due to the apprehension that the breach might inflict financial or reputational harm.

Additionally, poor management practices following a data breach, such as inadequate responses or compromised data security measures, can further exacerbate the situation and contribute to declines in share prices.

Conversely, there are scenarios in which a drop in stock price can be attributed to other factors, such as financial crises or regulatory alterations. Likewise, if a company confronts new regulatory requirements, investors may opt to sell their shares amid concerns that they might struggle to comply with the updated rules. Understanding the nature of these relationships is essential when evaluating the impact of data breaches on a company's financial standing and its effect on share prices.

Moreover, it is vital to consider that inadequate management practices can set a detrimental cycle that exacerbates crises and precipitates a decline in share values. These decisions, like implementing cost-cutting measures that compromise data security, can potentially catalyze new data breaches, further worsening the financial situation and resulting in a subsequent dip in share prices.



(**a**) Citigroup Inc. (**b**) JPMorgan Chase & Co. (**c**) LinkedIn.com

Figure 12. Historic stock prices (USD) of the two companies with the most incident occurrences (**a**,**b**), and the breach with the most records exposed (**c**). Vertical dashed lines represent a data breach event.

5. Discussion

In this section, we discuss some compliance and security aspects regarding the observed statistics based on the data. This includes a review of relevant laws and standards applicable to the companies included in the study, exploring security controls to mitigate the underlying causes of breaches, and examining response strategies employed when incidents occur.

5.1. Compliance and Pertinent Legislation

Data protection laws are regulations enacted to protect people's privacy, identity, reputation, and autonomy. These regulations have a global presence and include significant standards such as the European General Data Protection Regulation (GDPR), Brazil's General Law for the Protection of Personal Data (LGPD), and China's Personal Information Protection Law (PIPL) [55]. This legislative development reflects the evolving data protection and privacy regulation landscape globally.

Notably, the United States currently lacks a federal-level enacted data protection law. There is, however, ongoing discussion in Congress regarding a proposed bill known as the American Data Privacy and Protection Act (ADPPA) [56].

5.1.1. Standards by Sector

Nonetheless, the United States has a data protection framework for various economic sectors, as outlined in Table 5. This table also details the number of breaches in the dataset that are associated to each law or standard.

It is worth noting that the U.S. has additional laws regulating data protection, but these pertain to data categories that fall outside the scope of the dataset under study. For instance, laws like the Family Educational Rights and Privacy Act (FERPA) pertain to students' data, the Driver's Privacy Protection Act (DPPA) relates to driver's records, and the Federal Information Security Modernization Act (FISMA) applies to government data. Furthermore, laws such as the Children's Online Privacy Protection Act (COPPA) protect children's data. However, the dataset does not provide sufficient information to determine whether these laws were infringed in the reported breaches.

Table 5. Data protection standards and their applicability to the breaches in the dataset by type of data.

Act/Standard	Applicable to	# (%) Dataset Figure 10
SOX	Publicly traded companies	506 (100%)
GLBA	Financial	131 (25.89%)
Telecom. act	Communication	35 (6.92%)
PCI-DSS	Credit card data	25 (4.94%)
HIPAA	Healthcare	18 (3.56%)

5.1.2. Data Breach Notification Acts

In addition to those standards, all 50 States have established laws mandating private enterprises to inform individuals regarding security breaches that compromise personally identifiable information [56]. The majority of them also include government entities in the commitment.

However, some of those laws were not in effect during the time frame observed in the dataset. For instance, Alabama's Breach Notification Act, the most recent State law, was enacted in 2018. Similarly, South Dakota (2018) and New Mexico (2017) introduced their respective laws after the temporal scope of the dataset. Thus, their areas are depicted in gray in Figure 13, which displays the chronological order of the start of the effect of breach notification laws.

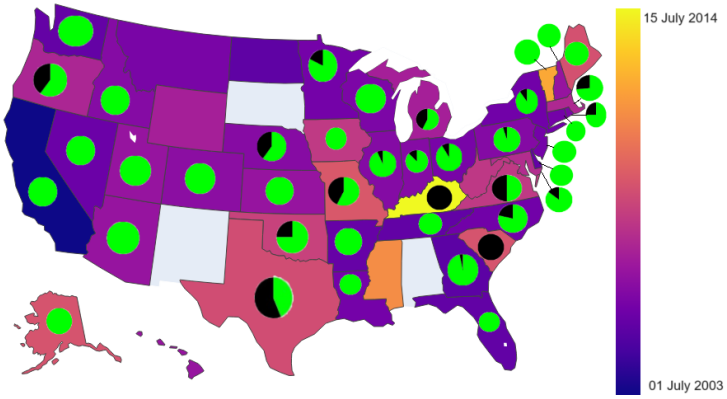


Figure 13. Breach notification statuses in U.S. States. The color map indicates the period in which the notification breach law took effect, and the pie charts indicate the distribution of breaches that occurred before (black) and after (green) the law took effect.

In this Figure, the pie charts denote the number of breach incidents that occurred before (black) and after (green) the commencement of the law's effect. States without a pie chart either experienced no breaches or had not enacted notification laws within the dataset's time frame. California was the inaugural U.S. State to have a breach notification Act in effect, from 1 July 2003.

The work of Coie [57] offers a comprehensive breakdown of the specifics of each State Act. It details critical aspects such as their definitions of data breaches and personal information, the timing and structure of breach notifications, their applicability, and other vital provisions that form the foundation of these State-level data breach notification laws.

5.1.3. Data Protection Acts

Not all states have enacted data protection laws, and some have yet to take effect, as seen in Table 6. These data may also be geographically visualized in Figure 14, helping illustrate the varying degrees of legislative coverage across different States.

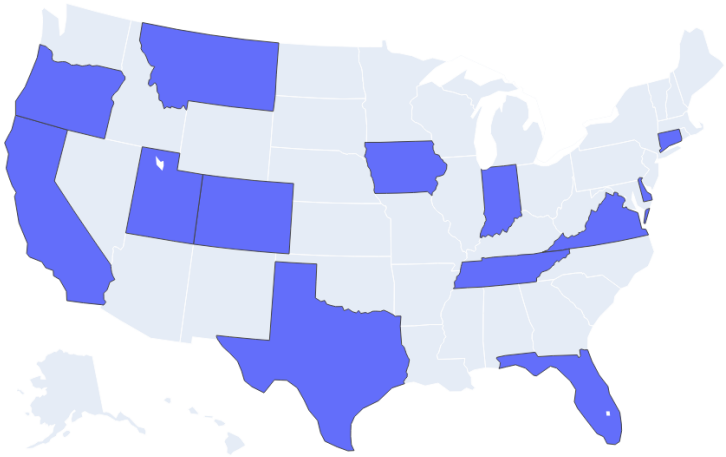


Figure 14. US States that have enacted a data protection law. None of these acts were in effect in the dataset time range.

It is important to highlight that the dataset under examination covers the period between 2005 and 2015 and that the first data protection law took effect in 2020. Therefore, none of the breached data analyzed in this paper were subject to regulation by a data protection law. Another observation is that New York, the second-most-breached State in the dataset, has not enacted any data protection law. In contrast, Montana did not have any reported breach in the analyzed data and has enacted the Montana Consumer Data Privacy Act (MTCDPA). However, it has not yet commenced its effects.

Table 6. State comprehensive data protection Acts and their applicability to the breaches in the dataset by type of data. None of these Acts were in effect in the dataset time range.

Act	State	Effect Date	# (%) Dataset Figure 2
CCPA	CA	01/01/2020	79 (15.61%)
CPRA	CA	01/01/2023	79 (15.61%)
VCDPA	VA	01/01/2023	8 (1.58%)
CPA	CO	07/01/2023	2 (0.4%)
CTDPA	CT	07/01/2023	10 (1.98%)
UCPA	UT	12/31/2023	2 (0.4%)
OCPA	OR	07/01/2024	5 (0.99%)
TDPSA	TX	01/01/2024	30 (5.93%)
FDBR	FL	07/01/2024	19 (3.75%)
MTCDPA	MT	10/01/2024	0 (0%)
ICDPA	IA	01/01/2025	1 (0.2%)
DPDPA	DE	01/01/2025	3 (0.59%)
TIPA	TN	07/01/2025	3 (0.59%)
Indiana CDPA	IN	01/01/2026	8 (1.58%)

Elements of the California Consumer Privacy Act (CCPA) include the rights of the data subjects, encompassing the right to be informed about the nature of collected data and its selling practices, to request data deletion, to opt out of data sales, to access their data, and to not be discriminated against in service and pricing when exercising their privacy rights.

5.2. Vulnerabilities Mitigation

To ensure the proper protection of customers' data and compliance with these regulations, a company must effectively implement security controls that mitigate the vulnerabilities that could lead to a breach. Different incident types require different security countermeasures as mitigation.

As seen in Figure 15, data breaches in 2023 were most frequently caused by phishing attacks, and the most costly originated with malicious insiders (INSDs) [58]. However, when categorizing these attack vectors into the types of the dataset, it was observed that the majority of them were related to hacking activities (HACK). The 'Accidental' label in the figure may be associated with the union of the DISC, PHYS, STAT, and PORT types of the dataset, while 'Physical' may be associated with the union of PHYS, STAT, and PORT. Social engineering and phishing, which may be considered a type of social engineering, are not strictly related to a category of the dataset.

This also complements the analysis of the dataset, informing the financial impact for different attack vectors, which is important to consider in a risk assessment and in prioritizing vulnerabilities mitigation. As an example, according to Section 4.4, HACK data breaches represent the second-most-frequent cause, and also correspond to significant average costs, as seen in Figure 15. This suggests a prioritization for mitigating HACK-related vulnerabilities.

This section reviews some suitable security controls that can mitigate these and other vectors. Such countermeasures could have reduced the likelihood and/or the impact of the breaches in the scope of this study.

To improve overall security, companies may adopt a structured framework, such as the NIST Cybersecurity Framework (CSF), which is an agnostic framework categorizing several security controls in five cores, as shown in Table 7. These core categories provide an organized approach to improving cybersecurity measures.

Especially for vulnerability mitigation, the Protect core function presents some valuable recommendations divided into categories: Identity Management, Authentication and Access Control, Awareness and Training, Maintenance, Protective Technology, Information Protection Processes and Procedures, and Data Security.

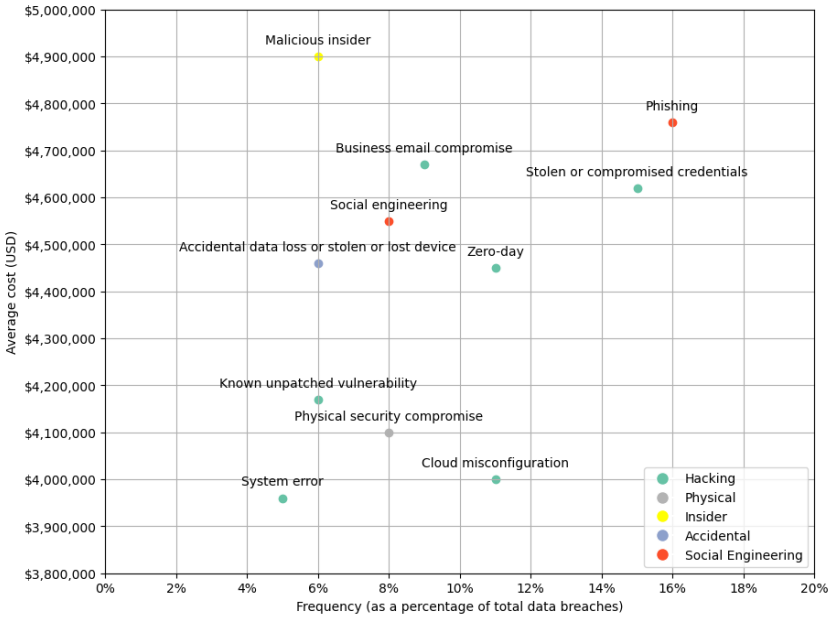


Figure 15. Frequency and average cost of initial attack vectors responsible for data breaches in 2023. Adapted from [58].

The latter, more pertinent to this study, is then divided into subcategories, which are as follows: data at rest are protected; data in transit are protected; assets are formally managed throughout removal, transfers, and disposition; adequate capacity to ensure availability is maintained; integrity-checking mechanisms are used to verify software, firmware, and information integrity; the development and testing environments are separate from the production environment; integrity-checking mechanisms are used to verify hardware integrity; and protections against data leaks are implemented.

Once again, the latter subcategory is more relevant to this work, and the NIST CSF references the Center for Internet Security (CIS) Critical Security Control, COBIT 5, ISA 62443-3-3:2013, ISO/IEC 27001:2013, and NIST SP 800-53 [59–61].

Specific to the NIST SP 800-53, which reviews security and privacy controls for information systems and organizations, the framework mentions the sections regarding information flow enforcement (AC-4), separation of duties (AC-5), the principle of the least privilege (AC-6), personnel screening (PS-3), access agreements (PS-6), boundary protection (SC-7), transmission confidentiality and integrity (SC-8), cryptographic protection (SC-13), covert channel analysis (SC-31), system monitoring (SI-4), and protection from information leakage due to electromagnetic emanation (PE-19). Regarding the latter, the TEMPEST is a valuable specification regarding equipment shielding against non-intentional leakage of radio or electric signals, sounds, and vibrations.

Table 7. NIST CSF core functions.

Core Function	Description
Identify	Help determine the current cybersecurity risk to the organization.
Protect	Use safeguards to prevent or reduce cybersecurity risk.
Detect	Find and analyze possible cybersecurity attacks and compromises.
Respond	Take action regarding a detected cybersecurity incident.
Recover	Restore assets and operations that were impacted by a cybersecurity incident.

As these countermeasures are closely related to the attack vector, we considered the `breach_type` of the dataset for reviewing them.

However, it is essential to note that listing all security measures for mitigating data breach-related vulnerabilities is impractical due to the vast number of attack vectors.

Therefore, this section provides general good practices against common attack vectors, not an exhaustive list of available security controls.

5.2.1. CARD

The Payment Card Industry Data Security Standard (PCI-DSS) is a critical debit card- and credit card-related data security standard. The PCI requires technical and operational controls to be put in place by any entity that stores, processes, or transmits credit card data.

The PCI-DSS delineates specific requirements for protecting payment card data. These requirements are detailed in Table 8, providing a comprehensive overview of the PCI-DSS standards and their associated requirements.

This is enforced by three ongoing steps: an assessment, identifying all locations of cardholder data, an inventory of assets, and analyzing them for vulnerabilities that could expose cardholder data. The following stage is to repair the vulnerabilities found, and, lastly, to report the assessment and remediation details and submit the resulting document to entities the company does business with.

Table 8. PCI-DSS requirements summarized.

Goals	Requirements
Build and maintain a secure network and systems.	Install and maintain a firewall.
	Do not use vendor-supplied defaults for system passwords and other security parameters.
Protect cardholder data.	Protect stored cardholder data.
	Encrypt transmission of cardholder data across open, public networks.
Maintain a vulnerability-management program.	Protect all systems against malware and regularly update antivirus software or programs.
	Develop and maintain secure systems and applications.
Implement strong access-control measures.	Restrict access to cardholder data by business need to know.
	Identify and authenticate access to system components.
	Restrict physical access to cardholder data.
Regularly monitor and test networks.	Track and monitor all access to network resources and cardholder data.
	Regularly test security systems and processes.
Maintain an information security policy.	Maintain a policy that addresses information security for all personnel.

Regarding encryption, PCI-DSS requires compliance with the Point-to-Point Encryption (P2PE) standard by using one of their listed validated solutions. Two relevant data breaches involving debit card and credit card information were the Home Depot and Target Corp. breaches, displayed in Figure 6, which represent the most voluminous breaches.

The most probable cause of the breach at Target was an infection by the Citadel malware [62]. This malware, which is based on its predecessor Zeus, executes a Man-in-the-Browser attack. Another malware used in the attack was BlackPOS, which aims at Points of Sale (POS) devices [63]. Section 5.2.2 provides more information on banking malware.

Consequently, approximately 40 million credit card and debit card records were leaked, including their encrypted PINs and other PII.

According to the studied dataset, Home Depot was also infected by BlackPOS [63], leaking 56 million payment records. The two companies were also PCI-DSS compliant at the time of the breach [64], although Table 8 shows some requirements that could have prevented a malware infection if successfully implemented, such as regularly updating antivirus software and maintaining secure systems and applications. Additionally, the Home Depot data breach could have been prevented using P2PE and network segregation. Hence, it is observed that PCI-DSS serves as a reliable foundation for credit card security, but for better security it should not be solely implemented.

One additional technology that may be used to improve transaction security is the Europay, MasterCard, and Visa (EMV) micro-processing chip, increasing complexity and costs for card counterfeiting, which is known as skimming.

Counterfeiting cases are steeply lowering in areas where EMV is implemented, but there has been a consequential rise in Card Not Present (CNP) crime [65]. A CNP crime is the unauthorized use of another individual's payment details for a transaction, mainly through online means. The payment information may be obtained after a data breach: for example, Bodker et al. [66] describe the script followed by criminals in a CNP crime, which allows a more reasoned consideration for mitigation strategies.

Furthermore, both Target and Home Depot breaches were initiated with a phishing attack [63], which reinforces the need for Security Education, Training, and Awareness (SETA), as discussed in Section 5.2.7.

5.2.2. HACK

As evidenced in the cases examined in this paper, such as eBay, Target, and Home Depot, phishing is a common threat vector used to initiate data breaches. A practical approach for reducing the success rate of these attacks is implementing a robust SETA program, which is discussed in more detail in Section 5.2.7. Such a program plays a fundamental role in enhancing employees' ability to recognize and thwart security threats, such as phishing attempts.

Naqvi et al. [67] reviewed the literature on phishing mitigation procedures through different vectors, such as e-mail and websites. Most proposed techniques rely on Machine Learning or training and awareness. Multifactor Authentication (MFA) may protect the user account even after successful phishing, as the user's identification and password obtained with the technique would not suffice for logging in, requiring an extra factor.

As related in Section 5.2.1, in Target and Home Depot breaches, after successfully phishing credentials, the attacker used banking malware to exfiltrate data. As the financial sector represents the most significant contribution portion in the dataset, we find it convenient to discuss this type of malware.

Black et al. [68] surveyed some of these malware (namely, Zeus V2, Citadel, Carberp, Vawtrak, Dridex, Dyre, and Rovnix), providing Indicators of Compromise (IoC) for identifying their infection and evaluating their similarities and differences.

However, it is relevant to note that certain malware strains are region-focused, such as Guildma, Grandoreiro, and Javali, which primarily targeted Brazilian entities [69]. A threat intelligence project may be needed to identify common malicious activities within the organization's operational domain.

Even after a successful malware infection, the data breach may be prevented if the company effectively applies other security measures, such as encryption and access control. This was not the case, for example, with LinkedIn, which, as discussed in Section 4.3, had millions of unsalted password hashes leaked, highlighting the importance of comprehensive security measures to protect sensitive data.

Several factors compounded LinkedIn's security vulnerabilities. Firstly, the company employed the SHA-1 hashing algorithm, which has been demonstrated to be vulnerable to various attacks. NIST SP 800-131A revision 2 has disallowed the use of SHA-1, permitting it for non-digital signature applications only. Currently, SHA-2 and SHA-3 are secure message-digesting algorithms.

Secondly, LinkedIn's security was compromised by the absence of a salt algorithm to enhance the security of hashed passwords. When the same password is processed using the same message-digesting algorithm, it consistently generates the same hash value, which increases predictability and susceptibility to brute-force attacks. Salting algorithms involve appending a unique string to the password before hashing it, significantly improving its security. Additionally, password leak bases should be continually monitored, in search of credentials in use at the organization.

Section 4.1 briefly introduced the Heartland Payment System breach, which relied on SQL injection. For application-level vulnerabilities, such as the one exploited at HPY, the Open Worldwide Application Security Project (OWASP) is a well-known reference. They regularly publish the Top 10 vulnerabilities in the application security scope, along with their mitigation strategies, such as input validation, Web Application Firewall (WAF), and software testing.

Although not in the studied dataset, attackers successfully intruded on cloud providers in the 2020 SolarWinds hack case, exposing and breaching their customers' data [70]. In this incident, adversaries inserted arbitrary code in the source code of a company product called Orion. Afterward, SolarWinds distributed the malicious code to its customers as part of the product, infecting over 18,000, including government entities and private companies [71]. This example reinforces the importance of Supply Chain Management (SCM) in cybersecurity.

In large and technologically complex companies, keeping the systems up to date may be challenging. As a consequence, attackers may exploit known vulnerabilities in the systems. Thus, it is fundamental to establish a patch-management program to timeously update and secure the organization's assets.

For zero-day attacks, which explore previously unknown vulnerabilities, a security patch has not yet been published by the product developer, and signature detection is ineffective. As an alternative, ML methods can detect such intrusions based on the perception of suspicious activities that differ from the expected baseline, which may enhance the detectability for novel attacks.

Data Loss Prevention (DLP) solutions also contribute to data security and avoidance of data breaches. Such technologies detect and deter unauthorized data transfers, including preventing PII data breaches. However, it is important to note that DLP is ineffective in detecting data exfiltration through steganographic techniques.

When there is a need to publish statistical metrics related to a dataset, but concerns about preserving the privacy of the individuals within the dataset are paramount, leveraging differential privacy can be a valuable and appropriate approach to addressing this challenge. Differential privacy, introduced by Dwork [72], provides a framework for releasing aggregate information about a dataset while adding noise or perturbation to the data so that individual records remain private and indistinguishable. This ensures that sensitive information is protected and that statistical knowledge can be derived without compromising the privacy of the data subjects, whilst maintaining the utility of the data [73].

Because of this balance between utility and privacy enhancement, differential privacy has applicability in several areas, and is used by the US Census and by big companies such as Google, Apple, and Microsoft [74].

5.2.3. INSD

An insider threat is anyone with authorized access to or knowledge of an organization's resources. The company trusts this person, who knows the company's fundamentals and has access to its assets. Because of that, a malicious insider can potentially cause great damage to the company imperceptibly. The average time taken for a company to detect an insider's malicious actions is 85 days [75].

An insider may be classified as unintentional or intentional. As unintentional insider threats are more suitable, within the scope of this work, to DISC, PHYS, PORT, and START, in this section, we discuss mainly intentional malicious insiders.

The main motivations for conducting an insider attack are financial benefits or espionage [76]. Insider-incident action is privilege abuse, while the actions are undertaken mainly via privilege abuse. Because of that, the least-privilege policy and Privileged Access Management (PAM) technology are helpful tools for preventing insider leakage.

The leading adopted technologies used for mitigating insider threats are Data Loss Prevention (DLP), Privileged Access Management, User and Entity Behavior Analytics

(UEBA), Security Information and Event Management (SIEM), Endpoint Detection and Response (EDR), and Insider Threat Management (ITM) [75].

Administrative security controls may also be implemented. A background check and an employee screening upon hiring may reveal a mischievous past history for the candidate, enabling the company to cancel the employment process. If a person passes this investigation, enforcing a Non-Disclosure Agreement (NDA) signing is an additional countermeasure, as it will legally constitute their liability.

After an employee is hired, other security measures should still be adopted. One is assessing the need to know for each employee, enforced through an access control mechanism. Granting more access to knowledge and data than the employee needs to perform their usual tasks exposes the information unnecessarily. A similar control is based on the principle of the least privilege, which grants a worker the minimum necessary privileges.

Separation of duties is another form of mitigating inside intentional threats, which divides critical tasks among several employees, as per their department in the organization, for example. A job rotation policy, although sometimes infeasible, may also help manifest fraud, sabotage, or espionage. Terminating the contract with the employer is another critical step in preventing data leakage, and the company must ensure that the user's accounts are disabled, preferably during the exit interview, in which any equipment belonging to the organization should be returned, and after which the ex-employee should be escorted out of the facility.

5.2.4. PHYS, PORT, STAT: Losses

Losing a device is greatly facilitated by its mobility, as an employee may take it anywhere and be robbed or mislay the portable equipment or document, possibly containing sensitive company data. In that regard, Bring Your Own Device (BYOD) has increased the potential for such occurrences. It refers to using employee-owned mobile devices to access business enterprise content or networks. Similar portability concepts are Choose Your Own Device (CYOD), Company-Owned and Personally Enabled (COPE), and Company-Owned Business Only (COBO), and they all raise security concerns.

Wani et al. [77] list some challenges these mobile devices bring to hospitals, which may also apply to companies in general. They categorize these challenges as related to technology, human factors, and policies. They also provide possible solutions to these challenges. Table 9 presents these challenges and solutions.

In addition to BYOD, teleworking and co-working spaces may pose a security threat to companies. These work models, which have emerged since the COVID-19 pandemic, also imply new security gaps similar to the BYOD-related ones.

Table 9. BYOD challenges and solutions [77].

	Technical	**People**	**Policy**
Challenges	Insecure device	Inappropriate behavior	Lack of policy
	Absence of locking	Lack of awareness	Compliance
	Insecure network	Poor user experience	Sanctions for breaches
	Suspicious app installed	Skills shortage	
Solutions	Mobile device management	Security culture	BYOD strategy and governance
	Containerization	Awareness and training	User agreement
	Identity and access management	Skills improvement	BYOD policy
	Endpoint security tools		
	Secure communication platforms		

Geofencing is another suitable security control in this scenario, which refers to triggered actions in response to a device leaving a pre-defined geolocation. Such actions could

be, for instance, disabling its network interface card or remotely wiping the device to prevent data leakage upon exiting an authorized area.

On this subject, Uz [78] evaluated the effectiveness of remotely wiping data, considering the deleted data may be forensically retrieved, as explained in Section 5.2.5.

Encryption is also recommended for data protection, and, for mobile devices, File-Based Encryption (FBE) is mandatory in Android since its 10th version [79]. For notebooks, BitLocker and VeraCrypt are some available options.

One appropriate access control method for BYOD is Attribute-Based Access Control (ABAC). As per this paradigm, the company can deny and concede access to an identity based on attributes of the request, such as location, hour of the day, and object being accessed.

5.2.5. PHYS, PORT, STAT: Disposals

When disposing of sensitive data, one must be aware of the possibility of an adversary searching the dustbin, which is known as dumpster diving, a social engineering attack.

With this method, the attacker may access any object the company discards, such as equipment and documents. There may be sensitive data among this disposed-of material, such as employees' noted passwords or customers' data. In that case, the malicious actor will have more information to conduct the attack.

It is noteworthy that since the California v. Greenwood case in 1988, the legality of the warrantless search-and-seizure of garbage left in public areas has been established [80]. Because of that, for one more layer of security against data breaches, companies should keep their waste bins locked in private areas.

As an additional countermeasure to this approach, a company should, at the end of the data life cycle, carry out an adequate disposal of information. To accomplish that, Data Classification and Asset Disposal Policies should be implemented and publicized to raise employees' awareness. To aid suitable editing of these and other policies, several esteemed security organizations provide policy templates, such as the SANS Institute (sans.org/information-security-policy/ accessed on 7 April 2024) and CIS (cisecurity.org/ accessed on 7 April 2024).

Before disposal, the media must be sanitized—that is, have its data rendered inaccessible for a given level of the attacker effort, depending on the classification of the data. Proper media-sanitization techniques are presented by NIST SP 800-88 [81] for different media types.For paper documents, for example, the standard states that they must be shredded in pieces small enough that there is reasonable assurance that the data cannot be reconstructed in proportion to the data confidentiality. To further hinder a malicious reconstruction, sensitive documents may be mixed with public paper in the shredder input. Regarding the size of the shredded pieces for each classification level, the German standard DIN 66399 [82] provides some valuable guidelines, some of which are summarized in Table 10.

Table 10. German DIN 66399 paper shredding sizes according to the data sensitivity.

Classification Level	Maximum Piece Width/Area
P-1 (least sensitive)	2000 mm^2 (particle area) or 12 mm (strip width)
P-2	800 mm^2 (particle area) or 6 mm (strip width)
P-3	320 mm^2 (particle area) or 2 mm (strip width)
P-4	160 mm^2 (cross-cut particle area)
P-5	30 mm^2 (cross-cut particle area)
P-6	10 mm^2 (cross-cut particle area)
P-7 (most sensitive)	5 mm^2 (cross-cut particle area)

Similar disposal approaches should be deployed to digital devices, such as Hard Drives (HD), Solid-State Drives (SSD), flash drives, and CDs/DVDs. Despite physical destruction and shredding still being possible for these types of media, and it indeed being

recommended for more sensitive cases, the nature of these devices allows for other erasure mechanisms, especially for the least sensitive data.

It is known that simply deleting files via the operating system is not an effective way to purge data, as data carving techniques can retrieve said files [83]. Other techniques, such as zero filling, in which all data are overwritten with zeroes, are effective against commonly available data retrieval mechanisms, according to NIST SP 800-88. Additional filling rounds may be performed to increase security.

Specifically for magnetic HDs, the degaussing technique may be used. It consists of applying a magnetic field to the hard drive, which changes the magnetic patterns on the device, consequently destroying the data.

The degaussing approach will not be practical for SSDs, which are not magnetic. For this type of media, a secure way of dealing with data remanence is crypto-shredding, also named crypto-erasure. In this procedure, the data stored in the device are encrypted with a secure algorithm, and then the decryption key is discarded, rendering the data unrecoverable.

In addition to disposals, these sanitization techniques should also be applied when donating or selling the devices if the sensitivity of the data allows the transfer of the property of the media.

5.2.6. PHYS, PORT, STAT: Thefts and Inappropriate Accesses

This section primarily discusses physical security aspects that may be implemented at a company facility to prevent a data breach. We understand that the mitigation approaches related to thefts of the company assets in possession of an employee outside of the company's premises are embraced in Section 5.2.4.

The physical-security design in a company starts in the architectural-arrangement stage of the facility construction. Crime Prevention Through Environmental Design (CPTED) strategies may be employed during this phase. Through this approach, criminals are deterred and more easily detected by the physical layout of the space.

The main CPTED principles are natural surveillance, access control, territorial reinforcement, and maintenance [84]. As an example, it is stated that fences should be at least 3 ft (about 1 m) high to deter casual trespassers and at least 8 ft (approximately 2.5 m) high to deter purposeful infiltrators [85].

Other physical controls should be implemented to prevent incidents, especially in more sensitive areas, such as data centers. Examples include the use of a keypad (preventive), guards (deterrent), security cameras (detective), and alarms (corrective).

Nonetheless, all these security measures will be rendered useless if the human factor is successfully explored. An adversary may, for example, covertly sneak through a door opened by authorized personnel, a practice known as tailgating, or they may convince someone to let them enter, for example, by saying that they forgot their badge and are in a hurry. The latter is a social engineering tactic known as piggybacking. An effective countermeasure to these intrusions is the use of a mantrap.

However, intruders do not always perpetrate physical incidents. Authorized guests, for example, may perform unauthorized activities, and, in that case, additional physical countermeasures must be put in place.

One possible gap is the direct observation of devices' screens and keyboards. In such cases, an adversary may obtain sensitive data such as passwords through shoulder surfing. To hinder this activity, it may be necessary to relocate the devices.

Another security measure regarding the employee's workstation is the implementation of a Clean-Desk Policy, which enforces that all desks within the company must be clear of objects and documents. After successfully implementing this policy, an intruder cannot steal a sensitive document from a worker's desk.

NIST SP 800-12 [86] Chapter 15 reviews other physical security practices. Physical safety, which aims to protect people's physical integrity, life, and health, is another relevant

topic in this discussion. However, as these incidents do not usually result in data breaches, which are the focus of this work, we do not include them in the discussion.

5.2.7. DISC

Unintended disclosure may be classified as the result of an unintentional insider threat, either due to negligence or recklessness.

An effective Security Education, Training, and Awareness program may be capable of reducing the incidence of these cases and promoting compliance in an organization.

Education aims to equip IT personnel with security skills through methods like cyberattack simulations, targeting a high level of expertise. Training focuses on enhancing security knowledge among all employees through classes, for example. Awareness efforts aim to capture the attention of all employees regarding security concerns through mediums like banners, addressing a basic level of security understanding [87].

It may be observed that security awareness programs are helpful in mitigating risks associated with the general utilization of technological resources by the general user, such as credentials compromise and social engineering attacks, like phishing. Conversely, security training and education focus on the prevention of cyber incidents rooted in technical vulnerabilities, such as weaknesses originating from misconfiguration, and they should be directed to IT personnel.

Alyami et al. [88] assessed the critical factors for deploying a successful SETA program, based on a survey with 65 respondents. They produced a ranked list of essential factors of success. Gamification is also seen as a reasonable way of enhancing engagement in the program.

According to PCR (Table 2), the DISC type categorization includes publicly posted information sent to the wrong party. In addition to SETA, a two-person control may also reduce the likelihood of these disclosures. With this approach, two people must authorize an action before its execution.

5.3. Containment, Recovery, and Response

The affected company must study a response strategy after an attacker and a data breach have circumvented the security controls. From the technical point of view, the company must quickly contain the data leakage to minimize the potential damage and then identify and eradicate the components of the incident. NIST SP 800-61 [89] provides a more in-depth guide for computer security incident handling.

This NIST publication divides the incident response process into five steps: Preparation, which occurs before an incident and corresponds to preventive security measures; Detection and Analysis, in which the attack vector and TTPs are identified; Containment, Eradication, and Recovery, which comprises an initial restriction of the malicious activity and a subsequent cleanse of malicious artifacts (though keeping them for forensic analysis); followed by the Restoration of the systems' operation. Finally, post-incident activities include discussing and documenting the incident to understand it better and prevent similar future intrusions.

For a better comprehension of the causes of the incident and of eventual system modifications made by the intruder, forensic tools and techniques may be helpful. NIST SP 800-86 [90] provides guidelines for integrating forensic techniques into incident response, including data collection, examination, and reporting from different sources, such as files, operating systems, networks, and applications. When performing digital forensics, it is important to maintain a chain of custody and preserve the integrity of the evidence, not removing it.

Specifically, in the data leakage domain, ref. [91] proposed a data breach response methodology based on ISO 27035 [92] and NIST 800-61 [89]. Their study emphasized the importance of automating this process, especially due to the short time required to notify a breach for legislative compliance.

Hillmann et al. [93] conducted 12 interviews with customers regarding their expectations regarding a data breach response. They concluded that expectations vary according to several factors, such as breach severity, data leakage type, and company sector. Hence, a company must adapt its response strategy to the specific scenario to maximize the chance of meeting its customers' expectations.

As a more general guide, the Federal Trade Commission (FTC) presents recommendations for an adequate data breach response (www.ftc.gov/business-guidance/resources/data-breach-response-guide-business accessed on 7 April 2024), such as assembling an incident response team, fixing vulnerabilities, and removing improperly posted information on the web.

For notification, especially for complying with the legislation mentioned in Section 5.1, the FTC mentions the importance of notifying law enforcement and affected businesses and individuals, specifying what happened, what information was stolen, how the attackers used the information, what remediation measures were taken, and how customers may contact the organization regarding the breach. They also provide a model letter for data breach notification.

6. Conclusions and Future Work

Data breaches are a growing threat to organizations of all sizes and industries. This in-depth analysis of data breaches in publicly listed companies in the United States improves our comprehension of the evolving landscape of cybersecurity threats within the corporate sector. This research contains several notable findings, which provide valuable knowledge of the patterns and implications of data breaches in these organizations.

Our study discovered that the financial sector has emerged as the primary target for malicious actors. This emphasizes the critical need for deploying strong cybersecurity measures within the industry, as it continues to be a prime focus for cyberattacks. Understanding the vulnerabilities that make the financial sector susceptible to breaches is paramount for securing the sensitive data handled within these organizations. Thus, we have also shown the most common causes for breaches in this economic sector.

Not only in the financial sector, we identified that incidents related to portable devices (PORT) and to malicious outsiders (HACK) were the most prevalent types of breaches. This highlights the importance of organizations taking proactive countermeasures to protect their data and mitigate the risk of such incidents.

Several frameworks, standards, and laws have been discussed to achieve this protection and some vulnerability mitigation controls. Adopting these practices is fundamental for preventing PII exposure. By addressing the vulnerabilities and threats revealed in this research, organizations can better protect their sensitive data and minimize the potential financial and reputational damages associated with data breaches. However, they are not exhaustive nor definite, and companies should adopt an approach of continuous due diligence and due care.

By discussing compliance, prevention, impacts, and response to data breaches, this paper enhances the understanding of data-exposure patterns and advances the discussion on applicable strategies for reducing the probability of occurrence and consequent costs. Addressing cyber incidents from the combined perspective of regulatory aspects, implementing security controls, and response planning is fundamental for appropriately mitigating the related risks.

We propose, as future work, a similar approach to studying data breaches, adopting a global data breaches dataset, along with a discussion of data protection laws at the federal level for different countries and their compliance aspects. Furthermore, expanding the research scope to encompass publicly and privately held companies may provide a more comprehensive view of the data breach scenario.

Author Contributions: A.L.M.S. and R.d.O.A. suggested the methodology and validated the experiments; G.A.P.R. and G.F.V. performed the formal analysis and provided the data visualization; G.D.A.N. supervised the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The dataset utilized in this study is openly available through the corresponding publication [6].

Acknowledgments: The authors would like to express their thanks for the technical and computational support provided by the LATITUDE Laboratory at the University of Brasilia, as well as the TED 01/2019 from the Office of the Solicitor General of the Union (Grant AGU 697.935/2019), the TED 01/2021 from the National Secretariat for Social Assistance—SNAS/DGSUAS/CGRS for the SISTER City Project—Intelligent Secure and Real-Time Effective Systems for Smart Cities (Grant 625/2022), the "System for Control and Unification of Projects for the Government of the Federal District—Sispro-DF" Project (Grant 497/2023), the General Attorney's Office for the National Treasure (Grant PGFN 23106.148934/2019-67), and FAP/DF. The authors would also like to thank the support of the Professional Post-Graduate Program in Electrical Engineering (PPEE).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following aviations are used in this manuscript:

BYOD	Bring Your Own Device
CARD	Debit and credit card data breach
CNP	Card Not Present
CCPA	California Consumer Privacy Act
CPA	Colorado Privacy Act
CPRA	California Privacy Rights Act
CPTED	Crime Prevention Through Environmental Design
CSF	Cyber Security Framework
CTDPA	Connecticut Data Privacy Act
CIS	Center for Internet Security
DISC	Data breach caused by unintentional disclosure
DLP	Data Loss Prevention
DPDPA	Delaware Personal Data Privacy Act
EMV	Europay, MasterCard, and Visa
FDBR	Florida Digital Bill of Rights
FTC	Federal Trade Commission
GLBA	Gramm–Leach–Bliley Act
HACK	Data breach caused by hacking activity
HD	Hard Drive
HIPAA	Health Insurance Portability and Accountability Act
ICDPA	Iowa Consumer Data Protection Act
Indiana CDPA	Indiana Consumer Data Protection Act
INSD	Data breach caused by malicious insider
ML	Machine Learning
MTCDPA	Montana Consumer Data Privacy Act
NYSE	New York Stock Exchange
OCPA	Oregon Consumer Privacy Act
P2PE	Point-to-Point Encryption
PCI-DSS	Payment Card Industry Data Security Standard
PHYS	Data breach involving paper documents
PII	Personal Identifiable Information
PORT	Data breach involving portable devices
PRC	Privacy Rights Clearinghouse
SETA	Security Education Training and Awareness
SOX	Sarbanes–Oxley Act

SSD	Solid-State Drive
STAT	Data breach involving stationary devices
TDPSA	Texas Data Privacy and Security Act
TIPA	Tennessee Information Protection Act
TTP	Tactics, Techniques, and Procedures
UCPA	Utah Consumer Privacy Act
UNKN	Data breach caused by unknown vector
VCDPA	Virginia Consumer Data Protection Act

References

1. Tripathi, M.; Mukhopadhyay, A. Financial loss due to a data privacy breach: An empirical analysis. *J. Organ. Comput. Electron. Commer.* **2020**, *30*, 381–400. [CrossRef]
2. Petkauskas, V. Mother of All Breaches Reveals 26 Billion Records. 2024. Available online: http://cybernews.com/security/billions-passwords-credentials-leaked-mother-of-all-breaches (accessed on 7 May 2024).
3. Wang, P.; D'Cruze, H.; Wood, D. Economic costs and impacts of business data breaches. *Issues Inf. Syst.* **2019**, *20*, 162–171.
4. Yimam, D.; Fernandez, E.B. A survey of compliance issues in cloud computing. *J. Internet Serv. Appl.* **2016**, *7*, 5. [CrossRef]
5. Khan, F.S.; Kim, J.H.; Moore, R.L.; Mathiassen, L. Data breach risks and resolutions: A literature synthesis. In Proceedings of the 25th Americas Conference on Information Systems, Cancún, Mexico, 15–17 August 2019; pp. 1–10.
6. Rosati, P.; Lynn, T. A dataset for accounting, finance and economics research on US data breaches. *Data Brief* **2021**, *35*, 106924. [CrossRef] [PubMed]
7. Layton, R.; Watters, P.A. A methodology for estimating the tangible cost of data breaches. *J. Inf. Secur. Appl.* **2014**, *19*, 321–330. [CrossRef]
8. Sood, G.; Cor, K. Pwned: The risk of exposure from data breaches. In Proceedings of the10th ACM Conference on Web Science, Boston, MA, USA, 30 June–3 July 2019; pp. 289–292.
9. Hammouchi, H.; Cherqi, O.; Mezzour, G.; Ghogho, M.; El Koutbi, M. Digging deeper into data breaches: An exploratory data analysis of hacking breaches over time. *Procedia Comput. Sci.* **2019**, *151*, 1004–1009. [CrossRef]
10. Raghupathi, W.; Raghupathi, V.; Saharia, A. Analyzing Health Data Breaches: A Visual Analytics Approach. *AppliedMath* **2023**, *3*, 175–199. [CrossRef]
11. Lee, I. Analysis of insider threats in the healthcare industry: A text mining approach. *Information* **2022**, *13*, 404. [CrossRef]
12. Churi, P.; Pawar, A.; Moreno-Guerrero, A.J. A comprehensive survey on data utility and privacy: Taking Indian healthcare system as a potential case study. *Inventions* **2021**, *6*, 45. [CrossRef]
13. Gong, M.; Wang, S.; Wang, L.; Liu, C.; Wang, J.; Guo, Q.; Zheng, H.; Xie, K.; Wang, C.; Hui, Z.; et al. Evaluation of privacy risks of Patients' data in China: Case study. *JMIR Med. Inform.* **2020**, *8*, e13046. [CrossRef]
14. Abouelmehdi, K.; Beni-Hessane, A.; Khaloufi, H. Big healthcare data: Preserving security and privacy. *J. Big Data* **2018**, *5*, 1–18. [CrossRef]
15. Djebbar, F.; Nordström, K. A Comparative Analysis of Industrial Cybersecurity Standards. *IEEE Access* **2023**, *11*, 85315–85332. [CrossRef]
16. Algarni, A.M.; Thayananthan, V.; Malaiya, Y.K. Quantitative assessment of cybersecurity risks for mitigating data breaches in business systems. *Appl. Sci.* **2021**, *11*, 3678. [CrossRef]
17. Sun, M.; Lu, Y. A Generalized Linear Mixed Model for Data Breaches and Its Application in Cyber Insurance. *Risks* **2022**, *10*, 224. [CrossRef]
18. Barati, M.; Yankson, B. Predicting the occurrence of a data breach. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100128. [CrossRef]
19. Masuch, K.; Greve, M.; Trang, S.; Kolbe, L.M. Apologize or justify? Examining the impact of data breach response actions on stock value of affected companies? *Comput. Secur.* **2022**, *112*, 102502. [CrossRef]
20. Tweneboah-Koduah, S.; Atsu, F.; Prasad, R. Reaction of stock volatility to data breach: An event study. *J. Cyber Secur. Mobil.* **2020**, *9*, 355–384. [CrossRef]
21. Piccotti, L.R.; Wang, H. Informed trading in the options market surrounding data breaches. *Glob. Financ. J.* **2023**, *56*, 100774. [CrossRef]
22. Chen, J.; Henry, E.; Jiang, X. Is cybersecurity risk factor disclosure informative? Evidence from disclosures following a data breach. *J. Bus. Ethics* **2023**, *187*, 199–224. [CrossRef]
23. Lulandala, E.E. Facebook data breach: A systematic review of its consequences on consumers' behaviour towards advertising. In *Strategic System Assurance and Business Analytics*; Springer: Singapore, 2020; pp. 45–68.
24. Molitor, D.; Raghupathi, W.; Saharia, A.; Raghupathi, V. Exploring Key Issues in Cybersecurity Data Breaches: Analyzing Data Breach Litigation with ML-Based Text Analytics. *Information* **2023**, *14*, 600. [CrossRef]
25. Schlackl, F.; Link, N.; Hoehle, H. Antecedents and consequences of data breaches: A systematic review. *Inf. Manag.* **2022**, *59*, 103638. [CrossRef]
26. Patterson, C.M.; Nurse, J.R.; Franqueira, V.N. Learning from cyber security incidents: A systematic review and future research agenda. *Comput. Secur.* **2023**, *132*, 103309. [CrossRef]

27. Khan, F.; Kim, J.H.; Mathiassen, L.; Moore, R. Data breach management: An integrated risk model. *Inf. Manag.* **2021**, *58*, 103392. [CrossRef]
28. Ullah, F.; Edwards, M.; Ramdhany, R.; Chitchyan, R.; Babar, M.A.; Rashid, A. Data exfiltration: A review of external attack vectors and countermeasures. *J. Netw. Comput. Appl.* **2018**, *101*, 18–54. [CrossRef]
29. Aslam, M.; Khan Abbasi, M.A.; Khalid, T.; Shan, R.U.; Ullah, S.; Ahmad, T.; Saeed, S.; Alabbad, D.A.; Ahmad, R. Getting Smarter about Smart Cities: Improving Data Security and Privacy through Compliance. *Sensors* **2022**, *22*, 9338. [CrossRef] [PubMed]
30. Neto, N.N.; Madnick, S.; Paula, A.M.G.D.; Borges, N.M. Developing a global data breach database and the challenges encountered. *J. Data Inf. Qual.* **2021**, *13*, 1–33. [CrossRef]
31. Ronquillo, J.G.; Erik Winterholler, J.; Cwikla, K.; Szymanski, R.; Levy, C. Health IT, hacking, and cybersecurity: National trends in data breaches of protected health information. *JAMIA Open* **2018**, *1*, 15–19. [CrossRef] [PubMed]
32. Park, S. Why information security law has been ineffective in addressing security vulnerabilities: Evidence from California data breach notifications and relevant court and government records. *Int. Rev. Law Econ.* **2019**, *58*, 132–145. [CrossRef]
33. Tsen, E.; Ko, R.; Slapnicar, S. *Dataset of Data Breaches and Ransomware Attacks over 15 Years from 2004*; The University of Queensland: Brisbane, Australia, 2020.
34. Biddle, N.; Edwards, B.; Gray, M.; McEachern, S. ANU Poll 2018: Data Governance. *ADA Dataverse* **2020**. [CrossRef]
35. Ikegami, K.; Kikuchi, H. Modeling the risk of data breach incidents at the firm level. In *Innovative Mobile and Internet Services in Ubiquitous Computing: Proceedings of the 14th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing (IMIS-2020)*; Springer: Cham, Switzerland, 2021; pp. 135–148.
36. Malliouris, D.D. Finance & Cyber Security: Uncovering Underlying and Consequential Costs of Security Breaches and Investments. Ph.D Thesis, University of Oxford, Oxford, UK, 2021.
37. Cornejo, G.A. Human Errors in Data Breaches: An Exploratory Configurational Analysis. Ph.D Thesis, Nova Southeastern University, Fort Lauderdale, FL, USA, 2021.
38. Obaydin, I.; Xu, L.; Zurbruegg, R. The Unintended Cost of Data Breach Notification Laws: Evidence from Managerial Bad News Hoarding. SSRN 3926962. 2021. Available online: https://www.researchgate.net/publication/354769133_The_Unintended_Cost_of_Data_Breach_Notification_Laws_Evidence_from_Managerial_Bad_News_Hoarding (accessed on 4 April 2024).
39. Huh, J.Y. We Care About Your Privacy (When It Matters): How Firms Strategically Respond to Data Breach Incidents. Ph.D Thesis, Duke University, Durham, NC, USA, 2022.
40. Carfora, M.F.; Orlando, A. Some Remarks on Malicious and Negligent Data Breach Distribution Estimates. *Computation* **2022**, *10*, 208. [CrossRef]
41. Benzell, S.; Hersh, J.S.; Van Alstyne, M.W.; Lagarda, G. How APIs Create Growth by Inverting the Firm. SSRN 3432591. 2022. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3432591 (accessed on 4 April 2024).
42. Zadeh, A.; Lavine, B.; Zolbanin, H.M.; Hopkins, D. A cybersecurity risk quantification and classification framework for informed risk mitigation decisions. *Decis. Anal. J.* **2023**, *9*, 100328. [CrossRef]
43. Niyonzigira, F. Exploring Nonprofit Organizations' Successful Compliance Strategies Against Cyber Threats: A Qualitative Study Inquiry. Ph.D Thesis, Capella University, Minneapolis, MN, USA, 2023.
44. Mulla, S.M.; Ghorpade, V.R. Evolution of Predictive Methodologies to Obstruct Ever-Growing Data Breaches. In Proceedings of the 10th International Conference on "Computing for Sustainable Global Development", New Delhi, India, 15–17 March 2023; pp. 1698–1703.
45. Rodrigues, G.A.P.; Serrano, A.L.M.; de Oliveira Albuquerque, R.; Saiki, G.M.; Ribeiro, S.S.; Orozco, A.L.S.; Villalba, L.J.G. Mapping of data breaches in companies listed on the NYSE and NASDAQ: Insights and Implications. *Results Eng.* **2024**, *21*, 101893. [CrossRef]
46. Kouadio, K.L.; Liu, J.; Liu, R.; Wang, Y.; Liu, W. K-Means Featurizer: A booster for intricate datasets. *Earth Sci. Inform.* **2024**, *17*, 1203–1228. [CrossRef]
47. Hamza, F. Not Just for Investment and Job Search: The Role of Earnings Announcement as a Driver of Cyber Risks. In *Information and Communication Technology in Technical and Vocational Education and Training for Sustainable and Equal Opportunity: Education, Sustainability and Women Empowerment*; Springer: Berlin/Heidelberg, Germany, 2024; pp. 143–154.
48. Stevens, G.M. Data Security Breach Notification Laws. 2012. Available online: https://journalistsresource.org/wp-content/uploads/2012/04/R42475.pdf (accessed on 4 April 2024).
49. Reidenbach, M.; Wang, P. Heartland payment systems: Cybersecurity impact on audits and financial statement contingencies. *Issues Account. Educ.* **2021**, *36*, 93–109. [CrossRef]
50. Klaus, T.; Elzweig, B. The impact of data breaches on corporations and the status of potential regulation and litigation. *Law Financ. Mark. Rev.* **2020**, *14*, 255–260. [CrossRef]
51. Cohen, D.T.; Hatchard, G.W.; Wilson, S.G. *Population Trends in Incorporated Places: 2000 to 2013*; US Department of Commerce, Economics and Statistics Administration, US Census Bureau: Suitland-Silver Hill, MD, USA, 2015.
52. Poornachandran, P.; Nithun, M.; Pal, S.; Ashok, A.; Ajayan, A. Password reuse behavior: How massive online data breaches impacts personal data in web. In *Innovations in Computer Science and Engineering: Proceedings of the Third ICICSE, 2015*; Springer: Singapore, 2016; pp. 199–210.
53. Roberts, S. Learning lessons from data breaches. *Netw. Secur.* **2018**, *2018*, 8–11. [CrossRef]

54. Minkus, T.; Ross, K.W. I know what you're buying: Privacy breaches on ebay. In Proceedings of the Privacy Enhancing Technologies: 14th International Symposium, PETS 2014, Amsterdam, The Netherlands, 16–18 July 2014; Proceedings 14, pp. 164–183.
55. Pimenta Rodrigues, G.A.; Marques Serrano, A.L.; Lopes Espiñeira Lemos, A.N.; Canedo, E.D.; Mendonça, F.L.L.d.; de Oliveira Albuquerque, R.; Sandoval Orozco, A.L.; García Villalba, L.J. Understanding Data Breach from a Global Perspective: Incident Visualization and Data Protection Law Review. *Data* **2024**, *9*, 27. [CrossRef]
56. Potter, A.; Campbell, K.; Baldin, A.; Chambers, H.; Toto, B.; Saturnino, F.; Prescott, V. *Comparing Comprehensive Us Privacy Laws: A Guide to Compliance*; Technical Report; OneTrust DataGuidance: Atlanta, Georgia, 2023.
57. Coie, P. Security Breach Notification Chart. 2014. Available online: https://www.perkinscoie.com/en/news-insights/security-breach-notification-chart.html (accessed on 4 April 2024).
58. IBM. *Cost of a Data Breach Report*; Technical Report; IBM Security: Cambridge, MA, USA, 2023.
59. *ISA 62443-3-3:2013*; Security for Industrial Automation and Control Systems. International Society of Automation: Pittsburgh, CA, USA, 2023.
60. *ISO/IEC 27001:2013*; Information Security, Cybersecurity and Privacy Protection. International Organization for Standardization: London, UK, 2013.
61. *NIST SP 800-53*; Security and Privacy Controls for Information Systems and Organizations. National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023.
62. Plachkinova, M.; Maurer, C. Security breach at target. *J. Inf. Syst. Educ.* **2018**, *29*, 11–20.
63. Shu, X.; Tian, K.; Ciambrone, A.; Yao, D. Breaking the target: An analysis of target data breach and lessons learned. *arXiv* **2017**, arXiv:1701.04940.
64. Rosenblum, P. Lessons from Home Depot: Expect Hackers to Crack More Retailers This Holiday Season. 2014. Available online: https://www.forbes.com/sites/paularosenblum/2014/11/06/lessons-from-home-depot-expect-hackers-to-crack-more-retailers-this-holiday-season/?sh=1f6436ea68bc (accessed on 22 October 2023).
65. Froud, D. The global implications of US EMV adoption. *Comput. Fraud Secur.* **2016**, *2016*, 5–7. [CrossRef]
66. Bodker, A.; Connolly, P.; Sing, O.; Hutchins, B.; Townsley, M.; Drew, J. Card-not-present fraud: Using crime scripts to inform crime prevention initiatives. *Secur. J.* **2022**, *36*, 693–711. [CrossRef]
67. Naqvi, B.; Perova, K.; Farooq, A.; Makhdoom, I.; Oyedeji, S.; Porras, J. Mitigation strategies against the phishing attacks: A systematic literature review. *Comput. Secur.* **2023**, *132*, 103387. [CrossRef]
68. Black, P.; Gondal, I.; Layton, R. A survey of similarities in banking malware behaviours. *Comput. Secur.* **2018**, *77*, 756–772. [CrossRef]
69. Bhardwaj, A.; Kaushik, K.; Maashi, M.S.; Aljebreen, M.; Bharany, S. Alternate Data Stream Attack Framework to Perform Stealth Attacks on Active Directory Hosts. *Sustainability* **2022**, *14*, 12288. [CrossRef]
70. Marelli, M. The SolarWinds hack: Lessons for international humanitarian organizations. *Int. Rev. Red Cross* **2022**, *104*, 1267–1284. [CrossRef]
71. Martínez, J.; Duría, J.M. Software supply chain attacks, a threat to global cybersecurity: SolarWinds' case study. *Int. J. Saf. Secur. Eng.* **2021**, *11*, 537–545. [CrossRef]
72. Dwork, C. Differential privacy. In Proceedings of the International Colloquium on Automata, Languages and Programming, Venice, Italy, 10–14 July 2006; pp. 1–12.
73. Seeman, J.; Susser, D. Between privacy and utility: On differential privacy in theory and practice. *ACM J. Responsible Comput.* **2024**, *1*, 1–18. [CrossRef]
74. Murakami, T.; Sei, Y. Automatic Tuning of Privacy Budgets in Input-Discriminative Local Differential Privacy. *IEEE Internet Things J.* **2023**, *10*, 15990–16005. [CrossRef]
75. Ponemon Institute. Cost of Insider Threats Global Report. 2020. Available online: https://www.exclusive-networks.com/ie/wp-content/uploads/sites/19/2020/12/UK-VR-Proofpoint-Report-2020-Cost-of-Insider-Threats.pdf (accessed on 4 April 2024).
76. ENISA. Threat Landscape Report 2016. 2016. Available online: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2016 (accessed on 4 April 2024).
77. Wani, T.A.; Mendoza, A.; Gray, K. Hospital bring-your-own-device security challenges and solutions: systematic review of gray literature. *JMIR MHealth UHealth* **2020**, *8*, e18175. [CrossRef]
78. Uz, A. The effectiveness of remote wipe as a valid defense for enterprises implementing a BYOD policy. Ph.D. Thesis, Université d'Ottawa/University of Ottawa, Ottawa, ON, Canada, 2014.
79. Groß, T.; Busch, M.; Müller, T. One key to rule them all: Recovering the master key from RAM to break Android's file-based encryption. *Forensic Sci. Int. Digit. Investig.* **2021**, *36*, 301113. [CrossRef]
80. Herdrich, M.A. California v. Greenwood: The trashing of privacy. *Am. UL Rev.* **1988**, *38*, 993.
81. *NIST SP 800-88*; Data Destruction. National Institute of Standards and Technology: Gaithersburg, MD, USA, 2014.
82. *DIN 66399*; Office Machines—Destruction of Data Carriers, Deutsches Institut für Normung e.V. DIN: Berlin, Germany, 2012.
83. Azeem, E.A. The Data Carving-The Art of Retrieving Deleted Data as Evidence. *Int. J. Electron. Crime Investig.* **2022**, *6*, 8. [CrossRef]
84. Tan, W.H.; Abas, H. Systematic Literature Review Crime Prevention through Environmental Design (CPTED) in Physical Security for IT Organization. *Open Int. J. Inform.* **2022**, *10*, 68–83.

85. Fennelly, L.J.; Perry, M.A. Encompassing effective CPTED solutions in 2020 and beyond: Concepts and strategies. In *Handbook of Loss Prevention and Crime Prevention*; Elsevier: Amsterdam, The Netherlands, 2020; pp. 45–77.
86. *NIST SP 800-12*; Guidelines for Managing the Security of 34 Mobile Devices in the Enterprise. National Institute of Standards and Technology: Gaithersburg, MD, USA, 2023.
87. Hu, S.; Hsu, C.; Zhou, Z. Security education, training, and awareness programs: Literature review. *J. Comput. Inf. Syst.* **2022**, *62*, 752–764. [CrossRef]
88. Alyami, A.; Sammon, D.; Neville, K.; Mahony, C. Critical success factors for Security Education, Training and Awareness (SETA) programme effectiveness: An empirical comparison of practitioner perspectives. *Inf. Comput. Secur.* **2023**, *32*, 53–73. [CrossRef]
89. *NIST SP 800-61*; Making Government Services Easier to Find. National Institute of Standards and Technology: Gaithersburg, MD, USA, 2012.
90. *NIST SP 800-86*; Guide to Integrating Forensic Techniques into Incident Response. National Institute of Standards and Technology: Gaithersburg, MD, USA, 2006.
91. Rabello, A.; Goulart, J.; Karam, M.; Pitanga, M.; Baldoino Filho, R.G.; Ricioni, R. Proposed Incident Response Methodology for Data Leakage. *ICSEA 2021* **2021**, 60.
92. *ISO 27035*; Information Security Incident Management—Training Courses. International Organization for Standardization: Geneva, Switzerland, 2023.
93. Hillmann, F.; Klauenberg, T.; Schroeder, L.; Diesterhöft, T.O. A User-centric View on Data Breach Response Expectations. *CIISR* **2023**, 19.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.