# A COMPREHENSIVE AND COMPARATIVE EXAMINATION OF HEALTHCARE DATA BREACHES: ASSESSING SECURITY, PRIVACY, AND PERFORMANCE

by

MOHAMMED AL KINOON
B.S. Eastern Florida State College, 2019
M.S. University of Central Florida, United States, 2022

A dissertation submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy
in the Department of Computer Science
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term
2024

Major Professor: David Mohaisen

# ABSTRACT

The healthcare sector is pivotal, offering life-saving services and enhancing well-being and community life quality, especially with the transition from paper-based to digital electronic health records (EHR). While improving efficiency and patient safety, this digital shift has also made healthcare a prime target for cybercriminals. The sector's sensitive data, including personal identification information, treatment records, and SSNs, are valuable for illegal financial gains. The resultant data breaches, increased by interconnected systems, cyber threats, and insider vulnerabilities, present ongoing and complex challenges. In this dissertation, we tackle a multi-faceted examination of these challenges. We conducted a detailed analysis of healthcare data breaches using the VERIS (Vocabulary for Event Recording and Incident Sharing) dataset. We delve into the trends of these breaches, investigate the attack vectors, and identify patterns to inform effective mitigation strategies. We conducted a spatiotemporal analysis of the VERIS and the Office of Civil Rights (OCR) datasets. We explored the geographical and temporal distribution of breaches and focused on the types of targeted assets to decipher the attackers' motives. Additionally, we conducted a detailed analysis of hospitals' online presence, focusing on their security and performance features. By comparing government, non-profit, and private hospitals in the U.S., we examined their security practices, content, and domain attributes to highlight the differences and similarities in the digital profiles of these hospital types. Furthermore, we expand our scope to include a comparative sector-based study investigating data breaches across various critical sectors. This broader view provides a contextual understanding of the healthcare sector's unique vulnerabilities compared to other sectors. Overall, this dissertation contributes fundamental insights into healthcare data breaches and hospitals' digital presence and underscores the urgent need for enhanced understanding and implementation of robust security measures in this vitally important sector, striving for a balance between technological advancement and data security.

To my family.

# ACKNOWLEDGMENTS

I want to extend my sincere gratitude to Allah Almighty for granting me the patience and strength necessary to overcome the challenges encountered during this journey. My Ph.D. journey would not have been possible without the dedicated support and valuable contributions of many, to whom I extend my genuine gratitude.

First and foremost, I am incredibly grateful to my supervisor, Prof. David Mohaisen, for trusting me and giving me a chance to pursue my Ph.D. at the University of Central Florida. I am forever thankful for his invaluable guidance, continuous support, and encouragement throughout my Ph.D. studies. I also want to thank him for his patience and availability outside of the office for the walks and coffee breaks that made my experience pursuing the Ph.D. at UCF wholesome.

I would like to extend my sincere thanks to my doctoral dissertation committee members, Prof. Cliff Zou, Prof. Sung Choi Yoo, and Prof. Xueqiang Wang, for their valuable feedback and suggestions for improvement at every step of my doctoral milestones.

I am also grateful to my friends and collaborators in the Security and Analytics Lab (SEAL): Abdulrahman, Ahmed, Sultan, Soohyeon, Saad, Ulku, Afsah, Hattan, Mohammad, Ahod, Jie, Ali, Aziz, and Ran. I would also like to thank my collaborators outside SEAL: Prof. DaeHun Nyang (Ewha Womans University), Prof. Songqing Chen (George Mason University), Prof. Rhongho Jang (Wayne State University), and Prof. Manar Mohaisen (Northeastern Illinois University).

Last but not least, I want to mention my parents, whose constant motivation and guidance were essential throughout my Ph.D. journey. Their support was crucial at every step, helping me overcome my challenges. Additionally, I'm deeply thankful to my siblings. Their support and encouragement were key to my strength and success throughout this journey.

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# CHAPTER 1: INTRODUCTION

HPAA data breach can be defined as "any security incident in which unauthorized parties gain access to sensitive data or confidential information, including personal data (Social Security numbers, bank account numbers, healthcare data) or corporate data (customer data records, intellectual property, financial information)" [46]. In healthcare, data breaches cause devastating damage to both healthcare organizations and patients due to the sensitive and personal information involved in this sector. Data breaches expose highly personal and confidential information such as medical records, diagnoses, treatment plans, insurance, and financial information. Cybercriminals can exploit the breached information maliciously, such as through identity theft and financial fraud, by using personal information such as Social Security Numbers (SSNs) and insurance details to perform fraudulent activity.

Healthcare data breaches cause severe financial consequences to health organizations and patients. Data breach victims of theft or fraudulent activities may face financial losses and difficulties resolving after the breach. Additionally, healthcare organizations may face legal consequences, financial penalties, loss of reputation, and potential litigation resulting from breaches. The healthcare industry continues to be a primary target for cybercriminals, as evidenced by recent trends and statistics. In 2023, the healthcare sector remains highly vulnerable to cyber threats. According to the Office for Civil Rights (OCR), the first three months of 2023 alone witnessed 145 data breaches [36]. To put this into perspective, in 2022, there were a staggering 707 reported incidents, resulting in the compromise of approximately 51.9 million records [65].

The Health Insurance Portability and Accountability Act (HIPAA) [68] establishes privacy and security standards to protect patient's health information, securing its handling and protection by healthcare providers, health plans, and healthcare clearinghouses. HIPAA categorizes data

breaches and other compliance failures into four tiers, which determine the severity of the incident and the corresponding penalties. Unlike some other data protection laws, HIPAA fines are assessed per violation, which can pertain to specific areas of non-compliance or individual compromised records. HIPAA classifies data breaches and compliance failures into four tiers, each associated with its own range of penalties. The penalties for Tier 1 breaches typically range from $100 to $50,000 per violation. Tier 2 penalties can be anywhere from $1,000 to $50,000 per violation. Tier 3 breaches, resulting from intentional disregard but corrected within a specified timeframe, may incur penalties between $10,000 and $50,000 per violation. Finally, the most severe Tier 4 breaches, characterized by sustained noncompliance, can lead to penalties starting at $50,000 or more per violation [42]. The specific penalty amount for HIPPA breaches varies based on the circumstances and the organization's response. Given the continuous and rapid increase in the number of data breaches in the healthcare sector, we find it crucial to emphasize the importance of conducting thorough investigations and analyses of these breaches.

## Statement of Research

In this dissertation, we propose three comprehensive studies that delve into the analysis of data breaches in the healthcare industry. Each study focuses on specific aspects providing a deep understanding of the critical risk of data breaches. We further elaborate on each study in the following.

***Measuring Healthcare Data Breaches*** **(§ 3).** Recently, healthcare data breaches have grown rapidly. Moreover, throughout the COVID-19 pandemic, the level of exposure to security threats increased as the frequency of patient visits to hospitals has also increased. Additionally, circumstances and constraints such as the curfew imposed on the public have resulted in a noticeable increase in Internet usage for healthcare services, employing intelligent devices such as smartphones. The Healthcare sector is being targeted by criminals internally and externally; health-

care data breaches impact hospitals and patients alike. To examine issues and discover insights, a comprehensive study of health data breaches is necessary. This study investigates healthcare data breach incidents by conducting measurements and analysis recognizing different viewpoints, including temporal analysis, attack discovery, security attributes of the breached data, attack actors, and threat actions. Based on the analysis, we found the number of attacks is decreasing, although not precluding an increasing severity, the time of attack discovery is long across all targets, breached data does not employ basic security functions, threat actions are attributed to various vectors, e.g., malware, hacking, and misuse, and could be caused by internal actors. Our study provides a cautionary tale of medical security in light of confirmed incidents through measurements.

***Security Breaches in the Healthcare Domain*: A Spatiotemporal Analysis (§ 4).** Over the past several years, data breaches have grown and become more expensive in the healthcare sector. Healthcare organizations are the main target of cybercriminals due to sensitive and valuable data, such as patient demographics, SSNs, and personal treatment records. Data breaches are costly to breached organizations and affected individuals; hospitals can suffer substantial damage after the breach while losing customer trust. Attackers often use breached data maliciously, e.g., demanding ransom or selling patient information on the dark web. To this end, this study investigates data breaches incidents in the healthcare sector, including community, federal, and non-federal hospitals. Our analysis focuses on the reasoning and vulnerabilities that lead to data breaches, including the compromised information assets, geographical distribution of incidents, size of healthcare providers, the timeline discovery of incidents, and the discovery tools for external and internal incidents. We use correlation to examine the impact of several dimensions on data breaches. Among other interesting findings, our in-depth analysis and measurements revealed that the average number of data breaches in the United States is significantly higher than in the rest of the world, and the size of the health provider, accounting for factors such as the population and number of adults in a region, highly influences the level of exposure to data breaches in each state.

***Understanding the Security and Performance of the Web Presence of Hospitals*: A Measurement Study (§ 5).** The recent transformation of healthcare medical records from paper-based to digital and connected systems raises concerns regarding patients' security and online privacy. For instance, sensitive personal information, such as patients' names, addresses, and SSNs, may be targeted due to the lack of proper security and privacy mechanisms. Using a total of 4,774 hospitals categorized as government, non-profit, and proprietary hospitals, this study provides the first measurement-based analysis of hospitals' websites and connects the findings with data breaches through a correlation analysis. We study the security attributes of three categories, collectively and in contrast, against domain name-, content-, and SSL certificate-level features. We find that each type of hospital has a distinctive characteristic of its utilization of domain name registrars, top-level domain distribution, and domain creation distribution, as well as content type and HTTP request features. Security-wise, and consistent with the general population of websites, only 1% of government hospitals utilized DNSSEC, in contrast to 6% of the proprietary hospitals. Alarmingly, we found that 25% of the hospitals used plain HTTP, in contrast to 20% in the general web population. Additionally, 8%-84% of the hospitals, depending on their type, had some malicious content, which is mostly attributed to the lack of maintenance. We conclude with a correlation analysis against 414 confirmed and manually vetted hospitals' data breaches. Among other interesting findings, our study highlights that the security attributes highlighted in our analysis of hospital websites are forming a very strong indicator of their likelihood of being breached. Our analyses are the first step towards understanding patient online privacy, highlighting the lack of basic security in many hospitals' websites and opening various potential research directions.

***Intersecting the Data Breach Landscape: A Comparative Study Across Diverse Industries*: A Measurement Study (§ 6).** In an era where digital information is pivotal, data breaches pose a critical threat across various sectors. Using the Vocabulary for Event Recording and Incident Sharing (VERIS) dataset, this study employs mixed methods to uncover distinct breach patterns,

quantifying financial and reputational damages. Analyzing cyberattack incidents from 2003 to 2022 with the VERIS dataset, we uncover a surprising peak in 2013, challenging assumptions about the continuous rise in cyber threats. The U.S. leads in security breaches, influenced by reporting laws such as the HITECH Act. Sector-specific vulnerabilities reveal *Healthcare* as highly susceptible (81.27%). Notable challenges include data disposal issues in *Healthcare* and *Retail*, IT misconfigurations in *Information* and *Professional Services* , and misdelivery errors in *Administration*. These findings highlight diverse cybersecurity challenges, primarily driven by human error. Our research emphasizes the need for robust security protocols and tailored cybersecurity strategies. Our findings offer actionable insights for organizations, policymakers, addressing enduring consequences and paving the way for preventative strategies in the digital age.

**Organization.** This dissertation is organized as follows: Chapter 2 review existing literature in the field. Following this, Chapter 3 draws from the work published in [8], titled “Measuring Healthcare Breaches". Next, Chapter 4 is based on the research presented at the Computational Data and Social Networks - 10th International Conference, CSoNet 2021, cited as [9], and titled “Security Breaches in the Healthcare Domain: A Spatiotemporal Analysis". In Chapter 5, we delve into the work titled “Understanding the Security and Performance of the Web Presence of Hospitals: A Measurement Study", as presented at the 32nd International Conference on Computer Communications and Networks, ICCCN 2023, referenced in [7]. Chapter 6 provides a comparative analysis of data breaches in various industries, aiming to provide insights into trends and vulnerabilities, specifically in comparison to the healthcare sector, in the study titled “Intersecting the Data Breach Landscape: A Comparative Study Across Diverse Industries". Chapter 7 provides a discussion of the findings of this dissertation and suggests strategies to minimize the occurrence of breaches. The final chapter, Chapter 8, summarizes the main findings and conclusions of the dissertation.

# CHAPTER 2: RELATED WORK

Previous studies have extensively examined healthcare data breaches in recent years, providing insights into their frequency and underlying causes. However, there has been relatively little research focused on analyzing the security configurations of healthcare providers. To provide a comprehensive understanding of our work, we conducted a review of previous studies pertaining to website content, security analysis, and healthcare data breach analysis.

## Data Breaches Analysis

Several studies have recently investigated data breaches in the healthcare industry [8, 9, 80, 90]. For instance, Seh *et al.* [80] conducted a comprehensive analysis of HIPPA data breach reports. Their study highlights that hacking incidents, unauthorized access (internal), theft or loss, and improper disposal of unnecessary data are the main disclosure types of protected healthcare information. Moreover, the authors applied the Simple Moving Average (SMA) and Simple Exponential Smoothing (SES) time series methods on the data to determine the trend of healthcare data breaches and their cost to the healthcare industry. Choi *et al.* [23] estimated the link between data breaches and hospital advertising spending, studying the period of the two years following the breach and finding hospitals had much higher advertising expenditures. Siddartha *et al.* [81] found that the healthcare industry is being targeted for two main reasons: being a rich source of valuable data and its weak defenses. Siddartha and Ravikumar [81] suggested that the security techniques employed in the healthcare industry miss data analysis improvements, e.g., data format preservation, data size preservation, and other factors. Luis *et al.* [90] defined DNS queries and TLS/SSL connections to identify the dangers encountered inside a hospital environment without disrupting the functioning network using two years of collected data. Another line of work, the 2022 Data Breach Investiga-

tions Report (DBIR) [34], investigates healthcare breaches among other industries. Based on the report, healthcare suffered 849 incidents, with 571 confirmed data disclosure in 2022. The report summarized various findings and determined that external actors are behind 61% of data breaches while 39% of data breaches involved internal actors. Furthermore, according to the same report, financial gain is the highest motive for attackers at 95%, followed by espionage at 4%. Raghupathi *et al.* [75] conducted a recent study that investigated data breach occurrences in healthcare provider environments specifically related to patient data using the U.S. Department of Health and Human Services publicly available dataset. Their study found a correlation between the occurrence of data breaches, breach locations, breach types, and the presence of business associates. Moreover, their study identified Hacking as the most common type of data breach, and Network servers are the most popular location for information breaches. To enhance our understanding of security vulnerabilities and malware threats, several studies have made significant contributions. Mohaisen and Alrawi's work on automated classification of malware samples, including Zeus, laid the groundwork for advanced detection methodologies [63]. Mohaisen et al. [64] developed AMAL, a behavior-based malware analysis system, and introduced Chatter [64], a novel method for classifying malware families via system event ordering, showcasing advancements in detection techniques and signature identification. Alasmary *et al.* [5, 6] research on detecting IoT malware using a graph-based approach and ShellCore's automation for malicious IoT software detection via shell commands representation mark significant advancements in securing IoT devices.

### Hospitals Websites Analysis

Over the past few years, there has been a drastic increase in the development and utilization of online services and web applications. Parallel to this rise has been an increasing concern over the privacy and security of these online services and applications, e.g., different components can

be compromised, putting their users at risk. Chung *et al.* [24] offered the first in-depth analysis of incorrect certifications in the online Public Key Infrastructure (PKI), showing that most PKI certificates are invalid. The same study scrutinized the origin of the invalid SSL certificates and summarized that the preponderance of the invalid certificates was generated by end-user devices, with a periodical renewal of new self-signed certificates. SSL certificates have been investigated for website risk and vulnerability analysis [4,14,18,19,21,24,26,50–53,61,62,100]. For instance, Meyer *et al.* [60] examined the SSL certificates' content and information to distinguish between phishing and benign websites.

Alabduljabbar *et al.* [2,3,10–13] explored the SSL certificate-based structural differences between free and premium content websites and highlighted that 35.85% of the free websites' certificates have significant security issues, with 17% invalid, 7% expired, and 12% with mismatched domain names. Bach *et al.* [17] examined the content of hospital websites in three different countries and assessed them as information repositories or as interactive online communication means in three countries: (Bosnia-and-Herzegovina), recent (Croatia) and established EU member countries (Slovenia). In a recent study conducted by Yu *et al.* [99], a comprehensive analysis was performed on 19,483 hospital websites from 152 countries and provincial jurisdictions across Asia, Europe, North America, Latin America, Africa, and Oceania. The researchers utilized these crawled websites to investigate the presence of trackers, ultimately revealing some concerning findings. The findings showed that 53.5% of these websites used tracking scripts/cookies. Additionally, 33 websites were flagged as malicious, and 699 sites transmitted sensitive data to external servers through session replay services. This highlights concerns about privacy, security, and data protection on hospital websites.

# CHAPTER 3: MEASURING HEALTHCARE DATA BREACHES

The United States Department of Health and Human Services defines a data breach as an intentional or non-intentional use or disclosure of confidential health information. A data breach compromises privacy and security, resulting in a sufficient risk of reputation, financial, and other harm to the affected individuals [97]. Over the past few years, concerns related to healthcare data privacy have been mounting since healthcare information has become more digitized, distributed, and mobile [49]. Medical records have been transformed from paper-based into electronic health records (EHR) to facilitate various digital systems. Medical EHR can be described as "a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports" [59]. EHR enhances patient care by enhancing diagnostics and patient outcomes, improving patient participation, enhancing care coordination, practicing efficiencies, and cost savings [70]. Despite the numerous benefits of EHR, the transformation has inflated the security and privacy concerns regarding patients' information. The growing usage of the Internet of Things (IoT) and intelligent devices affects the methods of communication in hospitals and helps patients quickly access their medical treatment whenever necessary.

Nevertheless, the usage of such technologies is a fundamental factor that can cause security risks and lead to data breaches [83]. Broadly, healthcare data breaches are external and internal. External breaches are malicious, including at least one or more threat actions from cyber criminals, such as hacking, malware, and social attacks. On the other hand, internal data breaches typically occur

This work has been published at the 22nd World Conference on Information Security Applications (WISA 2021)

due to malfeasance by insiders, human errors, and negligence from employees. Data breaches have increased in the past decade. In comparison with other industries, healthcare is the worst affected [56]. Cybercriminals are targeting healthcare for two fundamental reasons: it is a rich source of valuable data, and its defenses are weak [32]. Medical records contain valuable information such as victims' home addresses and Social Security Numbers (SSNs). Adversaries utilize such information for malicious activities and identity theft or exchange those medical records for financial profit on the dark web.

**Contributions.** For a better understanding of the landscape of healthcare data breaches against various attributing characteristics, we provide a detailed study of the VERIS (Vocabulary for Event Recording and Incident Sharing) dataset. Among other characteristics, we temporally analyze data breaches and their growth over time. To understand the attacks' intent, we analyze the type of breaches over security attributes and characterize the threat actions, highlighting the attack vector employed for the breach. We hope that those characterizations will shed light on the trend and the attack vectors, thus providing directions for mitigating those breaches.

## Summary of Completed Work

In this work, we conducted an in-depth analysis of healthcare data breaches using an authentic dataset. We focused on analyzing the timeline of data breaches and their impact on health organizations. This work provides valuable insights into the nature of data breaches in the healthcare sector, highlighting the importance of addressing security vulnerabilities to protect patients, customers, and employees. The findings underscore the need for robust cybersecurity measures in health organizations to mitigate the risks posed by internal and external threats.

## Data Source and Temporal Analysis

The object of this paper is to conduct a measurement of healthcare data breaches to understand trends and motives. To accomplish that, we used trusted and reliable data called VERIS. In the past, there were numerous initiatives to accumulate and share security incidents. Nonetheless, commitment and participation have been minimal. Reasons behind that are many, including (i) the difficulty of categorization, and (ii) the uncertainty of what to measure [91]. To facilitate data collection and sharing, VERIS is established as a nonprofit community designed to accommodate a free source of a common language for describing security incidents in a structured and repeatable way [91]. Due to the prevailing lack of helpful information, the VERIS dataset is an effective solution to the most critical and persistent challenges in the security industry. VERIS tackles this problem by offering organizations the ability to collect relevant information and share them responsibly and anonymously.

**VERIS and Incident Attributes.** VERIS's primary purpose is to create an open-source database to design a foundation that constructively and cooperatively learns from their experience to ensure a more reliable measurement and management risk system. VERIS is a central hub whereby information and resources are shared to maximize the benefits of contributing organizations. During the incident collection process, the VERIS community focuses on successfully implementing an intersection, namely the 4A's, which indicate the following: who is behind the incidents (actors), the action used by the adversary (actions), devices affected (assets), and how are they effected (attributes). An example of the 4A's for an incident can be as follows: internal (actor), hacking (action), network (asset), and confidentiality (attribute). VERIS designers estimate the needed information to be collected about an incident based on the level of threat, asset, impact, and control. Understanding these risk aspects enables organizations to improve their management systems and make informed decisions. The power of VERIS is the collection of evidence during and after the

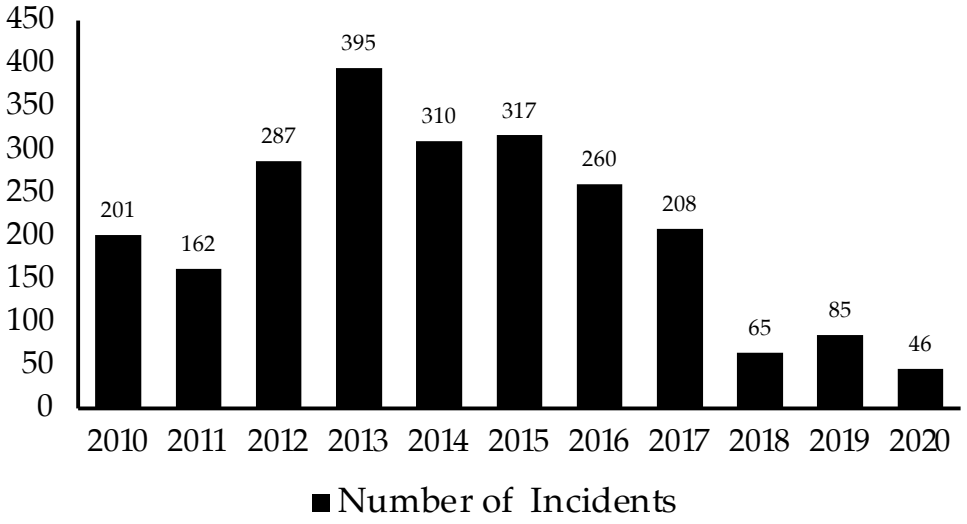



**Figure 3.1:** The yearly distribution of Data Breach Incidents.

incidents, besides providing helpful metrics to maximize risk management.

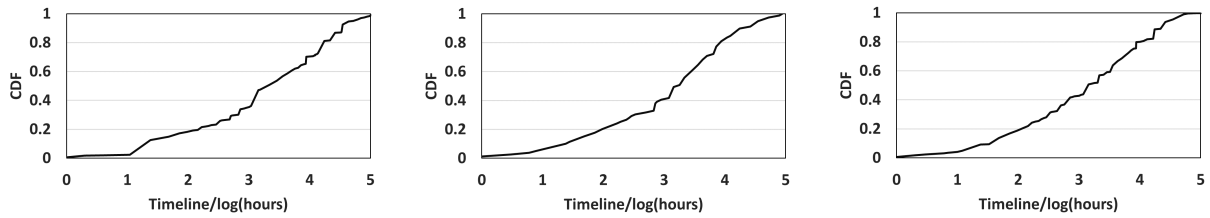
**Distribution of the Incident's Timeline.** We analyzed the timeline mapping of the incidents across the years. The VERIS dataset contains incidents that took place between the years 1971 and 2020. While a long period of time is considered, the time frame from 1971 until 2010 seemed to contain a low number of incidents, with only 272 (total) of them, per the VERIS dataset. Thus, to understand the actual trend in the active region, we limit ourselves to the years 2010 onward. This analysis is essential because it provides us with insights into the active period of breaches and attacks, and could hint at the underlying ecosystem. To this end, and upon this analysis, we found out the following (1) the per year number of incidents follows a normal distribution, with the peak in 2013. (2) 2013 was the highest year in the number of incidents, with 395 (16%), followed by 2015 with 317 (13%), and 2014 with 310 ($\approx$13%). (3) Contrary to the common belief that the number of attacks is increasing, we found that the number of breaches has been decreasing since 2013, per VERIS reporting, as shown in Figure 3.1.

**Takeaway:** The decrease in incident reports post-2013 might not indicate a genuine drop in cybersecurity threats but could reflect gaps in how incidents are reported or evolving complexities in breaches. This trend underscores the critical need for improved detection and reporting systems in cybersecurity, ensuring that any decline in reported incidents is a true measure of enhanced security, not a result of oversight or underreporting.

**Timeline Discovery for Data Breaches.** Health organizations encounter various difficulties in attempting to keep patients' medical records safe. The *timeline discovery* affects both the patient and the hospital. The longer it takes for an organization to discover a data breach, the more significant harm it can cause. The damage cannot only result in data loss or the disclosure of information but also includes businesses. In the literature, it was shown that organizations take 197 days to identify a data breach and 69 days to contain it, on average [45]. That amount of time to detect a data breach is considered long and costs organizations millions of dollars. An organization containing the data breach incident in less than 30 days from the date it happened can save up to $1 million compared to others who fail to do so, per the same study. In healthcare, hospitals and organizations can suffer many consequences due to a data breach, including lawsuits from the affected individuals, as well as reputation and trust loss. In addition, healthcare organizations incur significant costs in fixing the problem and protecting patients from additional harm.

We examined the response time for incidents affecting victims, patients, customers, and employees. In the following, we present the results and contrast.

**Results.** We began by converting the timeline discovery into one unit (hours). Then, we calculate the cumulative distribution function (CDF) of incidents. Due to the extensive range of timelines, we used the logarithmic function to the discovery time range for simplicity and visibility, as shown in Figure 3.2. Based on this analysis, we noticed that the discovery time of incidents for employees is significantly faster than for customers and patients. As we can see in Figure 3.2, we discovered




(a) Patients (b) Employees (c) Customers

**Figure 3.2:** CDF for the timeline discovery of different victim types.

that 20% of the incidents for employees were discovered within four days or less. It took five days or less to discover the same percentage for customers and up to six days to discover that for patients. Such results indicate the difference between the different categories of breach discovery time and perhaps the priorities associated with their discovery and protection, although all are relatively high. To further establish that, for 50% of the incidents, the discovery time was 2, 2.5, and 3 months for customers, employees, and patients, respectively. The patients represent most victims with 41%, and the discovery time for their data breaches extends to years (14 years to discover 100% of all incidents). While discovering 100% of incidents for customers require a longer time: up to 21 years. On the contrary, the discovery time of incidents for employees is much less because discovering 100% of the incidents for this category is about ten years.

**Takeaway:** Incidents discovery, even in cases involving highly protected victims, can span several years, which points to a fundamental laxity in the security measures of healthcare organizations. Our findings highlight the need for a comprehensive strategy to maintain healthcare data security. Key steps include adopting sophisticated security solutions, cybersecurity training for personnel, and formulating effective incident response protocols. These measures are essential to expedite breach detection and ensure the robust protection of patient information.

## Security Attributes

The VERIS dataset uses pairs of the six primary security attributes: confidentiality/possession, integrity/authenticity, and availability/utility as an extension of the CIA triad. In this section, we attempt to investigate the compromised security attributes during the incidents by conducting the following: (i) analyzing the confidentiality leakage that occurred during data breaches, (ii) presenting the different data types, and noting which is the most targeted by adversaries, (iii) determine the state of the compromised data at the time of the incidents.

**Data Confidentiality.** Confidentiality refers to the limit of observations and disclosure of data [91]. We start by examining the data confidentiality leakage that occurred during data breaches. This analysis is necessary because it examines the amount of compromised data and their varieties throughout the incidents. Using the VERIS dataset, we found that 1,045 out of total data, 1,937 incidents had *information disclosure*, representing 54% of the total incidents, 882 had a *potential information disclosure*, representing 46%, while only two incidents that had *no information disclosure* at all and eight incidents are *unknown*.

We analyzed data that attackers often target. Based upon this analysis, we discovered the following: medical information exposed to higher disclosure compared to the other types of information, encompassing 1,413 incidents, representing 73%, while personal information appeared in second place, with 345 incidents, representing 18%. Lastly, payment information appeared in third place, having 61 incidents, representing 3%. Other targeted information include *unknown* (44; 2%), *banking* (33; 2%), *credentials* (23; 1%), and *others* (18; 1%).

> **Takeaway:** Medical and personal data are the prime targets in 91% of breaches, highlighting a critical need for enhanced security measures in healthcare data protection against identity theft and privacy violations.

**Status of Breached Data.** During the exposure or compromise process, we investigated the state of the data and whether it was encrypted, transmitted, or stored unencrypted during the attack. This categorization aims to understand the security controls while the data is at rest or in motion due to transformation. As a result of this investigation, we noticed 36% of the data was *stored unencrypted*, 30% *stored*, 25% *unknown*, 3% *printed*, 2% *transmitted unencrypted*, and 4% with other attributes.

> **Takeaway:** A significant portion of the data breached was either stored or transmitted without encryption, highlighting a major gap in fundamental security practices and making it vulnerable to exploitation.

**Data Integrity and Authenticity.** Integrity refers to an asset or data to be complete and unchanged from the original state, content, and function [91]. Examples of loss of integrity include but are not limited to unauthorized insertion, modification, and manipulation. We wanted to discover the varied nature of integrity loss. Each time incidents occur, there can be at least one integrity attack. However, many losses can be associated with a single incident. Following the analysis, we noticed that most data integrity losses are due to altering behaviors containing 93 incidents, representing 31% of the overall. Software installation comes in second with 91 incidents, representing 30% of the known reasons. Other integrity related attacks include *fraudulent transmission* (18%), *data modification* (11%), *re-purposing* (3%), and *others* (6%).

Authenticity refers to the validity, conformance, correspondence to intent and genuineness of the asset (or data). Losses of authenticity include misrepresentation, repudiation, misappropriation, and others. Short definition: Valid, genuine, and conforms to intent [91]. Based upon this analysis, we observed that the authenticity state was poorly reported at the time of the incidents.

**Data Availability.** Availability refers to an asset or data being present, accessible, and ready for

use when needed [91]. A loss to availability includes deletion, destruction, and performance impacts such as delay or acceleration. We will show the variety of data available that might happen during the incidents. This analysis is necessary to understand the nature or type of availability or utility loss. Based on this analysis, we found that 769 incidents contained a loss of data regarding their effect on availability, representing 90% of the total incidents with the reported attribute. *Obfuscation*, and *interruption* are reported as remaining causes affecting availability, with 9% and 1% of all incidents, respectively.

> **Takeaway:** Despite limited reporting, more than 20% of all the studied incidents suffer from integrity and authenticity attacks due to a range of factors, magnifying the potential of attacks without data leaving the organization.

### Analyzing the Threat Actors

Threat actors are entities that can cause or contribute to an incident [91]. There can be at least one of the three threat actors involved, but on some occasions, there can be more than one actor involved in a particular incident. Threat actors' actions can be malicious, non-malicious, intentional or unintentional, causal or contributory [91]. VERIS classifies threat actors into three main categories, namely: external, internal, and partner. This classification excludes the contributory error that unintentionally occurs. For instance, if an insider unintentionally misconfigures an application and leaves it vulnerable to an attack. The insider would not be considered as a threat actor if the applications were successfully breached by another actor [91].

On the other hand, an insider who deliberately steals data or whose inappropriate behavior (e.g., policy violations) facilitated the breach would be considered a threat actor in the breach [91]. This section will explain and analyze each category of threat actors with their presence in incidents

from our dataset. This analysis is essential because of the following reasons: (i) it provides us with an understanding of the reasons or motives that can lead actors to act, (ii) the analysis can provide knowledge for organizations to consider proper precautions to defend against how threat actors operate. Several motives can be a reason for a data breach, such as fear, ideology, grudge, espionage, convenience, fun, and financial. Based upon this analysis, we noticed that the financial motive is the primary motive for adversaries, followed by looking for fun.

**External Actors.** External threats originate from outside of an organization and its third-party partners [91]. Examples include criminal groups, lone hackers, former employees, and government entities. It is also comprised of God (as in "acts of"), "Mother Nature," and random chance. Typically, no trust or privilege is implied for external entities. We found out that 97% of the external actor motives are financial, and 1% are for fun. Figure 3.3 shows the different motives of the actor's external motives.

**Internal Actors.** Internal threats originate from within the organization, which encompasses full-time company employees, independent contractors, interns, and other staff. Insiders are trusted and privileged (some more than others). Upon further analysis, we found that 52% of the internal motives for adversaries are financial, while 27% are for fun. Figure 3.3 presents the distribution of motives for internal motives.

**Partner Actors.** Partners include any third party sharing a business relationship with the organization, including suppliers, vendors, hosting providers, outsourced IT support, and others. Trust and privilege often exist between these partners [91]. Based on this analysis, we found out that most of the motives behind the incidents are financial 67%; fun and convenience are 17% and 8%, respectively. The remaining results for the internal motives distribution are shown in Figure 3.3.

**Results: Data Breaches Victims.** We analyzed the most targeted victims from adversaries accord-

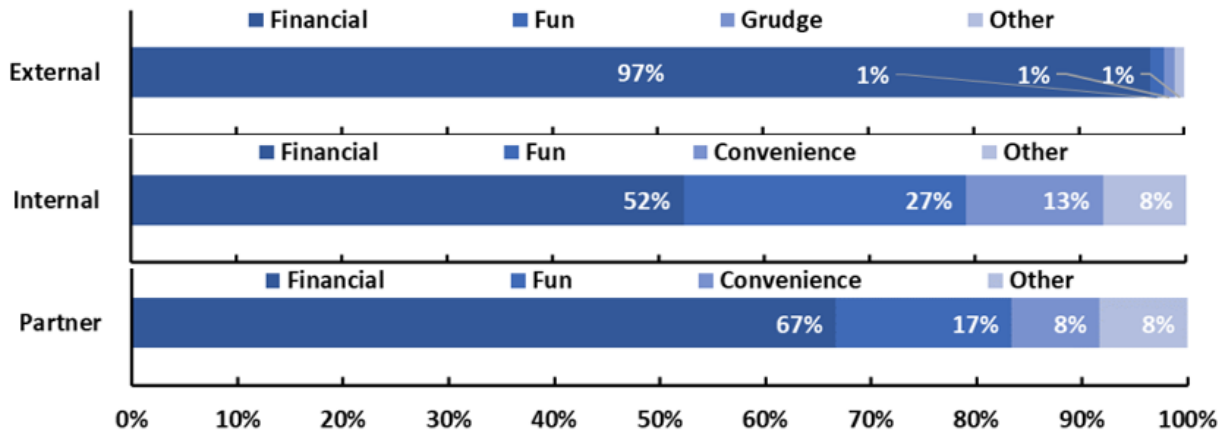



**Figure 3.3:** Threat actor's motives for external, internal, and partner actors.

ing to the number of incidents. Reasons often differ as to why these victims have been targeted, and it also depends on several other aspects, such as location, specific personal information, or a high number of patients in a hospital. We found that most of the targeted victims were patients (88%), the customer came in second (5%), and 5% for employees. Other types of victims include students (interns) working inside healthcare organizations or third-party companies that share data with a specific entity. Figure 3.4 shows the most targeted victims in the incidents.

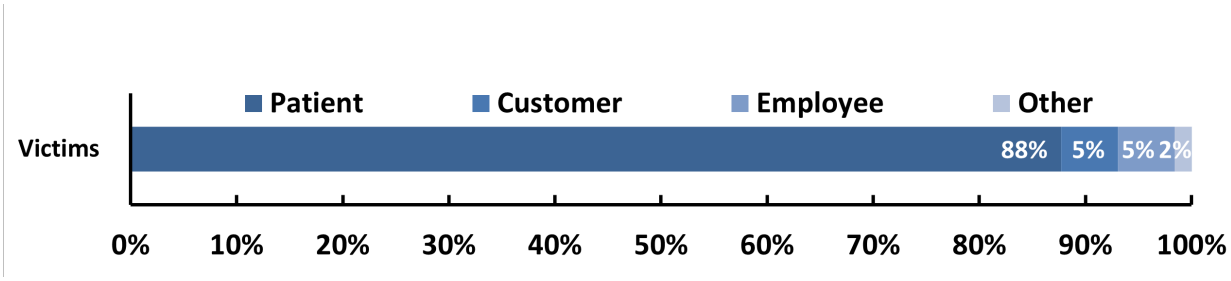



**Figure 3.4:** Distribution of incidents by targeted victims.

**Takeaway:** Patients are overwhelmingly the most targeted group in healthcare data breaches, accounting for 88% of incidents. This highlights the urgent need for stronger data protection measures for patient information. Additionally, considering the three threat actors combined, financial gain emerges as the primary motive for these adversaries to launch their attacks, further emphasizing the need for robust security protocols.

### Analyzing Threat Actions

In this section, we introduce our measurement and analysis of the threat actions used by adversaries during a data breach. This investigation intends to provide insight and the causes of threat actions and their occurrences in our dataset. The following section discusses the two types of threat actions: the different action varieties and the most used vectors by adversaries during an attack. The VERIS dataset classifies threat actions into seven primary categories: malware, social, hacking, misuse, physical, error, and environmental. Analyzing threat actions is essential due to the amount of risk associated with each of them every time an incident occurs. Generally, an incident usually contains a least one of the threat actions; however, most of the incidents will comprise multiple actions that often come with numerous categories.

**Terminology Definitions.** Below, we define several types of threat actions.

**Malware** Malicious software or malware is a computer code designed to disable, disrupt, or take control of the computer system by altering its state or function without the owner's informed consent [91]. Malware exploits technical flaws or vulnerabilities in hardware or software.

**Hacking** Refers to all attempts to intentionally access or harm information assets without (or exceeding) authorization by circumventing or thwarting logical security mechanisms. It includes brute force, SQL injection, cryptanalysis, denial of service attacks, etc. [91].

**Social** Social engineering criminals strive to exploit the users of these technologies by pretending to be something they are not to persuade others. Attackers utilize the trust to their advantage by misleading users into disclosing information that compromises data security. Social engineering tactics employ deception, manipulation, intimidation, and other techniques to exploit the human element, or users, of information assets, including pretexting, phishing, blackmail, threats, scams, etc. [91].

**Misuse** The use of entrusted organizational resources or privileges contrary to their intended purpose is misuse. This includes administrative abuse, policy violations, and use of non-approved assets. These actions can be malicious or non-malicious. [91].

**Physical** Encompass deliberate threats that involve proximity, possession, or force. These include theft, tampering, snooping, sabotage, local device access, assault, etc. [91]. Natural hazards and power failures are classified into physical actions. However, VERIS restricts these events to intentional incidents only caused by human actors.

**Error** Error broadly encompasses anything done (or left undone) incorrectly or inadvertently. It includes omissions, misconfigurations, programming errors, malfunctions, etc. [91]. It does not include any intentional incidents.

**Environmental** The environmental category includes any natural events such as earthquakes and floods and hazards associated with the immediate environment or infrastructure in which assets are located. The latter encompasses power failures, electrical interference, pipe leaks, and atmospheric conditions.

**Results: Threat Actions Analysis.** We measured the existence of each threat action category by calculating their varieties and vectors used in the incidents. We observed that ransomware represents 82% of the malware threat, followed by others 8%. VERIS "other" to define any enumeration

not represented by one of the categories in the data set. For the social threat actions category, with a percentage of 69%, phishing plays a large part in threat actions. The use of stolen credentials represents 80% of the hacking threat actions. With an increase in the number of employees, errors increased. Loss errors represent the main factor in this threat actions category representing 28%, followed by a disposal error of 27%. It is worth noticing that theft is in the physical threat actions category with a percentage of 96%. Finally, privilege abuse in the misuse category, with a rate of 59%, is behind most of the threat actions in these two categories.

On the other hand, when we analyze the threat action vectors as shown in Figure 3.6, we found out that the direct install represents 45% of the malware threat actions. Email attachment is the second most common malware breach vector with 32%. The email vector represents 81% in the social category for other categories of threat actions, and web applications represent the primary vector with 81% of all hacking threat actions.

Carelessness is the primary vector with 92% of the error category. Although most data breaches using hacking by threat actors involve brute force or the use of lost or stolen credentials [33], At the same time, LAN access is the most effective vector in the misuse category with 65%. It is clear that email and web application vectors represent the highest percentages among other vectors, and this is associated with the shift of valuable data to the cloud, including email accounts and business-related processes [33].

> **Takeaway:** Ransomware emerges as the predominant malware method in 82% of incidents, highlighting its critical role in cyber threats. Additionally, the misuse of email and web applications as vectors, especially in the context of the cloud, underscores a significant shift in threat dynamics, indicating an urgent need for enhanced security measures in these areas.

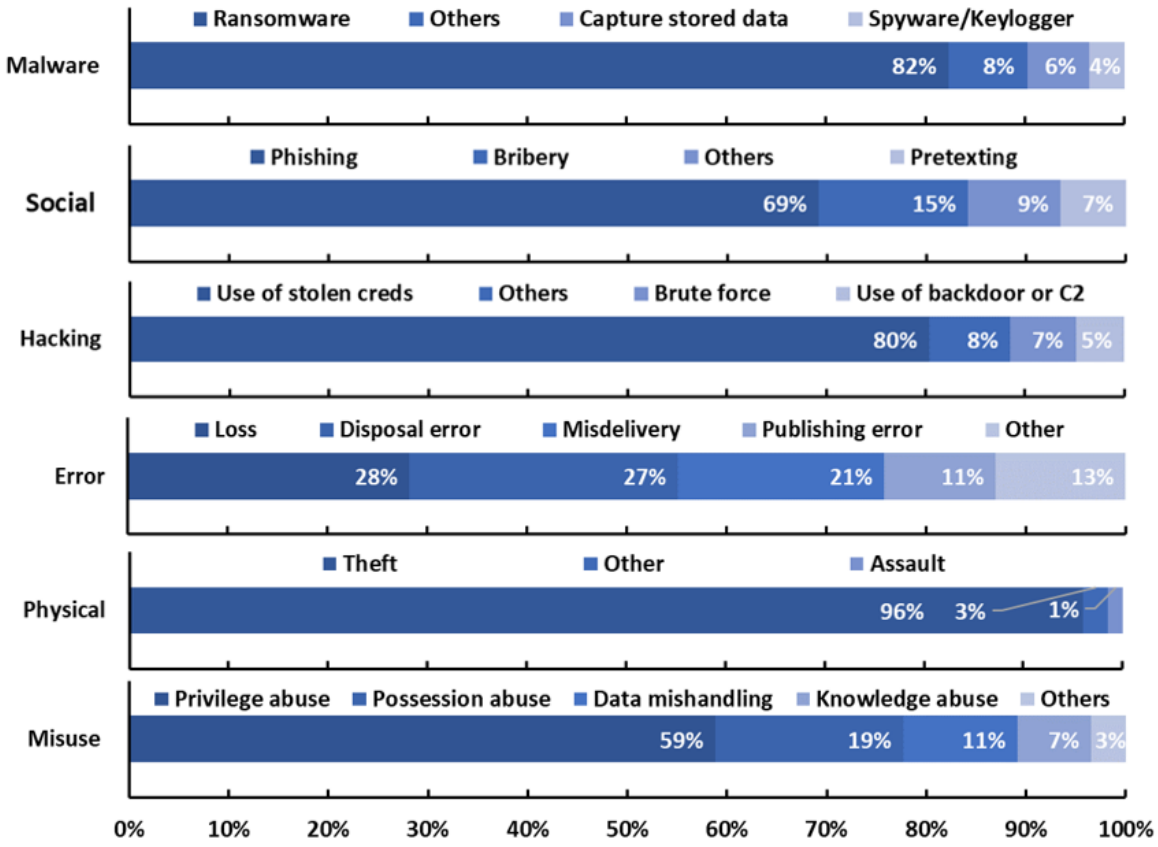



**Figure 3.5:** The variety of threat actions employed in data breach incidents.

Summary and Concluding Remark

While analyzing the timeline of the data set that comprised all the data breaches, the results showed that the highest number of incidents occurred from 2010-2020. Moreover, this long-term study revealed that health organizations are exposed to internal, external, and partner attacks. The financial is the primary motivation for the external, internal, and partner attackers. Without a doubt, there is a high cost associated with data breaches; the price for each stolen health record increases with time. Based on a long-term analysis of the data set, the actions used by the threat actors are classified into seven categories: malware, hacking, social, misuse, physical, error, and environmental.

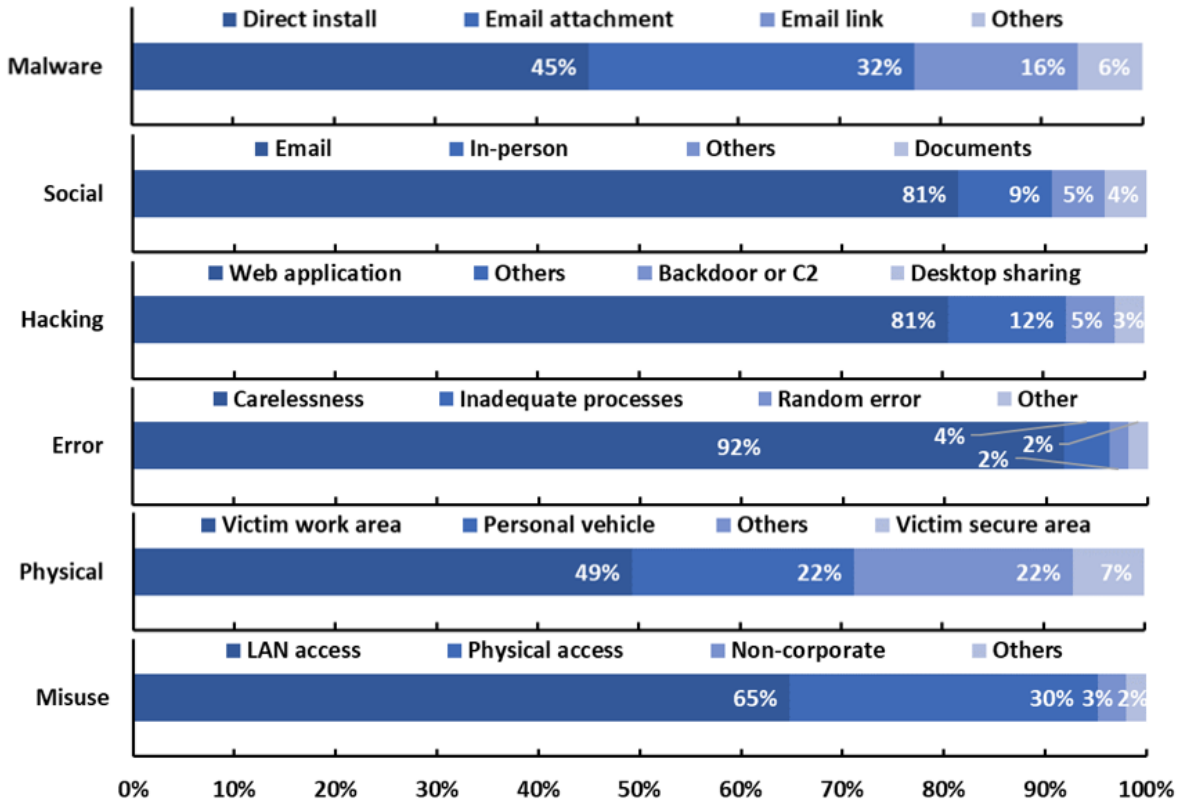



**Figure 3.6:** The threat action vectors employed in data breach incidents.

Ransomware motivated 82% of malware threat actors, and 45% of malware threat actions are directly installed. In the future, it would be worthwhile examining the correlation between security breaches and other indicators, including GDP, hospital size, etc.

# CHAPTER 4: SECURITY BREACHES IN THE HEALTHCARE DOMAIN: A SPATIOTEMPORAL ANALYSIS

Electronic health records (EHR) can be described as "a longitudinal electronic record of patient health information generated by one or more encounters in any care delivery setting. Included in this information are patient demographics, progress notes, problems, medications, vital signs, past medical history, immunizations, laboratory data, and radiology reports" [59]. The adoption of EHR improves the healthcare industry and patients alike, and the transformation of healthcare organizations from paper-based to digital has increased healthcare quality by improving patient care and participation, care coordination, diagnostics, patient outcomes, and practice efficiency. However, despite the numerous benefits of EHR, this transformation has led to numerous privacy and security issues which may arise from vulnerabilities (e.g., software vulnerabilities, insider threats, human error, etc.), increasing the possibility of cyber-attacks [48]. The alarming surge in healthcare data breaches has caused huge concerns in the healthcare sector due to the illegitimate and unauthorized disclosure of private healthcare data [8, 80].

Healthcare Data breaches can be classified as either internal or external, and they can occur as a result of theft of private health records, hacking, loss of sensitive patient data, and unauthorized access to patient's private information [98]. External cybersecurity incidents are committed by cybercriminals operating in the dark web, while internal data breaches result from something internal to an organization, such as disgruntled employees, malicious insiders, employee negligence, and human error. Patient medical records and personal information are often targeted in healthcare data

This work has been published at the 10th International Conference on Computational Data and Social Networks (CSoNet 2021)

breaches due to their sensitivity and value. External attacks aim to steal those records and demand a ransom or sell those records for hundreds of dollars per single patient on the dark web [82].

Data breaches are devastating and can cause significant damage to healthcare organizations; all the research in this domain demonstrates that the healthcare industry is the most targeted sector due to the attractive financial return of selling sensitive patient records on the dark web [95]. Additionally, the lenient security controls deployed by healthcare organizations further complicate matters and make the healthcare domain a favorite target for hackers. The cost of recovering from such breaches varies greatly by the nature of the incident and the number of compromised health records. To better understand the cost aspect, we can break down the cost of data breaches for healthcare entities into two categories: direct costs and indirect costs. Direct expenses include activating incident response teams, engaging forensic experts, outsourcing hotline support, and providing free credit monitoring subscriptions and discounts for future products and services. On the other hand, indirect costs include in-house investigations and communication, as well as the extrapolated value of customer loss resulting from turnover or diminished customer acquisition rates [40]. Given these facts, it's compelling to conduct extensive research studies into the causes, effects, and consequences of healthcare data security incidents. Perhaps more importantly, gaining insights into the different trends and the landscape, and understanding, analyzing, and measuring the statistics in data breaches is crucial for combating such incidents. This is the motivation of this paper, and we also wish to motivate the research community in this space to extend the body of knowledge by conducting more studies to be able to understand data breaches better and propose solutions in the fight against cybercrimes.

**Contributions.** To understand the landscape of healthcare data breaches against several attributing characteristics, we provide a detailed measurement-based study of the VERIS and the Office of Civil Rights (OCR) datasets. To understand attackers' intents and motives, we analyze the type

of assets targeted during breaches over various characteristics to investigate their effect. We also analyzed data breaches considering multiple views looking at their distribution, affected entities, breached information, location of the breach, etc.

## Summary of Completed Work

In this work, we focused on studying the factors influencing data breach incidents in the healthcare sector. Our study revealed that the number of adults and the population of a state played a significant role in the exposure to data breaches, with California, Florida, and Texas being the primary targets. Additionally, our analysis revealed that the media group was the most breached asset, followed by the Server and User group. Interestingly, a majority of the incidents occurred in small-size organizations (57%), while (43%) of the incidents took place in large organizations, indicating that larger healthcare organizations tend to have better security systems. This work provides insights into the factors influencing data breach incidents in the healthcare sector. The findings emphasize the need for robust security measures in smaller healthcare organizations and highlight the need for prompt detection and response to mitigate the impact of data breaches.

## Data Sources

One of the challenges with analyzing cybersecurity incidents, in general, and in the healthcare sector, in particular, is that most datasets are proprietary [94]. Additionally, most breached healthcare organizations shy away from disclosing their vulnerabilities after a breach due to a variety of concerns, including public image, reputation, and patient trust. The other challenge lies in the fact that each victim healthcare entity tends to take a different approach in analyzing and documenting a data breach [95]. This, in turn, complicates research efforts because data breach statistics are not

stored in a central online repository and are thus inaccessible to the broader research community. To address the above challenges and conduct our measurements and analysis of data breaches, we turn to the largest publicly available datasets of cybersecurity incidents, namely, the VERIS dataset and the OCR dataset, which we describe below.

**VERIS.** We obtained a reliable data source to conduct our research, namely, the Vocabulary for Event Recording and Incident Sharing (VERIS). Veris provides a common language for reporting data breaches incidents in an organized and repeatable manner [57]. Thus, Veris plays a significant role in providing a solution to one of the most critical and persistent challenges in the security industry; lack of quality information. Veris contributes to the solution of this problem by helping organizations collect helpful incident-related details and share them anonymously and responsibly with others. Veris's primary goal is to lay a foundation to constructively and cooperatively learn from our experiences to ensure the proper measurements and managing risk [22].

**Office of Civil Rights (OCR).** Our second dataset is obtained from the U.S. Department of Health and Human Services Office of Civil Rights. The U.S. Department of Health and Human Services (HHS) Office for Civil Rights (OCR) enforces federal civil rights laws, conscience, and religious freedom laws, the Health Insurance Portability and Accountability Act (HIPAA) Privacy, Security, Breach Notification Rules, and the Patient Safety Act and Rule, which together protect your fundamental rights of nondiscrimination, conscience, religious freedom, and health information privacy [71]. The OCR has its breach portal, where data breaches are reported. The website contains data breaches that are currently under investigation within the last 24 months by the OCR. There is also an archived dataset where resolved data breaches and/or those older than 24 months are archived. All the data breaches reported by the OCR are in the U.S. only. Additionally, all records in the subsequent data breaches affect 500 or more individuals; the OCR does not report minor data breaches that affect less than 500 individuals.

**Studied Dimensions and Variables.** This study aims to examine healthcare data breaches considering different aspects of threat characterization and modeling.

- **Geographical mapping:** This section analyzes the geographical mapping and distribution of incidents around the world. Analyzing the geographical mapping of the incidents is necessary for several purposes: (i) it provides us with an understanding of the areas most targeted by adversaries for affinity characterization, (ii) identifying locations around the world where the number of incidents varies due to valuable medical information, particular age group, banking details, etc. We can use this analysis for correlation and prediction capabilities.

- **State-level distribution:** This section measures the state distribution of incidents in the U.S. This analysis is necessary for (i) identifying the hot spots targeted by attackers and (ii) conducting correlation analysis between states.

- **Compromised assets**: This section details the targeted assets by breaches such as media, server, terminal, etc. Alongside, we will categorize the assets into groups, then dive into their varieties by an individual group against the number of incidents.

- **State-level correlation:** This section carries a correlation analysis of the number of incidents within the top ten states with characteristics such as population, Gross Domestic Product (GDP), number of adults, etc. This correlation provides us with essential insights into the reasoning and bearings for each state.

- **Healthcare provider size:** This section analyzes the number of breaches versus the size of organizations in terms of the number of employees. We intend to discover if the number of employees influences the frequency of data breach incidents.

- **Timeline discovery:** This section examines the response time for incidents affecting healthcare organizations. We will measure the amount of time taken until the discovery of inci-

dents. This analysis helps us determine the organization's security level and whether more extended discoveries cause more damage.

- **Discovery methods:** This section aims to identify the discovery mechanisms used by healthcare entities. Then, we will measure the reported tools and their use in data breaches in our dataset. This analysis can help with determining the appropriate tools needed to be implemented in organizations.

- **Adversary demography — The threat intent:** This section examines the intentions of attackers in data breaches, aiming to determine whether the incidents are targeted or are of an opportunistic nature.

## Measurement Results and Discussions

**The Global Distribution of Incidents.** Mapping incidents is provided in our dataset. The dataset uses the ISO 3,166 country codes for each country variable [37], where the codes are generated based on the physical location of the hospital targeted by the attack. Based upon this analysis, we discovered that 1,955 incidents out of the total incidents (2,407) took place in the United States, representing 81% of the total incidents. The United Kingdom comes in second, with 157 incidents, representing 7%, and Canada comes in third with 152 incidents, representing only (6%). Figure 4.1 presents the results for the remaining highest ten countries, while the rest of the world represents (2%) comprising 58 incidents.

As a result of the geographical mapping analysis, we decided to conduct our in-depth analysis study on the United States since most incidents occurred in this country. Several reasons explain why the majority of the incidents are in the United States. First, HIPAA requires healthcare entities to notify the Department of Health and Human Services (DHHS) whenever a data breach occurs.

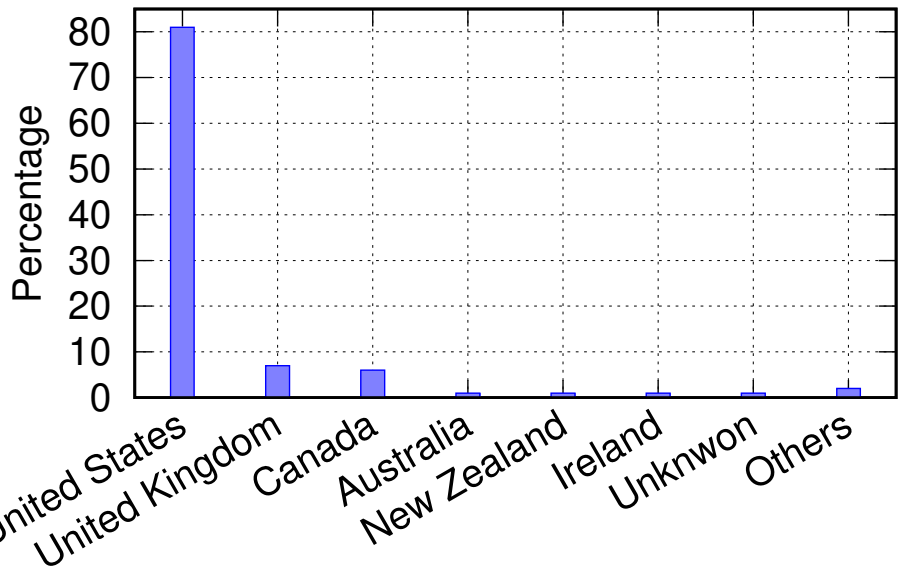



**Figure 4.1:** The geographical distribution of data breach incidents by country.

Second, covered entities must notify affected individuals following the discovery of a breach of unsecured protected health information [71]. In addition to that, covered entities must notify the Secretary of breaches of unsecured protected health information if the affected individuals are 500 or more [71]. Third, covered entities that experience a breach affecting more than 500 residents of a State or jurisdiction are, in addition to notifying the affected individuals, required to provide notice to prominent media outlets serving the State or jurisdiction [71]. Moreover, breach notification is also required for vendors and third-party service providers under the Health Information Technology for Economic and Clinical Health Act (HITECH) [58]. Finally, the HIPAA Security Rule requires healthcare organizations to create a risk management plan protecting all personal health data against security incidents (Office of Civil Rights 2015), which may explain the significant number of reported incidents in the United States [1].

**Number of Incidents by State.** Following the global distribution of incidents, we moved into

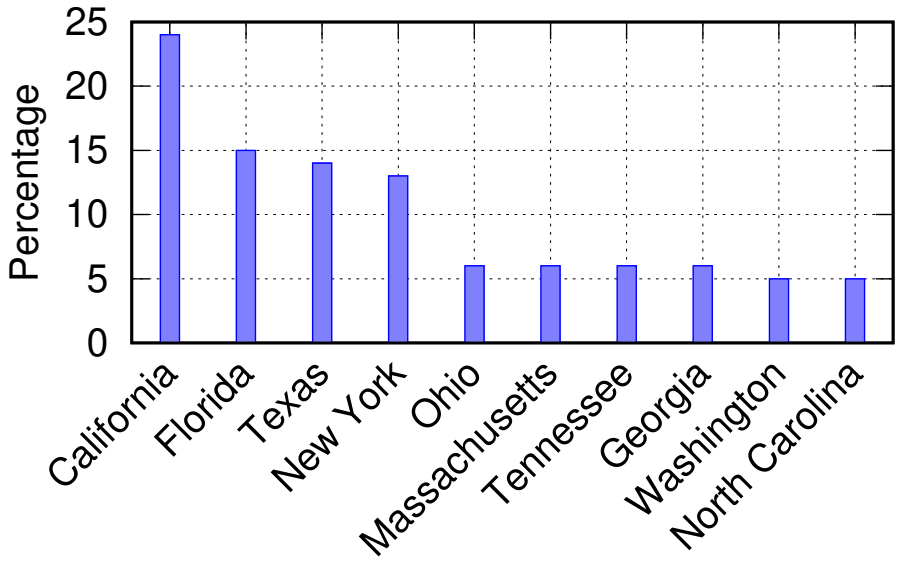



**Figure 4.2:** The distribution of data breach incidents across different states.

the mapping of incidents on the state level. We analyzed the number of incidents by state. As a result of this analysis, we noticed that California is the highest state, with the number of incidents comprising 241 incidents, representing 24% of the overall. Florida comes in second with 147 incidents, representing 15%, and Texas with 145 incidents, representing 14%. Figure 4.2 shows the remaining results of this analysis.

**Analyzing the Compromised Assets.** This section investigates the compromised information assets in the Veris dataset. We harnessed the power of natural language processing (NLP) models to help analyze the data gathered from breaches. Information assets fall into six main groups: media, server, terminal, network, user, and people. Each group comprises different varieties [78]. First, the network group includes access control readers such as badge and biometrics, camera or surveillance systems, intrusion detection systems (IDS), or intrusion prevention systems. Second, the media group comprises disk media such as CDs or DVDs, flash drives or cards, hard disk

drives, identity smart cards, and others. Third, the people group includes the administrator, auditor, cashier, customer, former employee, guard, and others. Fourth, the server includes authentication, backup, database, Dynamic Host Configuration Protocol (DHCP), DNS, mail, and others. Fifth, the terminal group includes an automated Teller Machine (ATM), detached PIN pad or card reader, gas "pay-at-the-pump" terminal, self-service kiosk, and others. Finally, the user group includes an authentication token or device, desktop or laptop, media player or recorder, mobile phone or smartphone, and many others.

The existence of assets depends on several reasons and conditions during each incident. We will measure each asset group based on their occurrences in the incidents, and then, we get into the measurement of their varieties to look into the most targeted type of each asset group. This analysis is essential, and its primary purpose is to adequately describe the incidents, assess control weaknesses and vulnerabilities, determine impact, and identify mitigation strategies.

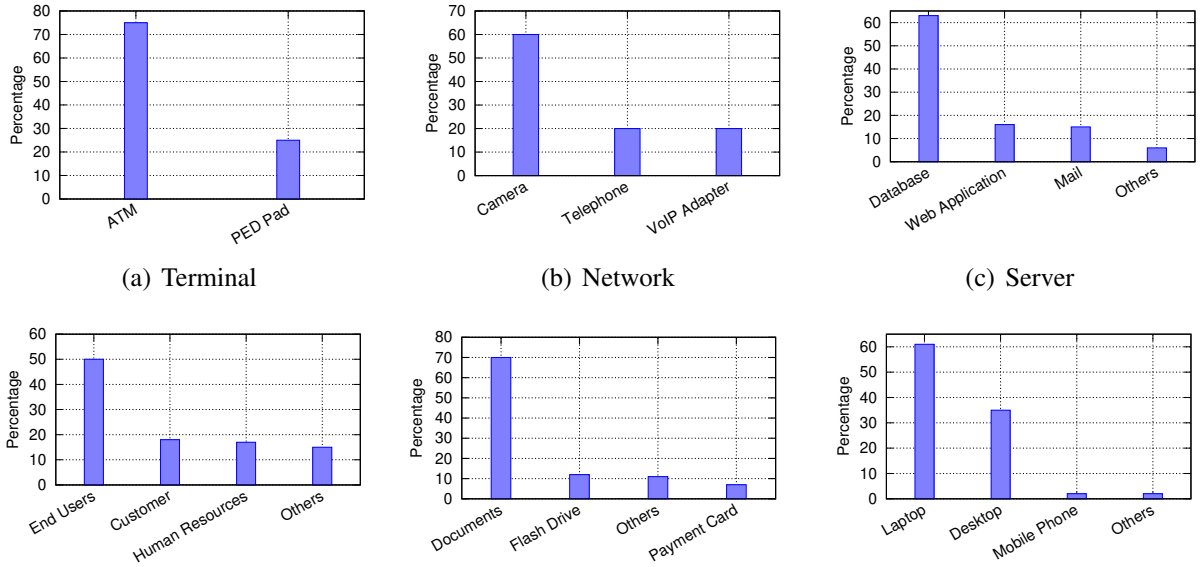
Usually, during a data breach incident, one or more assets get compromised by hackers [38]. A compromised asset refers to any loss of confidentiality, integrity, or availability during or after the incidents. In the following section, we seek to analyze and measure the asset groups and the total incidents for each group; then, we move to their different asset groups. Based on this analysis, we noticed that media assets are the clear leader comprising 564 incidents out of the overall, representing 33.97%, and the server comes in second, comprising 560 incidents, representing 33.73%. Table 4.1 shows the remaining asset categories and their number of incidents.

After measuring the number of incidents for each asset group as a whole, we moved into measuring their varieties. Based on the analysis done, we found that 61% of the incidents in the user group are through laptops, followed by the terminal group with 75% of the incidents through ATMs. In the server asset group, we found out that 63% of the incidents happened through exploiting the database. While for the people asset group, 50% of the incidents are because of the end-user. Most

**Table 4.1:** The distribution of incidents by asset group type during data breaches.

| Asset Group Type | # Incidents | Percentage |
|---|---|---|
| Media | 564 | 33.97% |
| Server | 560 | 33.73% |
| User | 493 | 29.69% |
| People | 34 | 2.04% |
| Network | 5 | 0.30% |
| Terminal | 4 | 0.24% |
| Overall | 1660 | 100 % |

of the incidents that happen in the network are throughout cameras, representing 60%. Lastly, 70% of the incidents in the media group are through documents. In Figure 4.3, we present the remaining results for the other asset groups and their varieties.




(a) Terminal (b) Network (c) Server

(d) People (e) Media (f) User

**Figure 4.3:** The varieties of information asset groups involved in data breach incidents.

**State Level Correlation.** This section will conduct a state-level correlation between the number of reported incidents and hospitals, staffed beds, population, and gross domestic product (GDP)

**Table 4.2:** State level correlation. Numbers of incidents (I), hospitals (H), employees (E), staffed beds (B), GDP (G), population (P), and adults (A) are considered.

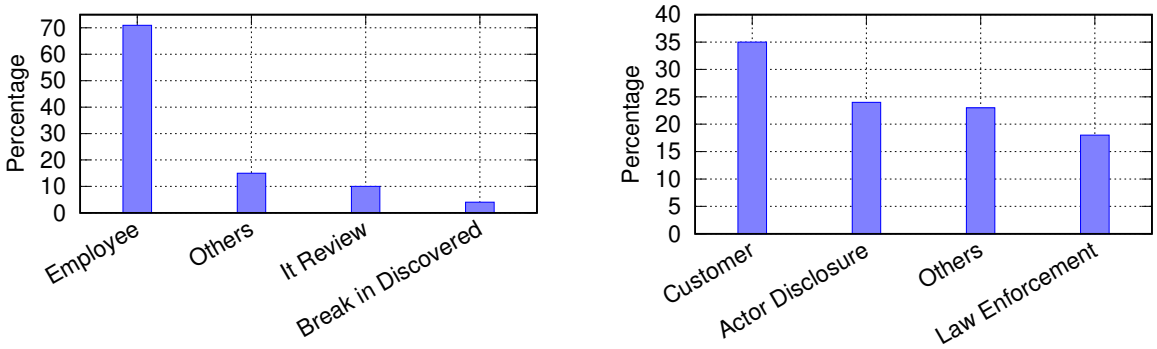
| | **I** | **H** | **E** | **B** | **G** | **P** | **A** |
|---|---|---|---|---|---|---|---|
| **I** | 1.00 | | | | | | |
| **H** | 0.88 | 1.00 | | | | | |
| **E** | 0.92 | 0.91 | 1.00 | | | | |
| **B** | 0.94 | 0.92 | 0.97 | 1.00 | | | |
| **G** | 0.95 | 0.86 | 0.92 | 0.89 | 1.00 | | |
| **P** | **0.96** | 0.95 | 0.94 | 0.94 | 0.95 | 1.00 | |
| **A** | **0.96** | 0.88 | 0.94 | 0.96 | 0.89 | 0.90 | 1.00 |

for the top 10 states. GDP is the gross domestic product and is represented in billion U.S. dollars. To address the following question, we conducted a state-level analysis considering these factors related to the reported incidents in our dataset. We decided to run this analysis on the highest ten states in terms of the number of reported incidents. We started by collecting the specified statistics for each state, including population, GDP, staffed beds, and hospitals. The relationship between two variables can be a positive relationship (1), no relationship (0), and an inverse relationship (-1). Upon this analysis, we discovered that the population and adults are highly correlated with the number of incidents (0.96). Followed by the GDP (0.95). The remaining results of the correlation are shown in Table 4.2.

**Organizations size.** The following section investigates the size of healthcare entities and how an organization's size might contribute to a data breach. Using Veris, we performed the analysis by looking into the scope of healthcare organizations at the time of the incident. We classified healthcare organizations into two main groups: small and large. A small group includes a size of up to 1,000 employees, while a large organization would have over 1,000 employees. Upon this analysis, there were a total of 1,361 incidents divided into two groups. Our analysis revealed that 57% of the incidents are in the small group, while 43% are in large groups.

**Timeline Discovery.** Timeline discovery of data breaches varies depending on the type of industry, geography, and level of security of an organization. According to a recent study conducted by the IBM security team in the healthcare sector, the average time to discover a data breach is 329 days, and 93 days are required to regain control. Unfortunately, prior work fails to provide an in-depth analysis of the timeline discovery of the data breaches, including discovery tools for external and internal incidents. To fill this gap, we analyzed the timeline discovery of the reported incidents and went over the tools used for incident discovery for both internal and external discovery methods. This analysis is essential to address the lessons learned during the incidents and remediation process and provide organizations with insights and corrective actions to improve their detection and defensive capabilities. Our analysis found that organizations fail to identify data breaches early enough, resulting in more damage. From the reported incidents, we discovered that 3% of the incidents took minutes until discovery, 9% took hours, 15% took days, 6% took weeks, 52% took months, and 15% took years. In the coming section, we will address different discovery methods and whether there is a difference between internal attacks and external attacks.

**Internal and External Discovery Methods.** Discovery methods fall into two main categories; internal and external. Organizations use tools to discover an incident depending on the type of data breach. External and internal data breaches are different, and each one of them requires special discovery tools. First, healthcare organizations use numerous tools to discover incidents for internal incidents, such as Host IDS or file integrity monitoring, network IDS, and IPS alerts. In contrast, practices including law enforcement, actor disclosure, and customer notifications can help discover external incidents. Our analysis found that most of the internal incidents are discovered by employees, representing 71% of the total incidents. In contrast, customers discover 35% of the external incidents, and actor disclosure comes in second, representing 24%. The remaining results of this analysis are shown in Figure 4.4.




(a) Internal discovery.

(b) External discovery.

**Figure 4.4:** Comparison of data breach discovery methods.

**Targeted vs Opportunistic.** To understand the nature of the data breach incidents and whether they are intentional or non-intentional, we conducted a measurement analysis to investigate the number of targeted incidents and opportunistic ones. This classification is uniquely relevant to deliberate and malicious actions. There are two main categories: targeted and opportunistic. First, opportunistic incidents occur when the victim exhibits a weakness that the actor has the knowledge to exploit. Second, targeted incidents happen when the adversary chooses the victim as a target, and then the actor will investigate possible vulnerabilities to exploit. Using our exclusively given records in our dataset, we found that more than half of healthcare data breaches are opportunistic, representing 80%, while, on the other hand, 20% are targeted.

Analysis of the OCR Dataset

***Type of Breach.*** We analyzed the causes of healthcare data breaches based on the reported incidents and observed that most incidents occur due to hacking or IT-related disclosure, comprising 1,069 incidents, representing 31% of the overall incidents. Unauthorized access and disclosure

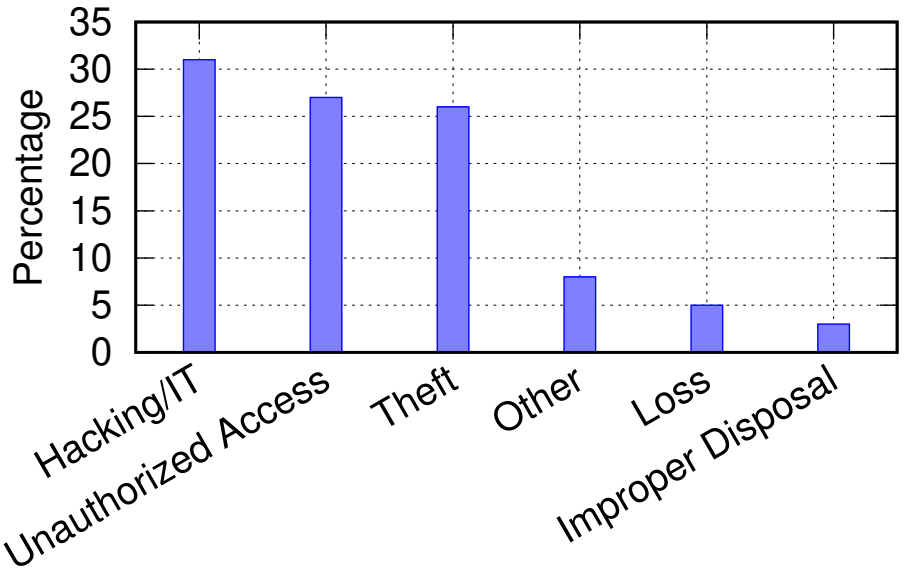



**Figure 4.5:** The distribution of breach types within the healthcare sector.

came in second, holding 934 incidents overall, representing 27%. Finally, the theft category came in third place, comprising 909 incidents, accounting for 26% of the total incidents.

***State Distribution.*** The following section addresses the distribution of the incidents for the U.S. states. Using the OCR data, we measured the incidents for each state; this analysis is essential for trends and comparison. Following this analysis, we have observed that states with a large population, high Gross Domestic Product (GDP), and large adult population are more targeted than others, as shown in section 4. California was the most affected, totaling 357 incidents, followed by Texas with 279 incidents, while Florida was the third largest with 215 incidents.

***Distribution of Incidents by Year.*** Using the ORC dataset, and over the period between 2009 and the time of conducting this study in 2021, we measured the reported incidents in the dataset affecting 500 or more victims and reported them to the HHS OCR. Following this analysis, we notice that the number of incidents surged over time, indicating a lack of implementing stringent

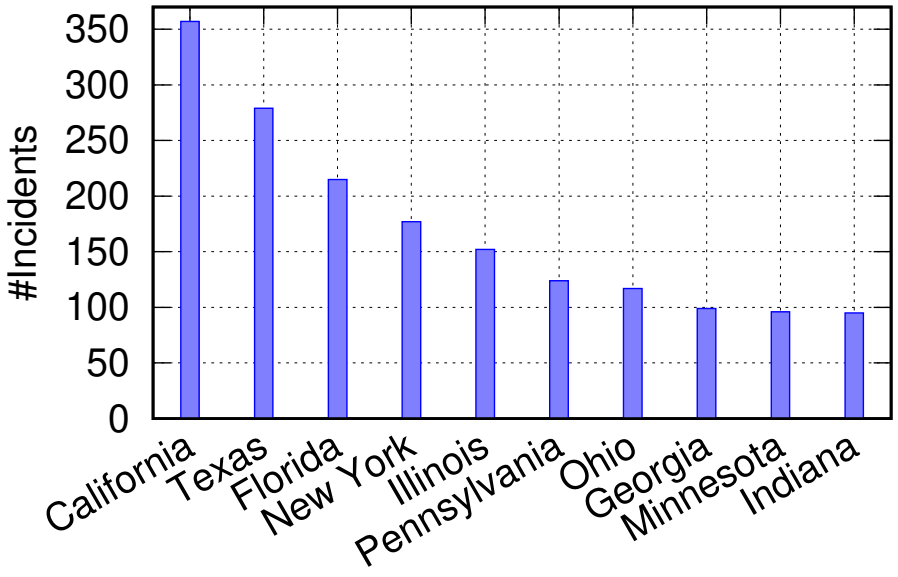



**Figure 4.6:** The distribution of data breach incidents across different states.

security controls by organizations in the healthcare industry. As shown in Figure 4.7, there was a massive increase in the number of incidents in 2019, as it was the year with the highest number of breaches in the whole dataset.

***Covered Entity.*** We analyzed the distribution of incidents by organization type. According to the OCR dataset, there are three main targeted entities. First, healthcare entities that provide healthcare services and engage in professional review activity through a formal peer review process for the purpose of furthering quality health care, a committee of that entity, a professional society, a committee or agent thereof, including those at the national, state, or local level, physicians, dentists, or other health care practitioners that engage in professional review activity through a formal peer review process to further quality health care [76]. Second, a business associate is a person or entity that performs certain functions or activities that involve the use or disclosure of protected health information on behalf of or provides services to a covered entity [67]. Third, health plan,

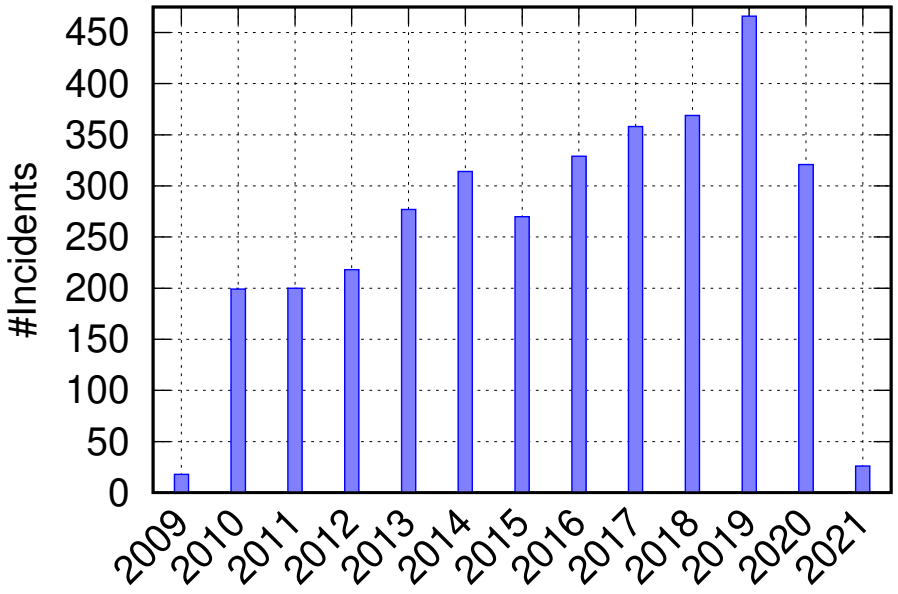



**Figure 4.7:** The yearly distribution of Data Breach Incidents.

which constitutes individual or group health plans that provide or pay the cost of medical care [54]. Following this analysis, we observed that healthcare entities are most targeted during the incidents, having 2,450 incidents which represent 73% of the total incidents; business associates and healthcare plans came in second and third, comprising 451 and 439 incidents and representing 14% and 13%, respectively. Figure 4.8 depicts the results of this analysis.

***Business Associates.*** We further analyzed the existence of incidents when a business associate is present or not. According to HIPAA, any covered entities and business associates enter into a contract to ensure the safety of protected healthcare information. A business associate may use or disclose protected health information only as permitted or required by its business associate contract or as required by law [88]. Our analysis revealed that 2,532 incidents had no business associates included, representing 76%, while only 819 incidents had a business associate, representing 24% of the incidents as shown in Figure 4.9.

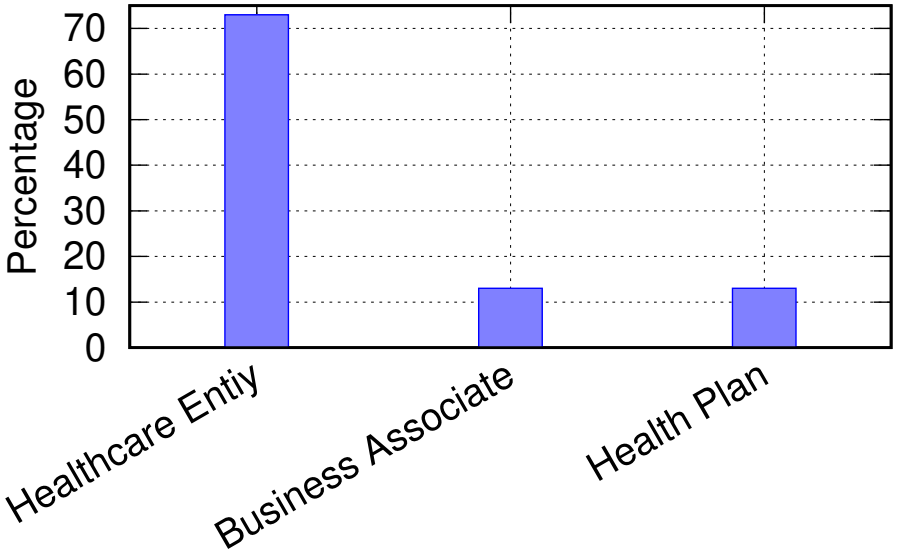



**Figure 4.8:** The distribution of covered entities.

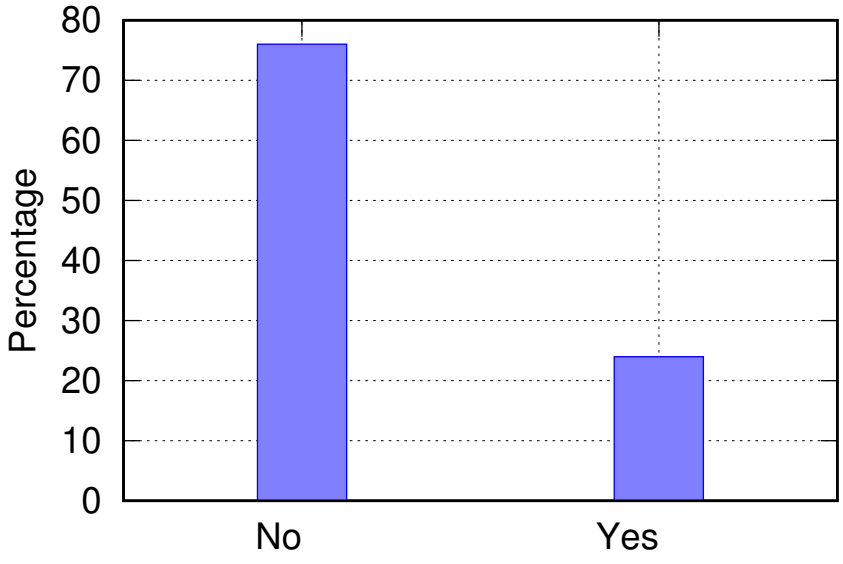



**Figure 4.9:** The distribution of business associates.

***Location of Breached Information.*** When a data breach occurs, private and confidential patient information gets disclosed due to either unauthorized access or human error. The healthcare system keeps records of valuable information and medical records containing sensitive personally identifiable information (PII) such as address history, financial information, social security numbers, and patient medical treatment records. Hackers often target this sensitive information due to its outstanding value. Hackers can easily use that data to set up a line of credit or take out a loan under patients' names. Unfortunately, healthcare organizations often lack the stringent security measures (e.g., encryption, robust anti-virus software, multi-factor authentication, etc.) required to secure medical records. To this end, we analyzed the most targeted information to gain insight into the type of medical and personal data prioritized by hackers in healthcare data breaches. We observed that paper/films are the most breached information comprising 662 of the overall incidents, representing 20%. Closely, the network server came in second, comprising 643 incidents, accounting for 19%. The other category came in third, comprising 641 incidents, representing 19% as well. The remaining attributes and results of this analysis are presented in Figure 4.10.

## Summary and Concluding Remark

Our study revealed that the number of adults and the state population highly influence the exposure to data breach incidents, with California, Florida, and Texas being the lead targets. We show that the media group was the most breached asset, followed by the Server and User group. Our timeline discovery revealed that most of the incidents, approximately 52%, were discovered within months, while 15% of the incidents took years to be discovered. Employees discovered the majority of the incidents for internal incidents. In the future, it would be interesting to conduct research harnessing the power of machine learning to enable information sharing on data breaches. We explored the relationship between the size of healthcare entities and their contribution to data breaches, analyzing

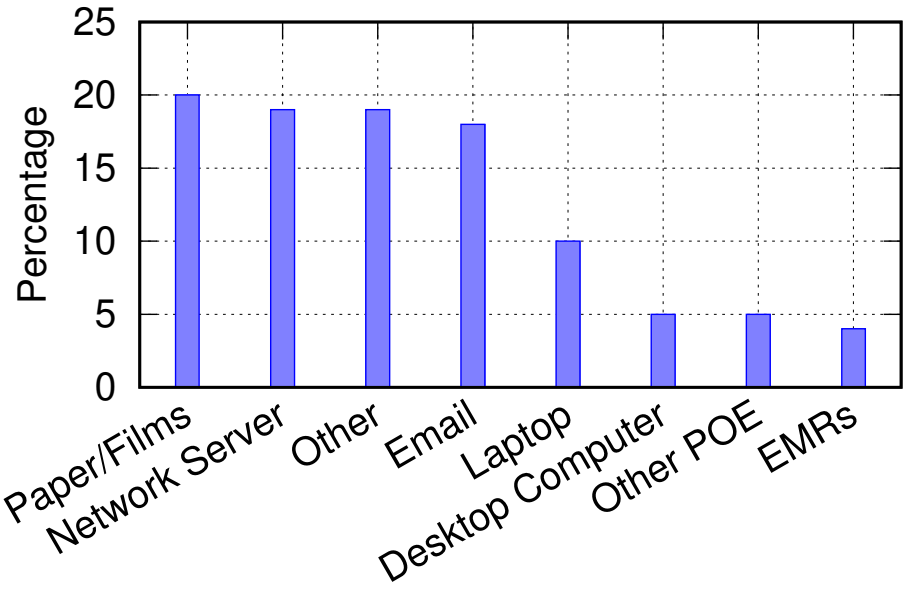



**Figure 4.10:** The distribution of breached information.

how an organization's size impacts the likelihood and severity of breaches. Moreover, we investigated the timeline of data breach incidents, aiming to identify patterns and trends in detection and reporting within healthcare organizations.

# CHAPTER 5: UNDERSTANDING THE SECURITY AND PERFORMANCE OF THE WEB PRESENCE OF HOSPITALS: A MEASUREMENT STUDY

Electronic Health Records (EHR) are longitudinal electronic patient health information records, which include patient demographics, progress notes, health problems, medications, vital signs, medical history, immunizations, laboratory data, etc. [59]. The adoption of EHRs has led to improved accessibility of healthcare information for patients and providers, resulting in higher quality patient care and more efficient coordination between hospitals. However, despite these benefits, the transformation to EHRs has also raised privacy and security concerns, particularly when EHR data is retrievable through website systems. EHRs centralize sensitive patient data, which can make them a prime target for cybercriminals seeking to steal or exploit this information. Additionally, EHRs can be accessed and shared across multiple healthcare providers, increasing the risk of data breaches and unauthorized access. Moreover, the implementation of EHR systems can introduce new vulnerabilities that may be exploited by cybercriminals.

For instance, vulnerabilities and software exploits in the healthcare domain have become a central focus of targeted cyber attacks [48], which can result in devastating data breaches. Given the sensitivity of private healthcare data, the unauthorized and illegitimate disclosure of this information can have catastrophic consequences [8, 15, 80].

Understanding the effect of healthcare data breaches is essential, and efforts in the literature classified those breaches into internal and external breaches. Internal breaches are commonly caused

This work has been published at the 32nd International Conference on Computer Communications and Networks (ICCCN 2023)

by human errors, particularly among healthcare employees. In contrast, external breaches, which are the more critical type, are caused by an unauthorized third party involved in the theft of private health records through hacking of the web-based user and healthcare provider-facing systems [98]. Cybercriminals typically commit these incidents, making their effect an open question, with no accurate assessment of their cost. For instance, adversaries involved in external breaches may aim to steal sensitive records and demand a ransom or sell those records for hundreds of dollars per single patient on the dark web [82, 95].

Given the importance of understanding data breaches in healthcare and the role of web technologies in enabling a significant part of the attack surface, this study is dedicated to analyzing the commonalities and differences among three types of hospitals: government public hospitals, non-profit hospitals, and proprietary hospitals. Namely, we analyze the websites and patients' portals for security configurations and common privacy practices. We note that compromising patients' portals allows the attacker to obtain sensitive information regarding the patient's records, including diagnoses, treatment records, hospital visits, and future appointments, alongside personal information. To the best of our knowledge, this work is the first in this direction, associating actual hospital potential exploitations and data breaches with website security and privacy configurations.

Our analysis is based on a total of 4,774 hospital websites grouped into three major hospital categories: government public hospitals, non-profit hospitals, and proprietary (private) hospitals. For our measurement assessments, we conduct both domain-level and content-level analyses to understand the similarities and differences among website attributes.

Our analysis is multi-faceted and covers a range of features by examining and comparing the website's domain SSL certificates, creation date, HTTP requests, page size, content type, average load time, and malicious activity association. The features explored in this analysis are particularly lightweight and do not require deep analyses of contents but rather focus on meta-attributes,

making our analysis techniques more generalizable to large-scale measurements. We further investigated the security attributes of these websites by exploring their association with malicious behaviors, including an assessment of the domain-based and content-based malicious behaviors of those websites and associated trends and characteristics.

To understand the implications of those characteristics, we further study their correlation with a manually vetted dataset of recently disclosed data breaches provided by the U.S. Department of Health and Human Services, the Office for Civil Rights (OCR) [36]. Leveraging information regarding the websites and associated breaches, we extracted the commonalities among hospitals' websites targeted with those data breach attacks towards their modeling and characterization. We believe that this work is the first step towards understanding website attributes that may lead to breaches and enable future research on vulnerability prediction and detection.

**Research Questions.** We aim to answer an overarching single question: **Is there any difference between the different categories of hospital websites with respect to the studied features across content, performance, and security?** We break the question down into the following quantifiable questions.

- **RQ1.** How different are hospitals with their use of domain, content, and transport layer features? We answer this question by comparatively exploring the domain-level features, including the domain name registrar, top-level domain distribution, domain creation distribution, and content-level features, including the content type and HTTP request features.
- **RQ2.** What are the main security characteristics of hospital websites, and how do they differ across types? We answer this research question by exploring the SSL certificate features and properties, maliciousness characteristics against engines, and data breach association.

**Contributions and Findings.** Given the lack of any systematic work on understanding the charac-

teristics of hospitals' presence on the web and their associated security and performance attributes, this study sets out to explore these hospitals' web presence across a range of attributes. Moreover, through a comparative analysis, this work uncovers the differences and similarities between the Government, Non-profit, and Proprietary hospitals in the United States. Our analysis is conducted across three dimensions: security, contents, and domains. To this end, our contributions are as follows:

1. **Domain-level Analysis.** Domain names are the gateway to websites, and they are essential to understanding various coarse-grained and easy-to-obtain features of those domains and entities behind them. To this end, we conduct a domain name registrar and top-level domain analysis to uncover websites/hospitals' characteristics and to contrast them. We uncover the affinities in the choice between websites and registrars, top-level domain choice, and domain creation dates. Among other interesting findings, we observe that the number of websites for government and non-profit hospitals has been declining in recent years, hinting at the aggressive proprietary healthcare system.

2. **Content-level Analysis.** We examined the contents of hospitals' websites for a deeper look into their utilized content types, size, and employed security features. Through this analysis, we found the gap in employing different content types, such as images and scripts, in those websites, which affects the various performance metrics, including loading times. More interestingly, and rather surprisingly, we found that 6% of proprietary hospitals use the Domain Name System Security Extension (DNSSEC), in comparison to less than 1% of the government and non-profit hospitals.

3. **SSL Certificate-level Analysis.** Certificates are essential for website authentication and to facilitate web content encryption at the transport layer, providing a secure application medium. We investigate the HTTPS protocol configurations, associated SSL features, and

the SSL certificate validity of hospitals' websites. We categorize websites based on certificate authority affinity, utilized algorithms, and certificate validity. Among other interesting findings, our investigation uncovers that more than 25.25% of hospital websites are still using the insecure HTTP protocol. Further, among websites that utilize HTTPS, up to 23% of the SSL certificates are invalid.

4. **Malicious Activities Analysis.** Because of their complex nature, unintended weaknesses might emerge due to the agglomeration of third-party code and the utilization of various shared pieces of infrastructure in hospital websites. To understand this dimension, we utilize various scanning tools to explore those websites' malicious activities at the domain and content levels. Among other interesting findings, we uncover that a large portion of websites contains malicious content and are associated with malicious behaviors.

5. **Data Breaches Analysis.** Data breaches are inevitable. But what (in the correlation sense) makes a website prone to data breaches? We explore this question by correlating and associating the hospitals to recently reported data breach incidents and uncover that non-profit hospitals are more likely to be involved in data breach incidents. We demonstrate the most important attributes contributing to data breach incidents, including hosting malicious codes, a large number of images, etc.

## Summary of Completed Work

This work analyzes the online presence of hospitals and investigates their websites in light of the increasing trend of attacks targeting hospital networks. We categorize hospitals into government, non-profit, and proprietary sectors, providing a comparative study that examines various structural and security features. We investigate the SSL certificate validity and related issues observed on

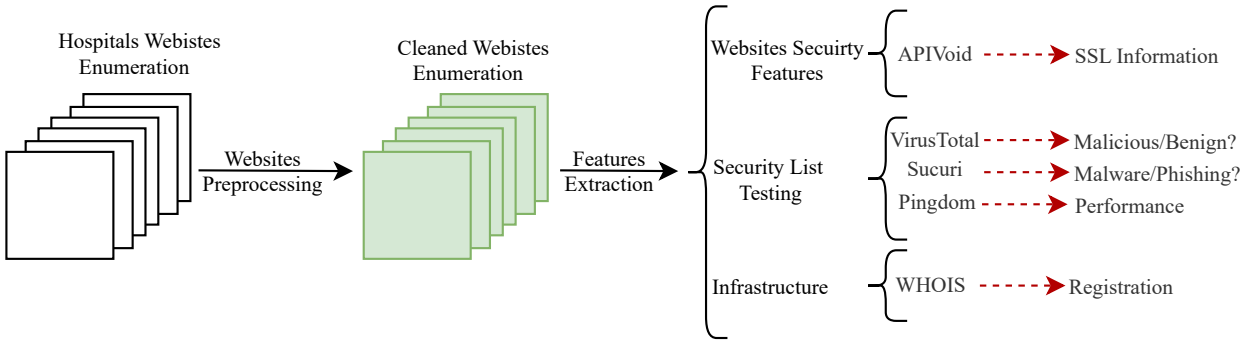



**Figure 5.1:** Our pipeline with the steps taken in website crawling and data augmentation against various dimensions: SSL, maliciousness, vulnerability, performance, and domain attributes.

hospitals' websites, as well as identify any malicious associated behaviors. Furthermore, we collect attributes and utilize them as features. The study aims to highlight the most important indicators of websites associated with data breach incidents and improve understanding of their security posture.

## Dataset, Pipeline and Research Questions

For this study, we utilized an authentic dataset of U.S. hospitals obtained from the Homeland Infrastructure Foundation-Level Data (HIFLD) [69]. The dataset contains hospitals distributed among the 50 U.S. states, Washington D.C., and U.S. territories of Puerto Rico, Guam, American Samoa, Northern Mariana Islands, Palau, and the Virgin Islands.

We categorized the hospitals in the dataset into three categories: government, non-profit, and proprietary. The government hospitals include federal, district, local, and state hospitals. Non-profit hospitals are those operated using charities according to the Internal Revenue Service (IRS) [41]. The proprietary hospitals are those owned and operated for profit by individuals, partnerships, or, in most cases, corporations [84]. Overall, we had 1,034 governmental hospitals, 2,187 non-profit hospitals, and 1,550 proprietary hospitals.

**Websites Preprocessing and Crawling.** Our study involved the use of website crawling to systematically gather information from websites and incorporate it for our further analysis. Figure 5.1 illustrates the overall process we followed, starting with website enumeration to identify the websites we needed to crawl. Next, we conducted website preprocessing, which involved removing irrelevant websites (i.e., websites with irrelevant contents to the scope of the study) and non-functioning websites. Finally, we performed feature extraction to extract attributes from the websites that we used for our further analysis.

To conduct our analysis, we introduced a data augmentation step. Upon crawling the websites associated with each hospital, we enriched the collected data with additional attributes, including the following.

- **SSL Attributes.** To extract SSL certificate information such as mismatched domains, SSL expiration dates, and certificate validity, we utilized APIVoid [16], a framework that offers cyber threat analysis and detection capabilities.
- **Maliciousness Attributes.** To analyze the maliciousness of hospital content, we employed VirusTotal API [92], an online service that aggregates results from over 70 scanning engines.
- **Vulnerability Attributes.** To examine each website for vulnerabilities and identify any malicious code, we utilized Sucuri [85], a service that tests websites against multiple known malware and blacklisting lists.
- **Performance Attributes.** To evaluate website performance and availability, we utilized Pingdom [74], a global monitoring software for websites.
- **Domain Attributes.** To determine ownership and DNSSEC information for each website, we utilized WHOIS [96], an Internet resource ownership database, and queried each website's creation date.

**Table 5.1:** The hospitals' website URLs correspond to domain registrar organizations. Notice that *Network Solutions* and *GoDaddy* are the most prominent in the list, with up to 67.65% associated URLs.

| Domain Registrar | Government | | Non-profit | | Proprietary | |
|---|---|---|---|---|---|---|
| | # | % | # | % | # | % |
| Network Solutions_LLC | 330 | 35.22 | 956 | 44.90 | 345 | 22.91 |
| GoDaddy.com_LLC | 314 | 33.51 | 621 | 29.17 | 527 | 34.99 |
| MarkMonitor_Inc | 1 | 0.11 | 19 | 0.89 | 156 | 10.36 |
| eNom_LLC | 28 | 2.99 | 51 | 2.40 | 159 | 10.56 |
| Register.com_Inc | 24 | 2.56 | 43 | 2.02 | 14 | 0.93 |
| NAMECHEAP_Inc | 13 | 1.39 | 37 | 1.74 | 24 | 1.59 |
| CSC CORPORATE_Inc | 2 | 0.21 | 51 | 2.40 | 19 | 1.26 |
| Tucows_Inc | 37 | 3.95 | 26 | 1.22 | 10 | 0.66 |
| Other | 188 | 20.06 | 325 | 15.27 | 252 | 16.73 |

**Table 5.2:** Top-Level Domain comparison between the Government, Non-profit, and Proprietary hospitals.

| Type | .org | .com | .gov | .net | .mil | .edu | .us |
|---|---|---|---|---|---|---|---|
| Government | 48.45% | 36.56% | 4.35% | 4.35% | 2.90% | 2.42% | 0.48% |
| Non-profit | 67.49% | 28.81% | 0.05% | 1.92% | 0.00% | 1.37% | 0.27% |
| Proprietary | 7.81% | 87.35% | 0.00% | 3.48% | 0.00% | 0.13% | 0.45% |

Overall, the steps of websites crawling and data augmentation allowed us to extract two types of information: ① website content data such as images, fonts, HTML, CSS, scripts, XHR, and redirects, and ② performance metrics such as page size, load time, and the number of requests.

### Websites Analysis

To understand the online presence and structural differences between hospitals of different categories, we conducted a range of analyses of their websites: domain-, content-, SSL certificate-, and malicious activities-based analyses. In the following, we review the findings from those analyses.

**Domain-level Analyses.** The domain name is a crucial asset for any organization, serving as a key element in their branding efforts and providing them with online presence and Search Engine Optimization (SEO) benefits. Therefore, in order to kickstart our website analysis, we begin by examining the domain name details, including the domain name registrar, the top-level domain, and the domain creation date.

**Domain Name Registrar.** The domain name registrar is an organization that manages the reservation of Internet domain names, as well as the assignment of IP addresses for those domain names [28], and certain registrars tend to be more lax with their security provisions and policies [30, 31, 89]. Analyzing the domain name registrar is crucial in evaluating a website's overall security and reliability. This is because the registrar provides important information about the organization's online presence and security measures. The level of security provisions and policies of the registrar can vary, which can impact the website's security and trustworthiness. Examining the domain name registrar can provide valuable insights into the organization's security approach and help assess potential risks associated with the domain name. In addition, understanding the registrar can shed light on the organization's online strategy and web hosting arrangements, which can further inform the analysis of the website's structure and performance. Table 5.1 shows the breakdown of domain registrar organizations by hospital type. Notably, *Network Solutions* and *GoDaddy* are the most prominent registrars, accounting for up to 67.65% of the domains. Additionally, we observe that although *Mark Monitor* and *eNom LLC* are relatively absent from government and non-profit websites, they contribute to 20.92% of proprietary hospital websites.

**Top Level Domain.** The Top-Level Domain (TLD) is the "extension" of a domain name. Besides branding, TLD plays an essential role in the Domain Name System (DNS) lookup and helps classify and communicate the purpose of domain names. Examples of TLDs include ".com," ".org," ".net," and ".edu." The TLD provides information about the website's purpose, organization type,

or geographic location. For instance, ".com" is commonly used for commercial websites, ".org" for non-profit organizations, ".edu" for educational institutions, and ".gov" for government websites. Understanding the TLD can provide insights into the website's intended audience and the type of content or services it provides. In addition, analyzing the TLD can help identify potential security risks associated with the website. For example, certain country-specific TLDs are known to be associated with malicious activities, and websites using such TLDs may be more likely to pose security threats. Moreover, some websites may use TLDs that are misspelled or similar to well-known TLDs in an attempt to deceive visitors and carry out fraudulent activities. By analyzing the TLD, we can identify such risks and take appropriate measures to mitigate them. Therefore, analyzing the TLD is a crucial step in evaluating a website's overall security and reliability. The Internet Corporation for Assigned Names and Numbers (ICANN) is responsible for the authority over all TLDs on the Internet and delegates these TLDs' responsibility to various organizations [27] Table 5.2 shows the TLD comparison between the hospital categories in our dataset. We observe that *".org"* is the most dominant TLD for the government (48.45%) and non-profit (67.49%), while it is relatively absent in the proprietary hospitals (only 7.81%). On the other hand, *".com"* is dominant for proprietary hospitals (87.35%) compared to (36.56% and 28.81%) in government and non-profit hospitals, respectively. We also notice that 92.15% of the hospitals' websites, in the aggregate, have *".com"* or *".org"*. Despite our common beliefs, we surprisingly uncover that only 4.35% of the government websites use the *".gov"* TLD.

**Domain Creation.** The domain creation date refers to the date on which a specific domain name was initially registered with a domain name registrar. It is a crucial piece of information because it can provide insights into the website's history and online presence. A website that has been registered for a longer duration may be more established and have a more significant online presence than a newer website. Analyzing the domain creation date can be useful in identifying potentially fraudulent or malicious websites. For instance, a website that has been recently registered may

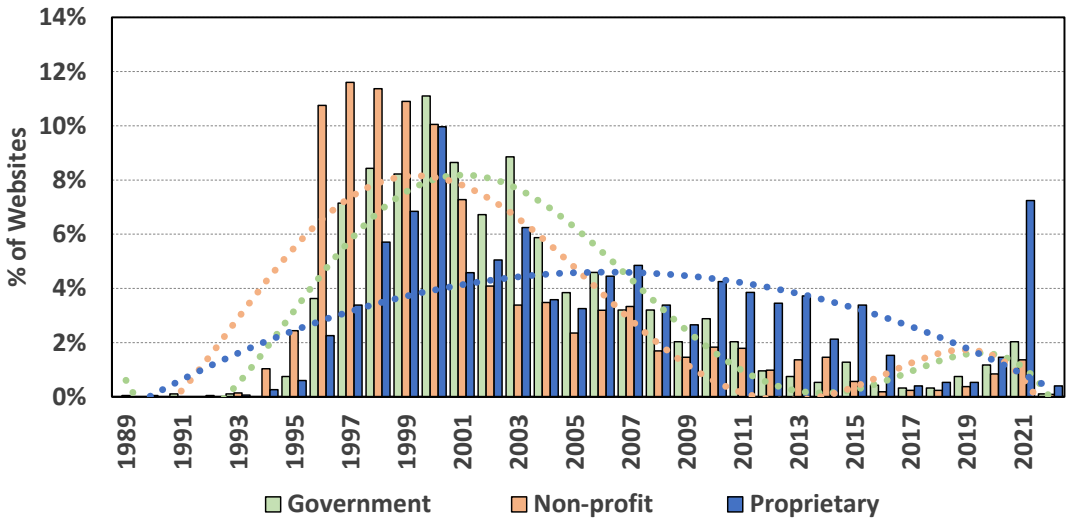



**Figure 5.2:** The domain creation date temporal analysis between the three hospital categories. Dot lines are moving averages.

**Table 5.3:** Content-type comparison between the Government, Non-profit, and Proprietary hospitals.

| Category | CSS | Font | HTML | Image | Redirect | Script | XHR |
|---|---|---|---|---|---|---|---|
| Government | 6.05% | 8.18% | 2.71% | 40.55% | 17.37% | 20.34% | 4.79% |
| Non-profit | 5.30% | 7.80% | 2.86% | 38.01% | 16.11% | 27.01% | 2.91% |
| Proprietary | 6.63% | 9.61% | 3.53% | 28.05% | 13.76% | 33.12% | 5.29% |

be more likely to be a part of a phishing scam or a fraudulent scheme. Therefore, analyzing the domain creation date can be a valuable step in evaluating the website's overall credibility and potential security risks. Figure 5.2 shows the domain creation date of hospitals in different categories. As shown, both government and non-profit hospitals' websites emerged in a similar period (1995 – 2009), with a declining trend after 2009. However, the emergence of proprietary hospital websites is steady, with a rapid increase in their numbers in 2021.

**Takeaway:** (**RQ1.**) While the number of websites for government and non-profit hospitals has been declining in recent years, proprietary hospitals have been growing significantly, particularly in 2021. Moreover, despite being government-supported, most government hospitals do not have *".org"* top-level domain.

**Content-level Analyses.** To analyze the content differences between the different categories of hospitals, we crawled the hospitals' websites using Pingdom [74], obtaining the HTTP request information and all associated files; scripts, images, HTML, and CSS files.

**Content Type.** On the structural level, Table 5.3 shows the distribution of the file type among the three hospital categories. XHR is an API used as an object to interact with servers and exchange data between servers and web browsers. Containing *"XHR"* is prominent among the government and proprietary hospital websites, with 10.08% combined. Overall, the file type distribution is similar for all categories, except for *"Image"* and *"Script"*. The *"Script"* content, which is defined as a computer program for adding dynamic capabilities to a website, is used most among the proprietary hospitals with 33.12%. The *"Redirect"* content, on the other hand, which is a website feature that sends a user from the current URL to another server, is applied more in the government (17.37%) and non-profit (16.11%) and the least in the proprietary category (13.76%).

**HTTP Request.** The HTTP (Hypertext Transfer Protocol) request is a message that is sent from a client (such as a web browser) to a server, requesting a particular resource or action. The request typically includes a URL (Uniform Resource Locator) that specifies the resource or action being requested, along with any additional information needed by the server to fulfill the request, such as headers and cookies.

By Analyzing HTTP requests, businesses and organizations can gain insights into the performance of their websites, including, for example, the speed of page load times, the number of requests per

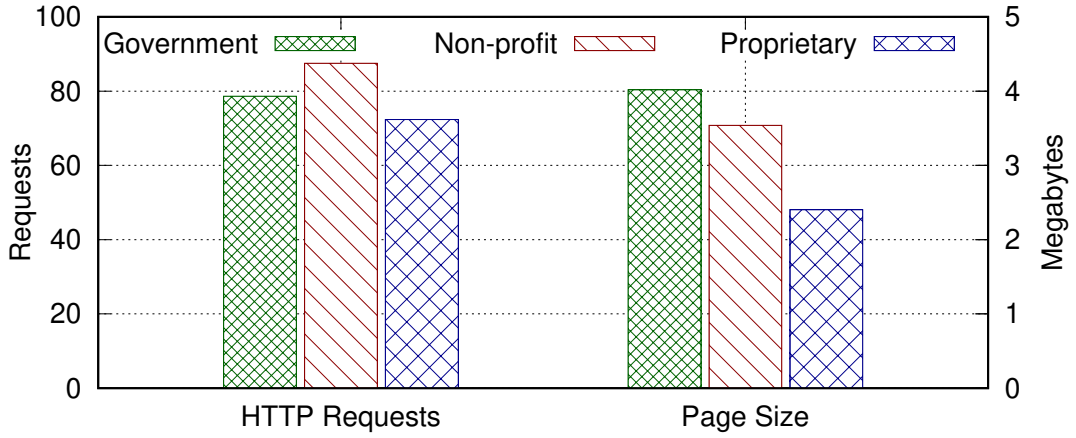



**Figure 5.3:** Request and response size comparison.

page, and the size of files being requested. This information can help identify areas where website performance can be further optimized.

Figure 5.3 shows the average HTTP requests per website across the three different categories of analyzed hospitals. We found that most websites generated 65 to 90 HTTP requests per visit, with the non-profit hospitals being the highest. Despite having relatively similar HTTP requests, the proprietary hospitals' average page size was 45% less than the government hospitals. Upon further analysis, we found that the proprietary hospitals' websites contain the least percentage of images in contrast to the government and non-profit hospitals, which help explain this trend.

> **Takeaway:** (**RQ1.**) Structurally, the content type distribution of the hospitals' categories are similar, except for *"Image"* and *"Script"* content types. Although the HTTP requests were relatively similar among hospitals, the average page size of the proprietary hospitals was 45% smaller than that of the government hospital.

**DNSSEC Prevalence.** DNSSEC (Domain Name System Security Extensions) is a security protocol that is used to protect against certain types of attacks on the Domain Name System (DNS),

**Table 5.4:** The corresponding certificate issuer organizations for the hospitals' websites. Notice that *Let's Encrypt* is the most prominent certificate issuer organization, with up to 25.21% associated URLs.

| Issuer Organization | Paid | Government | | Non-profit | | proprietary | |
|---|---|---|---|---|---|---|---|
| | | # | % | # | % | # | % |
| Let's Encrypt | ✗ | 321 | 31.04 | 559 | 25.56 | 295 | 19.03 |
| GoDaddy.com_Inc. | ✓ | 131 | 12.66 | 175 | 8.00 | 132 | 8.51 |
| Cloudflare_Inc. | ✗ | 25 | 2.41 | 108 | 4.93 | 179 | 11.54 |
| Sectigo Limited | ✓ | 67 | 6.47 | 133 | 6.08 | 38 | 2.45 |
| cPanel_Inc. | ✗ | 47 | 4.54 | 72 | 3.29 | 111 | 7.16 |
| DigiCert Inc. | ✓ | 35 | 3.38 | 154 | 7.04 | 14 | 0.90 |
| Trustwave Holdings, Inc. | ✓ | 1 | 0.09 | 9 | 0.41 | 148 | 9.54 |
| Entrust L1K | ✓ | 9 | 0.87 | 86 | 3.93 | 43 | 2.77 |
| Other | - | 148 | 14.31 | 385 | 17.60 | 149 | 9.61 |
| No SSL Certificate Found | - | 250 | 24.17 | 506 | 23.13 | 441 | 28.45 |

targeting the integrity and source authenticity.

DNSSEC works by adding digital signatures to the DNS records, which DNS resolvers can verify to ensure that the records have not been tampered with or forged. This helps to prevent attackers from redirecting users to fake websites, intercepting email messages, or other types of attacks that rely on manipulating DNS records. DNSSEC is designed to provide end-to-end security, meaning that the integrity of the DNS records can be verified from the root servers down to the individual domain name servers. It is supported by most modern web browsers and operating systems, and many top-level domains have already adopted DNSSEC to provide an additional layer of security for their users.

To gauge its prevalence in this critical category of websites, we investigated the DNSSEC deployment and found that 6% of the proprietary hospital websites enabled DNSSEC, as shown in Figure 5.4, which is higher than government (0.55%) and non-profit (0.75%) hospital websites.

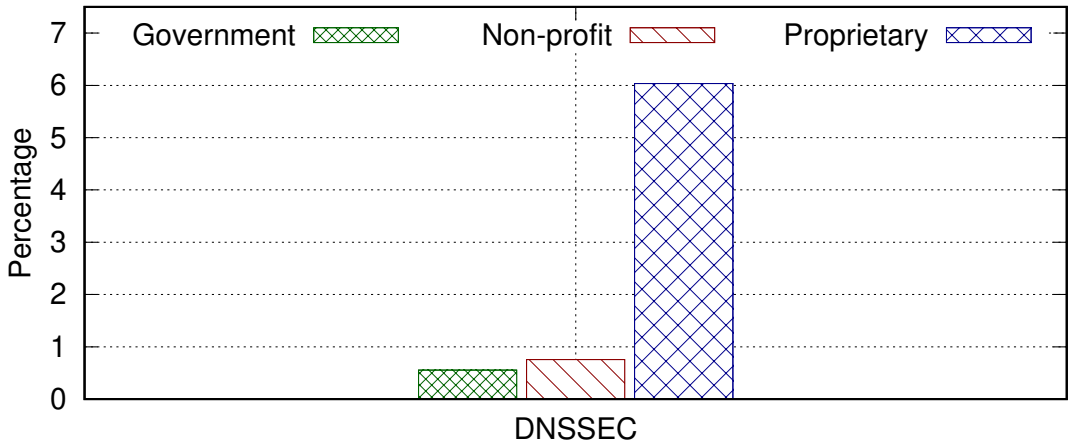



**Figure 5.4:** Domain Name System Security Extensions.

**Takeaway:** (**RQ2.**) We notice that 6% of proprietary hospitals used DNSSEC compared to under 1% in government and non-profit. While 1% is considered a small percentage, it is consistent with the DNSSEC deployment in `.com` and `.org` general websites (i.e., 0.75%–1%) [25], whereas the proprietary group has 6 to 8 times more than those levels, highlight better prevalence in this category.

**HTTPS and SSL Certificate Analysis.** The HTTP protocol is responsible for transferring website content from the web server to the endpoint browser. However, this protocol is insecure, exposing content to unauthorized access. Most websites have moved to use HTTPS, a secure version of HTTP, on top of the Secure Sockets Layer (SSL), which, among other functions, implements an encryption mechanism to protect the transferred content between web servers and endpoint browsers. Healthcare websites require users to enter sensitive personal information such as health-related data, insurance information, and medical history. HTTPS can help protect this information from being intercepted by attackers, thereby ensuring that patient data remains confidential.

To this end, we next look into SSL-related configurations: certificate authority, signature algorithm, and certificate validation. Among the studied hospitals, we noticed that 25.25% of the websites

**Table 5.5:** SSL signature algorithms' comparison.

| | Government | | Non-profit | | proprietary | |
|---|---|---|---|---|---|---|
| Algorithms | # | % | # | % | # | % |
| SHA256 with RSA | 751 | 95.00 | 1,535 | 91.31 | 891 | 80.34 |
| SHA256 with ECDSA | 26 | 3.30 | 108 | 6.42 | 179 | 16.14 |
| SHA384 with ECDSA | 10 | 1.27 | 35 | 2.08 | 37 | 3.34 |
| SHA1 with RSA | 1 | 0.13 | 2 | 0.12 | 1 | 0.09 |
| SHA384 with RSA | 1 | 0.13 | - | - | 1 | 0.09 |
| SHA512 with RSA | - | - | 1 | 0.06 | - | - |

are still using HTTP, in contrast to only around 20% in general web [93]. While there is a 5% of difference in the number of websites, that number is alarmingly high given the type of data associated with hospitals (i.e., at least one out of four hospitals uses an insecure protocol).

**Certificate Authority.** A certificate authority (CA) is an organization that validates the identities of entities, including websites, email addresses, etc., by binding entities to cryptographic keys through the issuance of electronic documents. Investigating the certificate authority organization (i.e., the issuer of the certificate), Table 5.4 shows that the majority of hospitals are using the free *Let's Encrypt* services [47] (i.e., free SSL certificates)., with up to 31.04% for the governmental hospital group. We also notice that hospitals' websites widely use free SSL certificates. Surprisingly, we did not find SSL certificates in 24.17% of government, 23.13% of non-profit, and 28.45% of proprietary websites.

**Algorithms.** Table 5.5 shows the SSL signature algorithms used by the government, proprietary, and non-profit websites. As shown, *SHA256 with RSA* is the most used scheme with 95.00% for government, 91.31% for non-profit, and 80.34% for proprietary hospital websites, respectively. This is mainly because hospitals intend to use traditional go-to algorithms adopted by service providers. On the other hand, we notice that fewer hospitals website use *SHA256 with ECDSA*

(Elliptic Curve Digital Signature Algorithm) algorithm that uses shorter keys for the same security level as in RSA with larger keys [3]. With 3.30% for government, 6.42% for non-profit, and 16.14% for proprietary hospital websites, respectively. We note ECDSA is a newer and more efficient algorithm and is mainly used in newer websites [3]. ECDSA is, however, more vulnerable to attacks than the older RSA under post-quantum adversaries, according to recent studies [77].

**Certificate Validity.** We further investigated the SSL certificate validity and potential issues. In the following, we discuss issues related to SSL certificate failures (see Figure 5.5).

*SSL Mismatched Domain.* A mismatched domain might be an indication of website impersonation or inconsistent website migration, and both highlight a lack of rigorous security practices. We found that 18.45% of the proprietary hospitals had SSL certificates with mismatched domains, versus 14.05% of the government hospitals and 17.67% of the non-profit hospitals. Even varying, all hospitals' websites had concerning ratios of mismatched domains.

*SSL Expired.* Our analysis uncovered that about 3.97% of non-profit hospitals have expired certificates, compared to 2.20% and 2.09% for government and proprietary hospitals. Similar to our previous analysis of out-of-date websites, this may lead to potential user information and data privacy risks.

*SSL Invalid.* The invalidity of SSL means that some fields in the certificate are incorrect. Surprisingly, all hospital categories had an alarming percentage of invalid SSL certificates, with 15.94% for government, 21.70% for non-profit, and 23.72% for proprietary hospital websites.

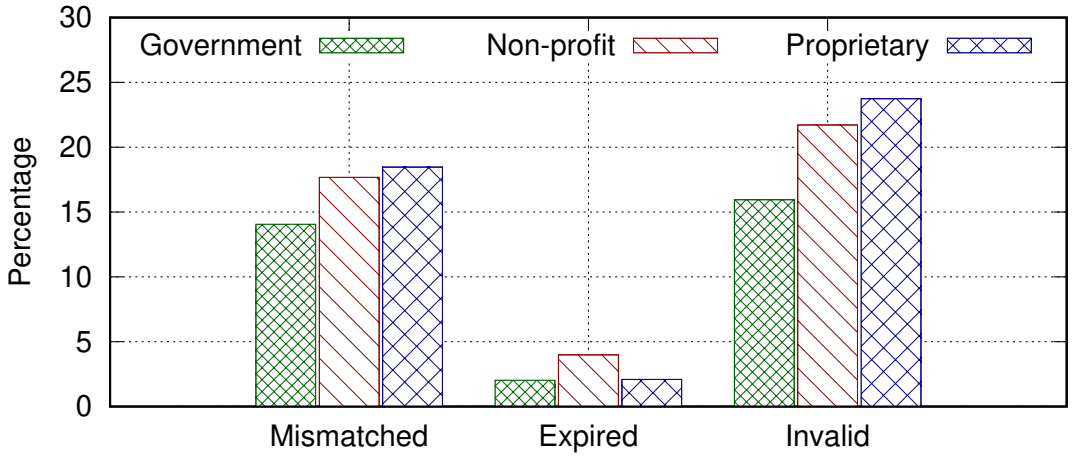



**Figure 5.5:** The SSL validity comparison of Government, Non-profit, and proprietary hospital websites.

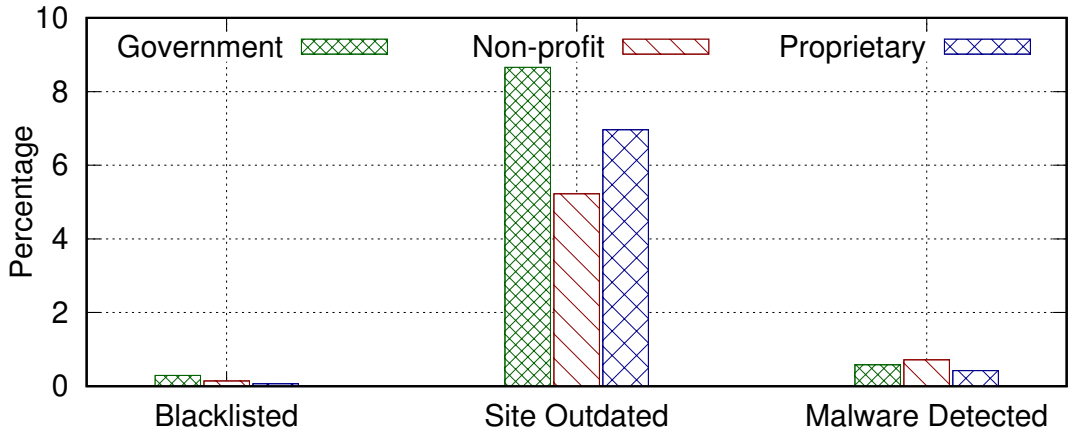



**Figure 5.6:** Comparing the maliciousness of Government, Non-profit, and proprietary hospitals' websites.

**Takeaway:** (**RQ2.**) More than 25% of hospitals' websites are using the plain HTTP protocol, which is alarmingly higher than ≈20% in the general websites [93]. Among websites that used HTTPS, 88.77% of them used *SHA256 with RSA*. Among the ≈75% hospitals with an SSL certificate, we found that 20.45% of the SSL certificates were invalid while 16.72% had a mismatched domain name, primarily in proprietary hospitals in both cases.

**Malicious Activities Analysis.** In addition to the structural differences and SSL certificate analyses, we study the malicious activities associated with the hospitals' websites. Malicious activities in this work include providing malicious or phishing content or the association of website resources with malicious attacks.

**Domain-based Malicious Activities.** We leveraged Sucuri [85] to explore domain-based malicious activities. Figure 5.6 shows that although only a small portion of hospitals' URLs are blacklisted or labeled as malware, about 8.66% of government, 5.21% of non-profit, and 6.96% of proprietary hospital websites are outdated, which raises concerns of data leakage.

**Content-based Malicious Activities.** Next, we analyzed the website content using VirusTotal API [92]. Figure 5.7 shows that, among the three hospital categories, 84.21% of proprietary hospitals contained malware, compared to only 13.15% and 7.89% in the government and non-profit hospitals, respectively. Moreover, we observed that 65.21% of proprietary hospital websites are suspected of having phishing-like behaviors. We note that the percentage for the government (30.43%) and non-profit (26.08%) hospitals are significantly smaller than that of the proprietary hospitals but still noticeably high. Besides, we observed that 56.25% of government and proprietary hospitals and 18.75% of non-profit hospitals are associated with malicious activities.

**Takeaway:** (**RQ2.**) Most hospitals have malicious features (domain or content), and many are vulnerable to data leakage due to a lack of maintenance (i.e., outdated websites). Among the hospital websites, 8.66% of the websites are outdated. In addition, a concerning portion of websites contain malicious content and are associated with phishing and malicious behaviors.

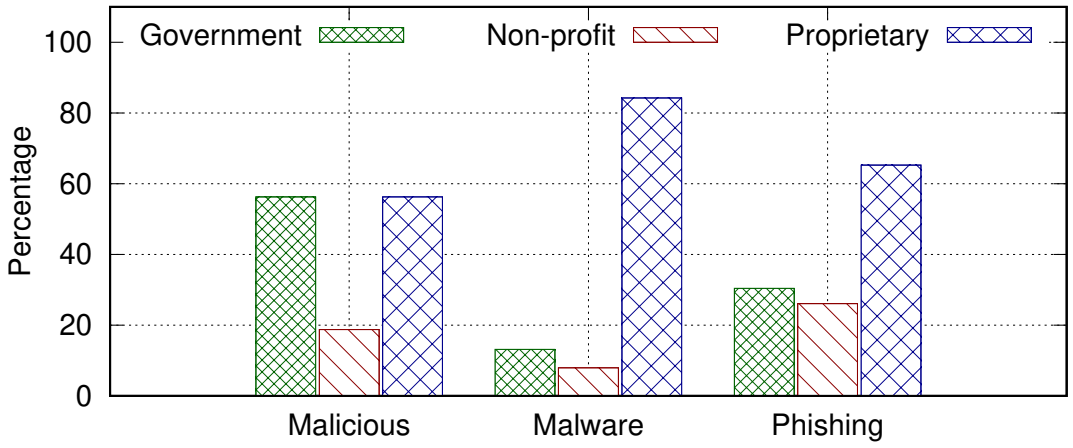



**Figure 5.7:** The potential maliciousness of Government, Non-profit, and proprietary hospitals.

Data Breaches Analysis

Analyzing the data breaches helps understand the correlation between web presence security and incidents. According to the Health Insurance Portability and Accountability Act (HIPPA), a data breach can be defined as an impermissible use or disclosure under the Privacy Rule that compromises the security or privacy of the protected health information [35]. In the healthcare domain, data breaches are devastating, as they cause damage to patients and healthcare organizations alike. Recent works have shown that healthcare is the most targeted industry by cyber criminals due to financial gain as attackers intend to sell patients' records on the dark web. To investigate historical data breach incidents in hospitals, we obtained the healthcare data breaches dataset from the U.S. Department of Health and Human Services, Office for Civil Rights (OCR). The OCR portal lists all data breaches of unsecured health information affecting 500 or more patients. The OCR portal categorizes data breaches into two categories; (i) incidents reported within the last 24 months and currently under investigation and (ii) the achieved breaches, which comprise the resolved breach reports older than 24 months. We note that it is challenging to associate the hospital names with the entities named in the data breaches, as they are not consistently organized (e.g., mixture or trunca-

tion). To resolve the issue, we started by using the hospitals' names as anchors and then leveraged Natural Language Toolkit (NLTK) [66] for punctuation removal, case normalization, stopwords removal, and lemmatization & stemming of the hospitals' names. A similar process was followed for the entity name among the data breaches dataset. Lastly, any hospital name and entity with two common words are filtered for manual analysis and vetting. Overall, we manually inspected 1,253 incidents, resulting in 414 accurate labeling of data breaches.

Hereafter, we analyze the data breaches, providing insights into the common online attributes of the breached hospitals.

**Associated Hospitals & Individuals.** Among the 414 data breach incidents in our dataset, 49 were government hospital-related, 156 were non-profit hospital-related, and 34 were proprietary hospital-related. It is worth mentioning that a hospital may be involved in several incidents. Our analysis indicates that the average number of affected individuals is 58,750 overall, including 60,458 for government, 64,977 for non-profit, and 50,815 for proprietary hospitals. Remarkably, the proprietary hospitals are involved in the least number of incidents and affected individuals.

**Data Breach Surface.** As shown in Figure 5.8, *"paper/films"* are the most commonly targeted for government (32%) and proprietary (21%) hospitals, despite only 6% for non-profit hospitals. Then, *"emails"* are mostly targeted in non-profit hospitals (34%), despite not being heavily targeted in government and proprietary hospitals (1%). Overall, *"network server"* is the second most common target after *"paper/films"*, inferring the importance of hospitals' online security.

**Data Breach Type.** As shown in Figure 5.9, we found the majority of incidents are *"hacking/IT"* representing 45.75% in the government hospitals, followed by proprietary and non-profit hospitals representing 30.22% and 22.58%, respectively. Further, we observed that *"unauthorized access/disclosure"* is the most common data breach type within the non-profit hospitals representing

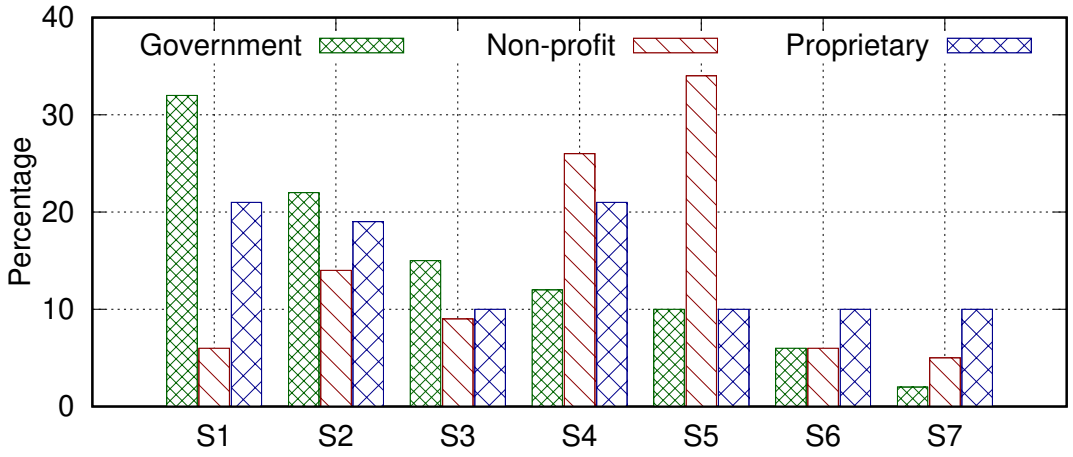



**Figure 5.8:** Comparing the data breach surfaces of Government, Non-profit, and proprietary hospitals. S1= Paper/Films, S2=Network Server, S3=EMR, S4=Other, S5=Email, S6=Laptop, S7=Desktop Computer.

50%, while 37.91% for proprietary and 32.98% for government hospitals.

**Data Breach Online Presence Attribution.** To understand the relationship between online presence, security properties, and data breach incidents, we used a gradient boosting model with non-negativity constraint (monotonously constraint) to learn important attributes of breach incidents.

Table 5.6 illustrates the 21 attributes used in our model, and Figure 5.10 shows the ten attributes that are directly (and mostly) correlated with the breached websites. As shown in the figure, when the website contains malware software (F20: Websites detected as malware by Sucuri API), it is (naturally) more likely to be involved in a data breach incident. Other features that highly correlated with websites' data breaches are (F15: The percentage of images retrieved by Pingdom API, and F13: The percentage of font retrieved by Pingdom API).

**Takeaway:** (**RQ2.**) Non-profit hospitals, with 156 reported breaches, are more vulnerable than other types. Key security attributes of hospital websites significantly influence their breach risk, with proprietary hospitals showing the least susceptibility.

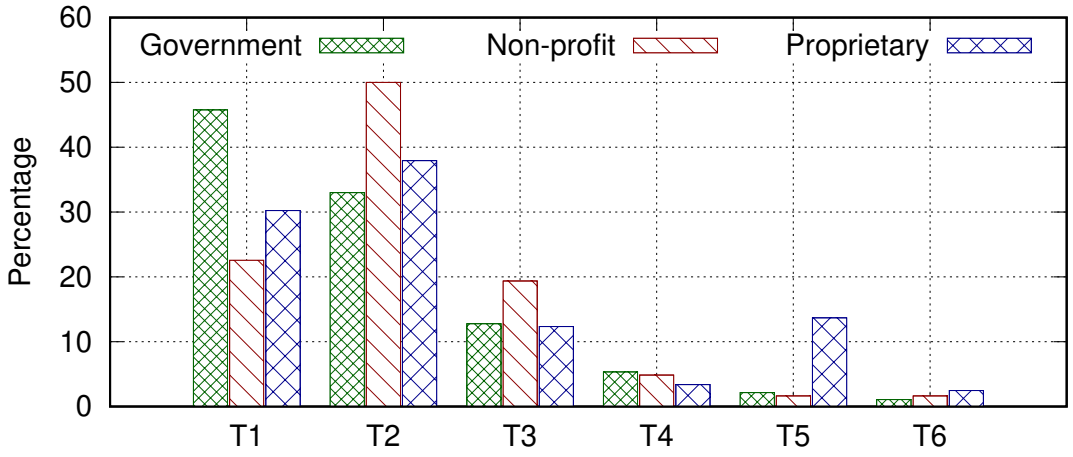



**Figure 5.9:** Comparing the data breach types of Government, Non-profit, and proprietary hospitals. T1=Hacking/IT, T2=Unauthorized Access, T3=Theft, T4=Loss, T5=Improper Disposal, T6=Other.

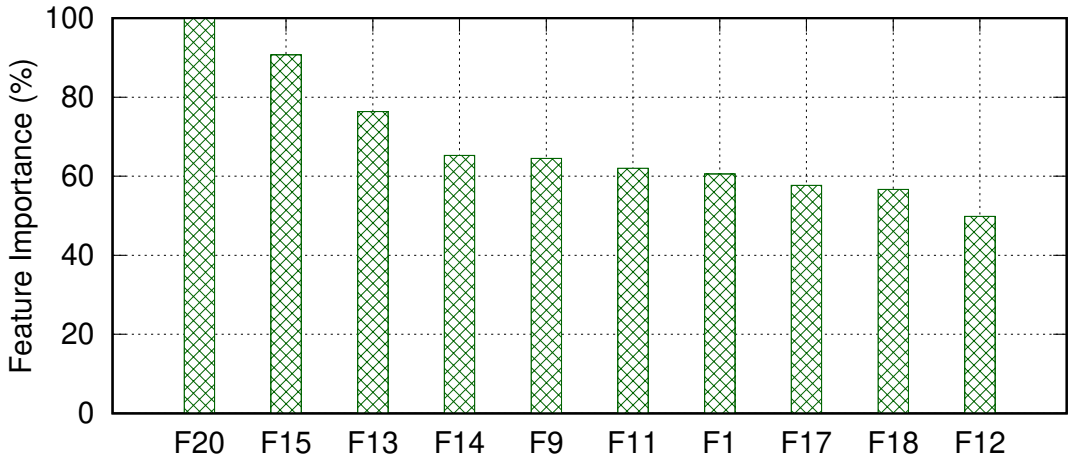



**Figure 5.10:** The domain- and content-level attributes importance (%) in distinguishing hospital websites associated with data breaches. The titles and descriptions of all features are shown in Table 5.6

## Summary and Concluding Remark

Recent reports showed an increasing trend of attacks targeting hospital networks to compromise sensitive patient data. In this paper, we investigated the online presence of hospitals by analyzing

**Table 5.6:** Attributes extracted for data breach analysis.

| | Title | Description |
|---|---|---|
| F1 | Certificate Invalid | The browser fails to verify website certificate |
| F2 | Certificate Unmatched | The website name does not match SSL certificate |
| F3 | Certificate_Expired | The website certificate becomes invalid |
| F4 | Validity_Days_Left | The remaining validity days of website certificate |
| F5 | Positives | The website domain is detected by VirusTotal API |
| F6 | Malicious_Site | Websites detected as malicious by VirusTotal API |
| F7 | Malware_Site | Websites detected as by malware VirusTotal API |
| F8 | Phishing_Site | Websites detected as by VirusTotal API phishing |
| F9 | Page_Size (MB) | The website average page size of in MegaByte |
| F10 | Load_Time (S) | The website average page load time of in seconds |
| F11 | Number of Requests | The website average number of requests |
| F12 | CSS | The percentage of CSS retrieved by Pingdom API |
| F13 | Font | The percentage of font retrieved by Pingdom API |
| F14 | HTML | The percentage of HTML retrieved by Pingdom API |
| F15 | Image | The percentage of images retrieved by Pingdom API |
| F16 | Redirect | The percentage of redirect retrieved by Pingdom API |
| F17 | Script | The percentage of script retrieved by Pingdom API |
| F18 | XHR | The percentage of XHR retrieved by Pingdom API |
| F19 | Blacklisted Flag | Websites detected as blacklisted by Sucuri API |
| F20 | Malware Flag | Websites detected as malware by Sucuri API |
| F21 | DNSSec Flag | Websites detected using DNSSec |

their websites. Benefiting from a categorization into government, non-profit, and proprietary hospitals, we conduct a comparative study that sheds light on various structural and security features. Of particular note, we investigated the SSL certificate validity, the related issues among hospitals' websites, and malicious associated behaviors. Leveraging the collected attributes as features, we demonstrate the most important attributes indicative of websites associated with data breach incidents and helpful in understanding their security. Our findings are among the first steps toward achieving patient security, alarmingly highlighting the lax security in many hospitals' websites.

# CHAPTER 6: CROSS-INDUSTRY ANALYSIS OF DATA BREACHES

Data is considered the core assets of organizations and institutions and serves as their vital resource. Data are crucial for different industries for many reasons, including decision-making, improving the efficiency and productivity, product personalizations, and many others. However, cybercriminals constantly target data due to the value and benefits that can be gained from such data. Thus, data breaches stand out as a dark shadow over various sectors, such as *Healthcare*, *Finance*, *Education*, *Transportation*, and many more. The landscape of data breaches is expansive and ever-changing, posing a complex challenge that affects all industries. As organizations rely more on digital technologies to manage sensitive data, the threat of data breaches becomes a significant and urgent issue. Data breaches have afflicted prominent organizations across industries. The security of the file transfer tool *MOVEit* was compromised, affecting over 200 organizations and exposing up to 17.5 million records in July 2023 [73]. In May 2023, it came to light that "T-Mobile" had experienced its second data breach within the same year. These incidents led to the disclosure of personal identification numbers (PINs), full names, and phone numbers of more than 800 customers [86]. The parent company of well-known fast-food chains, including KFC, Pizza Hut, and Taco Bell, *YUM Brands*, fell victim to a ransomware attack in April 2023 that resulted in the compromise of personally identifiable information (PII) [79].

A significant public debate arose surrounding "ChatGPT" due to a bug within an open-source library. This bug allowed certain users to access titles from the chat history of another active user, affecting approximately 1.2% of ChatGPT Plus subscribers who were active during a specific nine-hour window [72]. Finally, the popular fast-food joint, *Chick-fil-A*, experienced a data breach through its mobile app, exposing customers' personal Information [43]. Data breaches incur significant damage to both organizations and their customers. Organizations often find themselves burdened with substantial financial costs to mitigate the fallout from the compromised data, and

they invest years of effort to restore their damaged reputation. According to a report by IBM [44], the average total cost of a data breach in 2023 stands at $4.45 million, reflecting a 2.3% increase from the 2022 cost of $4.35 million. In response to these breaches, a notable 51% of organizations affected are directing their efforts toward strengthening security measures. This includes investments in incident response (IR) planning and testing, employee training, and adopting advanced threat detection and response technologies [44]. Concerns have arisen regarding the impact of data breaches on service quality, particularly in sectors such as *Healthcare*, where the quality of patient care is of utmost importance. It has been a common assumption that the quality of patient care tends to decline after a data breach. However, a study conducted by Choi *et al.* [55] challenges this prevailing belief, indicating that the quality of care in hospitals was not adversely affected following data breaches. Furthermore, the study suggests that hospital productivity remained resilient despite the disruptions caused by data breaches.

This paper delves into a detailed analysis of data breaches across twelve distinct sectors, shedding light on the various facets that contribute to the vulnerability of these sectors and the strategies employed by malicious actors to compromise data security. The sectors under examination encompass *Healthcare*, *Administration*, *Finance*, *Information*, *Education*, *Professional Services*, *Retail*, *Manufacturing*, *Other Services*, *Administrative*, *Accommodation*, and *Transportation*.

Our study explores a diverse set of attributes associated with data breaches, ranging from the confirmation of incidents to the motivations behind them. These attributes include action misuse, action misuse vectors, the classification of incidents as targeted or opportunistic, the types of victims affected, the size of organizations in terms of the number of employees, the internal and external motives driving data breaches, action errors, hacking incidents, malware attacks, physical breaches, and the types and vectors of these attacks. Furthermore, we consider the geographical dimension, examining the country of origin for these incidents and the yearly distribution trends.

**Contributions and Findings.** Given the lack of measurements analysis on understanding the characteristics of data breaches across different industries, this study explores data breaches across various attributes. Moreover, this work uncovers the differences and similarities between 12 primary known industries through a comparative analysis. To this end, our contributions are as follows:

1. **Breaches Classification.** This section outlines the approach to categorizing data breaches across industries using the North American Industry Classification System (NAICS). The classification process involved manually assigning data breaches to their respective industries based on NAICS guidelines.

2. **Preliminary Analysis.** Offers an initial examination of data breaches, focusing on their distribution over time and across various locations. This analysis also considers the size of affected organizations and distinguishes between targeted and opportunistic breaches.

3. **Threat Actions Analysis.** This section investigates the range of threat actions involved in data breaches, such as misuse, error, physical threats, hacking, social tactics, and malware. This comprehensive review aims to uncover the attack methods used against different industries and evaluate their impact.

4. **Discovery Methods.** Explores the techniques used to detect data breaches. The study assesses the success of existing detection methods and suggests improvements for more efficient discovery of breaches.

5. **Actor Motives.** This section delves into the motivations behind data breaches, differentiating between the objectives of internal and external actors. The analysis seeks to uncover the root causes of breaches across various sectors to inform more effective prevention strategies.

6. **Types of Victims.** Identifies the different victim profiles affected by data breaches in each

industry. By focusing on the victims, the study aims to highlight the broader impact of breaches and suggest protective measures for those most at risk.

## Dataset

To conduct our analysis, we utilized an authentic dataset known as the Vocabulary for Event Recording and Incident Sharing (VERIS) [91]. VERIS has a set of metrics designed to provide a common language for describing security incidents. We chose this dataset for our analysis for the following reasons: (i) VERIS offers a standardized framework for describing security incidents. This structure provides a common language that facilitates precise and repeatable reporting of incidents and (ii) the ability to categorize security incidents based on specific industries. This allows us to gain insights into security incidents tailored to particular sectors. VERIS responds to one of the most critical and persistent challenges in the security industry: the lack of quality information reporting for security incidents. To address this issue, VERIS aims to assist organizations by collecting and sharing valuable information related to incidents. Furthermore, it enables organizations to share this information anonymously and responsibly with others in the security community. The main goal of VERIS is to enhance risk measurement and management by providing a solid foundation that encourages constructive and cooperative learning from past experiences.

**VERIS Overview.** Over the past few years, the security industry has faced challenges in effectively gathering and sharing security incident data. There has been widespread participation; however, success in data sharing has been challenging to achieve due to the absence of a universally accepted taxonomy for incident reporting. The efforts to collect and share security incident data often encounter obstacles such as ambiguity regarding what to measure and incompatible classification systems that hinder progress, resulting in the lack of clarity and consistency, limiting the value of the collected data.

**The 4 A's Framwork.** In efforts to mimic the obstacles mentioned above, VERIS introduces the concept of the "4 A's," which are high-level categories for analyzing security incidents. The 4 A's represent who is behind an incident (Actor), methods used (Action), devices affected (Asset), and how they are affected (Attribute). VERIS uses numbers at intersections to reference specific combinations within these categories. The information risk can be described as the intersection between four intersecting landscapes, which are Threat, Asset, Impact, and Control. VERIS's efforts to provide effective risk management require information from each of these landscapes. To achieve this, VERIS leverages first-hand investigations and post-incident analysis as a primary data source. It focuses on providing metrics that are valuable for risk management. VERIS offers a unique strength by being rooted in real-world evidence and practical incident insights.

**NAICS SYSTEM.** VERIS utilizes the North American Industry Classification System (NAICS) to identify industries. NAICS is a framework used mainly to classify businesses into specific sectors based on their primary economic activity [20]. In the mid-1960s, NAICS was developed to replace the Standard Industrial Classification (SIC) system, aiming to provide a more modern and comprehensive approach to industry classification. NAICS framework was developed jointly by the statistical agencies of Canada, Mexico, and the United States to ensure a consistent and meaningful system for economic analysis across North America. This collaboration was necessary considering the interconnectedness of the three economies, especially in light of trade agreements such as NAFTA (North American Free Trade Agreement).

The NAICS framework offers numerous advantages. Firstly, it provides a tool for governments and research institutions to generate data tailored to specific industries, facilitating comparisons, insights into economic transitions, and the formulation of policies. Secondly, many governmental rules and guidelines utilize NAICS codes to determine their relevance. Lastly, companies employ NAICS codes to discern potential market opportunities, evaluate competitors, and monitor

**Table 6.1:** A list of industry codes and their corresponding definitions based on the North American Industry Classification System (NAICS).

| Sector | Definition |
|---|---|
| 11 | Agriculture, Forestry, Fishing, and Hunting |
| 21 | Mining, Quarrying, and Oil and Gas Extraction |
| 22 | Utilities |
| 23 | Construction |
| 31-33 | Manufacturing |
| 42 | Wholesale Trade |
| 44-45 | Retail Trade |
| 48-49 | Transportation and Warehousing |
| 51 | Information |
| 52 | Finance and Insurance |
| 53 | Real Estate and Rental and Leasing |
| 54 | Professional, Scientific, and Technical Services |
| 55 | Management of Companies and Enterprises |
| 56 | Admin, Support, Waste Management & Remediation Services |
| 61 | Educational Services |
| 62 | Health Care and Social Assistance |
| 71 | Arts, Entertainment, and Recreation |
| 72 | Accommodation and Food Services |
| 81 | Other Services (except Public Administration) |
| 92 | Public Administration |

economic patterns. Table 6.1 shows the breakdown of industries and the unique code assigned to each sector. Using the industry codes, we manually labeled data breach incidents by specific industries, resulting in 12 primary sectors: *Healthcare*, *Administration*, *Finance*, *Information*, *Education*, *Professional*, *Retail*, *Manufacturing*, *Other Services*, *Administration*, *Accommodation*, and *Transportation*.

## Dataset Breakdown

Table 6.3 provides a comprehensive breakdown of data breaches across various industries, classified by their NAICS code. For each sector, we display the total number of breaches, the count of near misses, suspected breaches, and confirmed breaches. Among the industries, we observed that the *Healthcare* and *Administration* industries have the highest data breaches, with 2,445 and 2,330, respectively. Thus, these numbers are significantly higher than the rest of the industries, indicating that these sectors might be prime targets or have vulnerabilities that are exploited more often. Moreover, for most industries, most reported breaches are confirmed. This indicates that the majority of reported breaches are not false alarms. Furthermore, the *Transportation* industry shows a relatively lower number of breaches (105), suggesting that this industry either has robust cybersecurity measures or is not as heavily targeted. Finally, industries such as *Information*, *Manufacturing*, *Other Services*, *Administrative*, and *Accommodation* all report zero near misses, suggesting underreporting or a very efficient initial detection system that quickly confirms or refutes suspected breaches.

## Preliminary Analysis

In this paper, we aim to conduct a comparative examination of the primary industries targeted by cybercriminals. In the next sections, we present our conducted analytical methods and highlight key insights derived from our findings.

**The Timeline Distribution of Incidents.** In examining the timeline mapping of incidents across various industries, we analyzed data from the VERIS dataset covering 2003 to 2022. Notably, 2003 to 2008 showed a comparatively low incident count, with just 147 incidents recorded. Thus, to understand the actual trend in the active region, we limit ourselves to 2008 onward. This decision

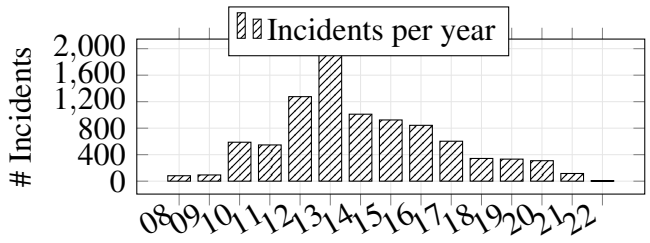



**Figure 6.1:** The distribution of incidents per year from 2008-2022.

allows for a more accurate understanding of the trends during the more active periods of cyber-attacks and breaches. Our findings reveal that the number of yearly incidents follows a normal distribution, peaking significantly in 2013 with 1,939 incidents, constituting 21.51% of the incidents during this period. Subsequent years, such as 2014 and 2015, also recorded high incidents, with 1,011 (11.22%) and 924 (10.25%) incidents, respectively.

However, contrary to the prevailing belief of a continuous increase in cyberattacks, we observed a downward trend in incident numbers post-2013, as illustrated in Figure 6.1. This trend challenges common perceptions and suggests potential improvements in cybersecurity measures, changes in attack strategies, or variations in reporting practices. This analysis is vital, as it provides insights into the active phases of cyber threats and helps understand the evolving nature of these threats and the effectiveness of the cybersecurity landscape over time.

> **Takeaway:** The analysis of cyberattack incidents from 2003 to 2022 reveals a critical peak in 2013, followed by an unexpected decline in subsequent years, suggesting significant changes in cybersecurity measures and attack strategies, challenging common beliefs about the continuous rise in cyber threats, and highlighting the dynamic nature of the cybersecurity landscape.

**The Global Distribution of Incidents.** In this section, we delved into the distribution of security breaches across the top 10 countries, utilizing the VERIS dataset and ISO 3,166 country codes

to identify the location of these incidents [37]. Our analysis aimed to shed light on the global cybersecurity incident landscape, identifying the most affected countries and exploring the underlying factors for these trends. Our analysis revealed that the *United States* contained 72.56% of the total incidents (6,487 out of 8,421). Following the U.S., the *United Kingdom* reported 574 incidents (6.42%), and *Canada* had 369 incidents (4.13%). The comprehensive results for the top ten countries are outlined in Table 6.2, with the remaining global incidents accounting for 9.61% (864 incidents). Moreover, to enhance the representation and understanding of our analysis, we created a geographical map depicting the distribution of incidents per country as shown in Figure 6.2, allowing for a more intuitive grasp of security breaches' global spread and providing a clear geographical perspective.

**Potential Reasons.** The prevalence of data breaches in the United States can be attributed to several regulatory and legislative factors. The Health Information Technology for Economic and Clinical Health (HITECH) Act plays a pivotal role in this regard, mandating that healthcare providers, healthcare plans, and other entities covered by the Health Insurance Portability and Accountability Act (HIPAA) to report data breaches to individuals affected, the media, and the Department of Health and Human Services (DHHS) in the event of a significant compromise of protected healthcare information [71]. This reporting requirement extends to breaches affecting over 500 individuals, necessitating notifications to the Secretary and media outlets in the relevant state or jurisdiction [71]. Complementing these requirements, the HIPAA security rule compels healthcare organizations to develop comprehensive risk management plans to safeguard personal health data. This robust regulatory framework in the United States likely contributes to the high number of reported incidents, as detailed in our analysis and supported by various sources, including the Office of Civil Rights and studies on healthcare data protection.

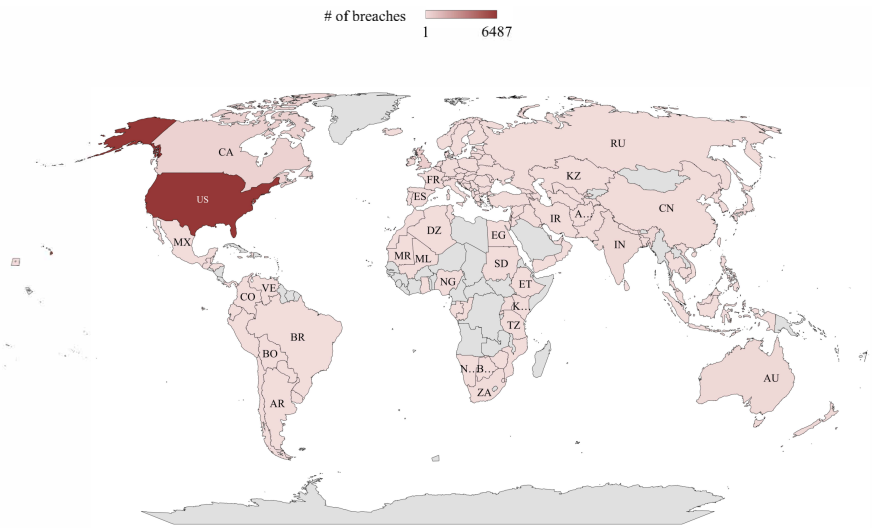



**Figure 6.2:** Geographical distribution of data breaches worldwide

**Takeaway:** The global analysis shows the U.S. leading in security breaches mainly due to strict reporting laws such as the HITECH Act, followed by the U.K. and Canada, highlighting the influence of regulatory frameworks on the visibility of data breaches.

**Organizations Size.** Table 6.3 presents the breakdown of data breaches across various indus-

**Table 6.2:** Top 10 countries ranked by the number of breaches.

| Country | # of breaches |
|---|---|
| United States | 6,487 |
| United Kingdom | 574 |
| Canada | 369 |
| Australia | 161 |
| India | 138 |
| New Zealand | 103 |
| China | 65 |
| Japan | 62 |
| Ireland | 60 |
| Germany | 57 |

**Table 6.3:** The distribution of Security Incidents by Industry and Organization Size, Highlighting Incident Status.

| NAICS Code | Industry | # of breaches | Small (1-1,000) | Large (1000+) | Unknown | Near Miss | Suspected | Confirmed |
|---|---|---|---|---|---|---|---|---|
| 62 | Healthcare | 2,445 | 913 | 734 | 798 | 1 | 20 | 2,424 |
| 92 | Administration | 2,330 | 272 | 1,238 | 820 | 2 | 15 | 2,313 |
| 52 | Finance | 804 | 263 | 388 | 153 | 1 | 2 | 801 |
| 51 | Information | 769 | 308 | 224 | 237 | 0 | 4 | 765 |
| 61 | Education | 520 | 172 | 225 | 123 | 2 | 7 | 511 |
| 54 | Professional | 406 | 224 | 74 | 108 | 1 | 4 | 401 |
| 44-45 | Retail | 467 | 220 | 160 | 87 | 1 | 4 | 351 |
| 31-33 | Manufacturing | 216 | 65 | 120 | 31 | 0 | 4 | 212 |
| 81 | Other Services | 205 | 115 | 24 | 66 | 0 | 0 | 205 |
| 56 | Administrative | 175 | 76 | 56 | 43 | 0 | 0 | 175 |
| 72 | Accommodation | 158 | 65 | 59 | 34 | 0 | 0 | 158 |
| 48-49 | Transportation | 105 | 21 | 55 | 29 | 0 | 0 | 105 |
| Total | - | 8,600 | 2,714 | 3,357 | 2,529 | 8 | 60 | 8,421 |

tries. In total, there have been 8,600 security incidents. When categorized by organization size, small-sized entities reported 2,714 incidents, large ones accounted for 3,357, while the size was unknown for 2,529 incidents. For small organizations, those with 1-1,000 employees, the *Healthcare* industry leads with the highest number of breaches, recording 913 incidents, followed by the *Administration* and *Information*industries with 272 and 308 incidents, respectively. For large organizations, those with over 1,000 employees, we observed that the *Administration* industry stands out with the most breaches, totaling 1,238 incidents, followed by the *Healthcare* sector with 734 incidents. In contrast, the *Finance* industry comes third, having 388 incidents. Finally, it is concerning for organizations of an unknown size to note the significant number of breaches. Our analysis shows that the *Healthcare* industry records 798 incidents, closely followed by the *Administration* sector with 820 breaches. It is important to mention that the lack of clarity regarding the size of these organizations underscores the importance of better reporting and categorization in cybersecurity data.

**Table 6.4:** The distribution (%) of opportunistic and targeted data breaches by industry.

| | Opportunistic | Targeted |
|---|---|---|
| Healthcare | 81.27 | 18.73 |
| Administration | 65.12 | 34.88 |
| Finance | 55.14 | 44.86 |
| Information | 63.74 | 36.26 |
| Education | 68.42 | 31.58 |
| Professional | 64.58 | 35.42 |
| Retail | 59.18 | 40.82 |
| Manufacturing | 61.90 | 38.10 |
| Other Services | 64.29 | 35.71 |
| Administrative | 73.91 | 26.09 |
| Accommodation | 70.00 | 30.00 |
| Transportation | 78.57 | 21.43 |

**Takeaway:** The breakdown of 8,600 data breaches by organization size reveals that small and large entities are significantly affected, with *Healthcare* being the most targeted industry. The high number of incidents in organizations of unknown size underscores the necessity for better data categorization and reporting in cybersecurity.

**Nature of Data breaches.** There are two types of data breaches: targeted and opportunistic breaches. In targeted breaches, the attackers set a specific target by researching and having a particular organization, system, or individual they want to breach. Their motivation can range from financial gain to corporate espionage or personal vendettas. While in opportunistic breaches, the attackers look for vulnerabilities in any system without a specific target in mind. If a system is easy to breach because of poor security measures, attackers will take advantage of it. The primary motive is to find any weak point available without preference for a specific target. As shown in Table 6.4, the *Healthcare* industry is the most vulnerable to opportunistic data breaches with 81.27%, indicating that many healthcare systems may have vulnerabilities or lack the latest security measures.

**Takeaway:** The *Healthcare* industry is the most susceptible to opportunistic data breaches (81.27%), highlighting significant security vulnerabilities.

### Discovery Methods

Early detection of data breaches is crucial for organizations to mitigate damage, protect their reputation, and prevent ongoing unauthorized access. However, despite its critical importance, numerous businesses face challenges in promptly recognizing these security incidents. Often, the realization of a breach occurs after significant damage has been done, highlighting a gap in effective real-time breach detection mechanisms [29]. The delay in identifying a data breach can lead to increased harm to organizational assets and customer data. This section aims to investigate organizations' various tools and methods to detect security breaches, focusing on both external and internal traffic. Our analysis is crucial for understanding the effectiveness of the mechanisms and guiding organizations in adopting the most efficient strategies to prevent future incidents.

**External Discovery Methods.** To detect breaches early, organizations deploy various external cybersecurity tools such as Intrusion Detection Systems (IDS), Security Information and Event Management (SIEM) Systems, Antivirus and Anti-Malware solutions, and Firewalls. These technologies are essential in early detection, as they monitor and flag anomalies, intrusions, and unauthorized activities within data and systems. We analyzed how security incidents are identified to gain insight into the breach discovery methods across different sectors.

Table 6.5 shows the results of our comprehensive analysis of how data breaches are externally identified across various industries. We observed that *actor disclosure* plays a significant role in detecting security breaches, specifically in the *Information* and *Other Services* sectors, accounting for 58.84% and 69.14%, respectively, suggesting a high engagement with external cybersecurity

**Table 6.5:** The distribution (%) of external data breach discovery sources across sectors. Abbreviations: Actor Disclosure (AD), Customer (Cust), Suspicious Traffic (ST), Law Enforcement (LE), Documents Found (DF), and Other (Oth).

| | AD | Cust | ST | LE | DF | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 25.04 | 36.38 | 0.16 | 12.91 | 17.64 | 7.87 |
| Administration | 36.09 | 23.51 | 21.19 | 5.96 | 4.64 | 8.61 |
| Finance | 20.63 | 46.03 | 0.00 | 13.89 | 1.59 | 17.86 |
| Information | 58.84 | 13.46 | 0.00 | 2.64 | 0.26 | 24.80 |
| Education | 50.94 | 27.04 | 0.63 | 8.18 | 2.52 | 10.69 |
| Professional | 43.33 | 12.22 | 2.22 | 8.89 | 1.67 | 31.67 |
| Retail | 12.64 | 36.26 | 0.00 | 20.33 | 4.95 | 25.82 |
| Manufacturing | 47.06 | 18.82 | 1.18 | 18.82 | 1.18 | 12.94 |
| Other Services | 69.14 | 17.28 | 0.00 | 7.41 | 0.00 | 6.17 |
| Administrative | 38.89 | 11.11 | 0.00 | 12.96 | 3.70 | 33.33 |
| Accommodation | 11.11 | 31.94 | 0.00 | 18.06 | 1.39 | 37.50 |
| Transportation | 41.30 | 32.61 | 0.00 | 13.04 | 0.00 | 13.04 |

experts in these industries. Moreover, *Customers* play a crucial role in detecting breaches, particularly in the *Finance* (46.03%) and *Retail* (36.26%) sectors, highlighting the importance of customer vigilance in identifying security issues.

Furthermore, *suspicious network traffic* is a notable method for discovering breaches in the *Administration* sector, contributing to 21.19% of discoveries; however, it appears less effective or used in other sectors. *Law enforcement's* involvement in breach discovery is significant in sectors such as *Healthcare* (12.91%) and *Retail* (20.33%), signifying the severe nature of breaches that require legal intervention. Discovering breaches through *documents found* is mainly prevalent in *Healthcare*, suggesting the ongoing risk posed by physical document management.

**Internal Discovery Methods.** Table 6.6 shows a detailed analysis of organizations' internal methods to detect data breaches across various industries. Our analysis found that most of the internal breaches are discovered based on *Employee Reporting* across all sectors, with significantly high us-

**Table 6.6:** The distribution (%) of different internal discovery methods used to detect data breaches across various industries. Abbreviations: Employee Report (ER), IT Review (ITR), Incident Response (IR), Log Review (LR), Fraud Detection (FD), Other (Oth).

| | ER | ITR | IR | LR | FD | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 68.31 | 12.36 | 3.82 | 4.04 | 0.45 | 11.01 |
| Administration | 63.05 | 12.81 | 4.43 | 2.96 | 1.48 | 15.27 |
| Finance | 66.67 | 4.50 | 6.31 | 2.70 | 7.21 | 12.61 |
| Information | 53.97 | 12.70 | 4.76 | 11.11 | 4.76 | 12.70 |
| Education | 77.23 | 4.95 | 3.96 | 0.00 | 0.00 | 13.86 |
| Professional | 77.78 | 0.00 | 8.89 | 0.00 | 2.22 | 11.11 |
| Retail | 79.41 | 1.47 | 2.94 | 4.41 | 2.94 | 8.82 |
| Manufacturing | 77.27 | 0.00 | 13.64 | 9.09 | 0.00 | 0.00 |
| Other Services | 86.96 | 4.35 | 4.35 | 0.00 | 0.00 | 4.35 |
| Administrative | 61.11 | 11.11 | 0.00 | 11.11 | 0.00 | 16.67 |
| Accommodation | 72.73 | 0.00 | 9.09 | 9.09 | 0.00 | 9.09 |
| Transportation | 60.00 | 0.00 | 6.67 | 0.00 | 6.67 | 26.67 |

age in *Education* (77.23%), *Professional Services* (77.78%), *Retail* (79.41%), and *Manufacturing* (77.27%). In addition, *IT Review* shows varying reliance across industries, with substantial usage in *Administration* (12.81%) and *Information* (12.70%) but completely absent in *Professional Services* and *Manufacturing*.

In contrast, *Incident Response* teams are significantly utilized in *Finance* (6.31%), *Professional Services* (8.89%), and *Manufacturing* (13.64%), indicating their importance in these areas. *Log Review* is most prominent in *Information* (11.11%) and *Manufacturing* (9.09%) but not in *Education* and *Other Services*. *Fraud Detection* methods find the most application in the *Finance* sector (7.21%), reflecting the need to safeguard financial data. This distribution of methods highlights the varying strategies employed by different industries, emphasizing the importance of employee awareness and sector-specific approaches in combating data breaches.

**Takeaway:** Data breach discovery analysis reveals a mix of external and internal strategies. Externally, actor disclosure and customer vigilance play vital roles, especially in the *Information* and *Finance* sectors. Internally, employee reporting is dominant across most industries.

### Threat Actions

In cybersecurity incidents, threat actions refer to the deliberate steps executed by adversaries that precipitate or amplify an adverse event. At least a single threat action precipitates each security breach; however, the complexity of such incidents often emerges from many actions that intersect across distinct categories. VERIS schema delineates these activities into seven fundamental categories: Malware, Hacking, Social Engineering, Misuse, Physical Actions, Errors, and Environmental [91]. In the following, we analyze the prevalence of these actions across various industries.

**Misuse Variety.** Misuse pertains to inappropriate administrative abuse and violations of policies to employ unauthorized assets. Such actions may be intentional or unintentional but are unified by their deviation from the authorized use. Notably, misuse is a threat vector from entities to which the organization has extended trust, including insiders and affiliates. In instances of misuse, the individual in question has legitimate access or privileges but improperly exploits them. Our analysis aims to classify the types of misuse implicated in security breaches. This classification will clarify the nature of the incidents, evaluate the underlying weaknesses in organizational controls, and formulate effective strategies for mitigation.

Table 6.7 presents the various types of action misuse observed during security breaches across industries. We observed that privilege abuse is the predominant issue across all industries, with the *Information* sector exhibiting the highest incidence at 89.36%, suggesting a significant concern regarding insider threats or unauthorized access within this field. Data mishandling also presents

**Table 6.7:** The distribution (%) of action misuse by industry. Abbreviations: Privilege Abuse (PA), Data Mishandling (DM), Knowledge Abuse (KA), Other (Oth).

| | PA | DM | KA | Oth |
|---|---|---|---|---|
| Healthcare | 80.51 | 10.71 | 5.14 | 3.64 |
| Administration | 79.86 | 11.54 | 8.14 | 0.45 |
| Finance | 78.13 | 9.38 | 11.46 | 1.04 |
| Information | 89.36 | 2.13 | 6.38 | 2.13 |
| Education | 54.29 | 25.71 | 8.57 | 11.43 |
| Professional | 73.91 | 10.87 | 10.87 | 4.35 |
| Retail | 72.22 | 8.33 | 16.67 | 2.78 |
| Manufacturing | 82.93 | 0.00 | 7.32 | 9.76 |
| Other Services | 70.00 | 30.00 | 0.00 | 0.00 |
| Administrative | 82.76 | 10.34 | 6.90 | 0.00 |
| Accommodation | 61.54 | 7.69 | 15.38 | 15.38 |
| Transportation | 66.67 | 0.00 | 22.22 | 11.11 |

a notable problem, especially in the *Education* and *Other Services* sectors, indicating potential weaknesses in data management protocols. Furthermore, Knowledge abuse is less prevalent but presents a challenge in the *Finance* industry at 11.46%, hinting at the misuse of insider information or expertise for malicious purposes.

Some sectors, such as *Manufacturing* and *Other Services*, reported no incidents in specific categories, which could either reflect robust security measures in those areas or a lack of reporting. *Retail* showcases a relatively even spread across misuse types, with privilege abuse still leading but with a significant fraction of incidents involving knowledge abuse. In contrast, the *Transportation* sector shows no data mishandling incidents, potentially indicating robust data protection strategies. Yet, it records a higher propensity for knowledge abuse and a noteworthy presence in the "Other" category.

**Misuse Vector.** Table 6.8 shows the vectors or access methods misused during security breaches. *Accommodation* leads with the highest percentage of breaches involving physical access at 85.00%,

**Table 6.8:** The distribution (%) of misuse action vectors by industry. Abbreviations: Physical Access (PA), LAN Access (LA), Remote Access (RA), Non-corporate Devices (NC), Other (Oth).

| | PA | LA | RA | NC | Oth |
|---|---|---|---|---|---|
| Healthcare | 25.23 | 70.45 | 1.44 | 2.34 | 0.54 |
| Administration | 14.48 | 81.07 | 1.56 | 1.11 | 1.78 |
| Finance | 21.10 | 70.64 | 4.59 | 0.92 | 2.75 |
| Information | 14.29 | 59.18 | 18.37 | 6.12 | 2.04 |
| Education | 18.18 | 72.73 | 0.00 | 3.03 | 6.06 |
| Professional | 29.41 | 60.78 | 9.80 | 0.00 | 0.00 |
| Retail | 43.48 | 43.48 | 6.52 | 4.35 | 2.17 |
| Manufacturing | 18.75 | 70.83 | 6.25 | 2.08 | 2.08 |
| Other Services | 61.11 | 27.78 | 5.56 | 5.56 | 0.00 |
| Administrative | 27.27 | 72.73 | 0.00 | 0.00 | 0.00 |
| Accommodation | 85.00 | 12.50 | 0.00 | 2.50 | 0.00 |
| Transportation | 41.18 | 47.06 | 0.00 | 5.88 | 5.88 |

highlighting a significant vulnerability in on-site security measures. In contrast, LAN access is predominantly exploited in the *Administration* and *Administrative* sectors, both at 72.73%, suggesting that internal network security is a critical concern. Notably, the *Information* sector is most affected by remote access breaches, accounting for 18.37% of its security incidents, underscoring its digital connectivity risks. Non-corporate access breaches are less frequent across all industries. Still, they are most prominent in *Information* at 6.12%, *Transportation*, and *Other Services*, both at 5.88%, indicating potential security gaps in external network management. Lastly, the *Education* sector demonstrates a diverse range of breach methods, with 6.06% classified under "Other," implying varied and less typical approaches to security breaches.

**Takeaway:** Privilege abuse is the predominant misuse in data breaches, with *Information* being the most affected sector. Additionally, sectors such as *Accommodation* and *Administration* face significant threats from physical and LAN access breaches. These insights highlight the security challenges and the necessity for targeted cybersecurity strategies across industries.

**Error Variety.** Table 6.9 presents the distribution of cybersecurity incidents caused by various action errors across multiple industries. Notably, the *Healthcare* and *Retail* sectors show significant vulnerabilities in data disposal, with rates of 25.51% and 30.00%, respectively, suggesting better protocols for handling and disposing of sensitive data. The issue of lost data is especially pronounced in *Healthcare*, where 28.62% of incidents are due to loss, highlighting the critical nature of securing physical devices and documents. In IT infrastructure, the *Information* and *Professional Services* sectors are affected by misconfigurations, with incidents occurring 47.26% and 42.86% of the time, indicating the complexity and potential for error in their IT systems. *Administration* stands out with a high rate of misdelivery at 55.39%, pointing to a recurring issue with sending information to incorrect recipients.

Publishing errors are a cross-industry concern, with the *Education* sector recording a 28.87% incidence rate, *Information* at 26.03%, and *Finance* at 16.67%, suggesting stricter controls on information release processes. The *Manufacturing* sector's most significant challenge comes from malfunctions, with a 6.67% share of incidents, possibly linked to using industrial machinery and control systems prone to failure. The 'Other' category indicates various unspecified errors that differ by industry, underscoring unique challenges that may not be captured by general classifications. Moreover, a consistent 0% in certain types of errors for various industries suggests that some errors are either exceptionally well-managed within those sectors or have not been detected during the data collection timeframe. This overview underlines that cybersecurity issues are multifaceted and that each industry faces distinct challenges, necessitating customized security approaches to mitigate risks associated with human error effectively.

**Error Vectors.** Our analysis revealed that carelessness is the predominant factor contributing to data breaches across industries, as shown in Table 6.10. This is particularly pronounced in the *Administration* sector, where it accounts for 94.57% of breaches. Random errors are far less

**Table 6.9:** The distribution (%) of action errors for cybersecurity incidents in various industries. Abbreviations: Misdelivery (MD), Loss (L), Publishing (P), Disposal (D), Misconfiguration (MC), Other (Oth).

| | MD | L | P | D | MC | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 23.82 | 28.09 | 9.62 | 25.04 | 4.89 | 8.55 |
| Administration | 54.78 | 13.69 | 12.68 | 6.53 | 4.14 | 8.18 |
| Finance | 39.90 | 14.14 | 16.16 | 5.56 | 15.15 | 9.09 |
| Information | 6.04 | 2.01 | 25.50 | 2.01 | 46.31 | 18.12 |
| Education | 33.10 | 7.59 | 28.28 | 7.59 | 15.86 | 7.59 |
| Professional | 16.00 | 5.00 | 18.00 | 9.00 | 42.00 | 10.00 |
| Retail | 20.00 | 6.67 | 5.00 | 30.00 | 26.67 | 11.67 |
| Manufacturing | 9.68 | 16.13 | 16.13 | 9.68 | 29.03 | 19.35 |
| Other Services | 26.47 | 14.71 | 26.47 | 8.82 | 14.71 | 8.82 |
| Administrative | 6.25 | 15.63 | 18.75 | 12.50 | 9.38 | 37.50 |
| Accommodation | 33.33 | 0.00 | 0.00 | 27.78 | 33.33 | 5.56 |
| Transportation | 13.04 | 8.70 | 21.74 | 8.70 | 17.39 | 30.43 |

common, with the highest incidence reported in the *Other Services* sector at 10.00%. The data indicates that breaches due to inadequate personnel are relatively rare, with many sectors reporting no breaches from this vector. Inadequate technology does not frequently lead to breaches either, except in the *Accommodation* sector, which reports a significant 20.00%, signaling a potential area for technological enhancement. The 'Other' category is minimally represented, suggesting that most error vectors fall within the defined categories. Our comprehensive analysis emphasizes the crucial role of human error in data security and highlights the need for improved diligence and procedural rigor to reduce the prevalence of data breaches.

**Takeaway:** Data disposal and loss issues are high in *Healthcare* and *Retail*, while IT misconfigurations significantly impact *Information* and *Professional Services* . The *Administration* is particularly prone to misdelivery errors. These findings underscore the varied cybersecurity challenges across industries, primarily driven by human error, highlighting the need for customized security approaches.

**Table 6.10:** The distribution (%) of the common error vectors leading to data breaches. Abbreviations: Random Error (RE), Carelessness (CL), Inadequate Personnel (IP), Inadequate Procedures (IPc), Inadequate Technology (IT), Other (Oth).

| | RE | CL | IP | IPc | IT | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 2.25 | 90.99 | 0.45 | 4.05 | 0.00 | 1.80 |
| Administration | 0.14 | 94.57 | 0.42 | 2.92 | 0.28 | 0.84 |
| Finance | 0.00 | 88.14 | 0.00 | 10.17 | 1.69 | 0.00 |
| Information | 1.32 | 92.11 | 0.00 | 3.95 | 0.00 | 2.63 |
| Education | 3.85 | 78.85 | 1.92 | 11.54 | 3.85 | 0.00 |
| Professional | 3.39 | 86.44 | 0.00 | 6.78 | 1.69 | 1.69 |
| Retail | 6.67 | 83.33 | 3.33 | 6.67 | 0.00 | 0.00 |
| Manufacturing | 7.14 | 92.86 | 0.00 | 0.00 | 0.00 | 0.00 |
| Other Services | 10.00 | 90.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Administrative | 0.00 | 84.62 | 0.00 | 15.38 | 0.00 | 0.00 |
| Accommodation | 0.00 | 80.00 | 0.00 | 0.00 | 20.00 | 0.00 |
| Transportation | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 |

**Physical Variety.** In the context of data breaches, physical security breaches occur when individuals unlawfully gain access to an organization's premises to conduct data theft or damage. Physical attacks stand in contrast to cyberattacks, which are executed remotely via digital means. Physical breaches involve direct, onsite actions, ranging from unauthorized access by individuals to sensitive areas to intricate strategies such as hardware tampering or destruction. Such breaches are characterized by immediate physical interaction and may include a variety of actions such as theft, hardware manipulation, eavesdropping, sabotage, direct access to devices, or even assault. A significant instance of this kind of breach was the case involving Albert Gonzales, who physically broke into Heartland Payment Systems' offices in Florida. Gonzales stole computer equipment containing sensitive data, such as payroll information, social security numbers, and banking details. The fallout from this breach was severe, costing the company approximately $140 million in damages, legal fees, and penalties [87].

We analyzed the physical data breaches to investigate the various physical attacks involved. Ta-

ble 6.11 details the variety of physical actions in data breaches across different sectors. Our analysis reveals that *theft* emerges as the most frequent physical action linked to data breaches in nearly all sectors, with exceptionally high occurrences in *Manufacturing*, where it accounts for 100% of the breaches, and in *Retail*, at 74.29%, suggesting that directly pilfering devices, documents, or data storage devices is a primary concern across these industries.

The *Accommodation* sector stands out with a high rate of *assault* incidents at 25%, indicating more aggressive methods of breaches. The instances of *snooping* and *surveillance* are generally minimal across most sectors; however, the *Information* sector reports a higher occurrence of surveillance activities at 8.33%, hinting at more sophisticated means of unauthorized *Information* gathering. Notably, *tampering* is a significant concern in sectors such as *Finance* and *Retail*, with 19.05% and 20.00%, respectively, indicating that tampering with equipment or data is a favored method in these industries.

Our analysis shows sector-specific trends in physical breach methods. For instance, the *Healthcare* sector primarily faces threats from theft (96.38%), while *Accommodation* shows a more varied profile with substantial incidents of assault and tampering. Despite these trends, it is noteworthy that data breaches caused by physical loss have decreased considerably in recent years [39]. The decline suggests a growing effectiveness of physical security measures and heightened awareness of physical breaches in various industries. Our findings highlight the need for sector-specific physical security strategies to safeguard against each industry's risks and challenges in protecting its physical infrastructure against data breaches.

**Physical Vector.** Following the various physical actions during data breaches, we analyzed the locations where data breaches can occur. Table 6.12 provides an in-depth look at the areas where physical data breaches most frequently occur across various industries.

**Table 6.11:** The distribution (%) of physical actions variety in data breaches across sectors.

| | Assault | Snooping | Surveillance | Tampering | Theft | Other |
|---|---|---|---|---|---|---|
| Healthcare | 1.65 | 1.15 | 0.33 | 0.33 | 96.38 | 0.16 |
| Administration | 3.62 | 2.90 | 1.45 | 2.17 | 89.86 | 0.00 |
| Finance | 1.59 | 0.00 | 1.59 | 19.05 | 76.19 | 1.59 |
| Information | 0.00 | 0.00 | 8.33 | 8.33 | 83.33 | 0.00 |
| Education | 1.64 | 0.00 | 0.00 | 1.64 | 95.08 | 1.64 |
| Professional | 3.57 | 0.00 | 0.00 | 0.00 | 96.43 | 0.00 |
| Retail | 5.71 | 0.00 | 0.00 | 20.00 | 74.29 | 0.00 |
| Manufacturing | 0.00 | 0.00 | 0.00 | 0.00 | 100.00 | 0.00 |
| Other Services | 10.53 | 0.00 | 0.00 | 0.00 | 89.47 | 0.00 |
| Administrative | 5.26 | 0.00 | 0.00 | 0.00 | 94.74 | 0.00 |
| Accommodation | 25.00 | 0.00 | 0.00 | 25.00 | 50.00 | 0.00 |
| Transportation | 0.00 | 0.00 | 0.00 | 11.11 | 88.89 | 0.00 |

Table 6.12 gives a detailed overview of the most common locations for physical data breaches across different sectors. Our analysis shows that the *Victim Work* area is the most frequent site for physical breaches, particularly in the *Professional* sector at 53.85% and in *Healthcare* at 48.08%, suggesting that internal workspaces often pose substantial risks due to easy access to information. Breaches in *Personal Vehicles* are also noteworthy, especially in the *Administration* sector at 29.17% and in *Healthcare* at 22.25%, indicating vulnerabilities related to transporting or storing sensitive data in personal vehicles.

We noticed lower breaches within the *Victims Secure* areas, with the *Information* sector at 10.00% and *Other Services* at 9.52%, implying better protection in these high-security zones. *Personal Residences* emerge as notable breach locations in *Education* and *Manufacturing*, with 8.62% and 14.29%, respectively, highlighting the risks involved in handling sensitive data at home. Breaches in *Partner Facilities* are prevalent in *Other Services* (23.81%) and *Administration* (11.76%), suggesting that data security risks extend to locations affiliated with business partners.

Our analysis highlights the nature of physical breach locations, emphasizing the need for industry-

**Table 6.12:** The distribution (%) distribution of physical action vectors during data breaches across sectors. Abbreviations: Victim's Work Area (VW), Personal Vehicle (PV), Victim's Secure Area (VSA), Personal Residence (PR), Participant Facility (PF), Other (Oth).

| | VW | PV | VSA | PR | PF | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 48.08 | 22.25 | 6.39 | 6.14 | 5.88 | 11.25 |
| Administration | 26.39 | 29.17 | 9.03 | 6.25 | 7.64 | 21.53 |
| Finance | 8.91 | 5.45 | 2.48 | 1.98 | 4.46 | 76.73 |
| Information | 20.00 | 20.00 | 10.00 | 0.00 | 10.00 | 40.00 |
| Education | 46.55 | 24.14 | 3.45 | 8.62 | 3.45 | 13.79 |
| Professional | 53.85 | 19.23 | 0.00 | 11.54 | 3.85 | 11.54 |
| Retail | 8.81 | 1.26 | 1.89 | 0.00 | 5.03 | 83.02 |
| Manufacturing | 28.57 | 23.81 | 0.00 | 14.29 | 0.00 | 33.33 |
| Other Services | 28.57 | 23.81 | 9.52 | 4.76 | 23.81 | 9.52 |
| Administrative | 47.06 | 17.65 | 0.00 | 0.00 | 11.76 | 23.53 |
| Accommodation | 19.05 | 14.29 | 0.00 | 4.76 | 9.52 | 52.38 |
| Transportation | 31.25 | 6.25 | 0.00 | 6.25 | 0.00 | 56.25 |

specific security strategies that account for each sector's unique challenges and vulnerabilities. The analysis underscores the importance of robust security measures within the conventional work environment and in personal and partner domains to safeguard against physical data breaches.

**Takeaway:** Theft is the main physical action in data breaches, notably in *Manufacturing* and *Retail*. *Accommodation* and *Finance* sectors are highly susceptible to assault and tampering. Breaches often occur in work areas and personal vehicles, highlighting their vulnerability. This underlines the urgency for strong, industry-specific physical security measures.

**Hacking Variety.** Hacking is unauthorized or excessive access to information assets by bypassing security measures and employing various methods such as brute force, SQL injection, cryptanalysis, and denial of service (DoS) attacks [91]. Table 6.13 overviews the various hacking methods prevalent in different industries, underscoring specific vulnerabilities and common hacking tactics. In the *Healthcare* sector, *stolen credentials* is the dominant issue, accounting for an overwhelming

**Table 6.13:** The distribution (%) of hacking types used during data breach incidents across industries.

| | Brute force | DoS | Stolen Creds |
|---|---|---|---|
| Healthcare | 6.56 | 4.92 | 88.52 |
| Administration | 10.77 | 49.23 | 40.00 |
| Finance | 5.56 | 31.48 | 62.96 |
| Information | 12.09 | 49.45 | 38.46 |
| Education | 10.81 | 24.32 | 64.86 |
| Professional | 19.23 | 30.77 | 50.00 |
| Retail | 14.29 | 14.29 | 71.43 |
| Manufacturing | 16.67 | 8.33 | 75.00 |
| Other Services | 6.25 | 68.75 | 25.00 |
| Administrative | 20.00 | 10.00 | 70.00 |
| Accommodation | 63.64 | 0.00 | 36.36 |
| Transportation | 14.29 | 28.57 | 57.14 |

88.52% of breaches. In contrast, the *Administration* sector is more frequently targeted by *DoS attacks*, representing 49.23% of its breaches.

The *Finance* and *Education* sectors are significantly affected by *stolen credentials*, with 62.96% and 64.86% of breaches, respectively. In addition, the *Information* sector exhibits an almost equal division between *DoS attacks* and *stolen credentials*. *Professional*, *Retail*, and *Manufacturing* sectors primarily suffer from *stolen credentials* breaches, though *Retail* also demonstrates a balanced mix of *brute force* and *DoS attacks*. In contrast, *DoS attacks* predominantly afflict the *Other Services* sector, accounting for 68.75% of its breaches, while *brute force* attacks are most prevalent in the *Accommodation* sector, making up 63.64% of its incidents.

**Hacking Vectors.** We analyzed the various hacking vectors employed during data breach incidents to provide insights into the common hacking vectors used in data breaches across industries. Understanding these vectors can help organizations and cybersecurity *Professional*s better prepare for and protect against potential threats. In the *Healthcare* sector, the primary method is through Web

**Table 6.14:** The distribution (%) of the various hacking vectors employed during data breach incidents across industries.

| | Backdoor | Phys. Acc. | Cmd. Shell | Web App. | Other |
|---|---|---|---|---|---|
| Healthcare | 4.62 | 2.31 | 0.00 | 83.85 | 9.23 |
| Administration | 36.93 | 1.14 | 0.28 | 59.66 | 1.99 |
| Finance | 4.72 | 1.89 | 0.00 | 88.68 | 4.72 |
| Information | 3.33 | 0.30 | 0.30 | 93.03 | 3.03 |
| Education | 1.11 | 2.22 | 1.11 | 90.00 | 5.56 |
| Professional | 10.20 | 0.00 | 2.04 | 80.61 | 7.14 |
| Retail | 1.16 | 0.00 | 0.00 | 93.02 | 5.81 |
| Manufacturing | 4.88 | 2.44 | 0.00 | 90.24 | 2.44 |
| Other Services | 3.03 | 0.00 | 1.52 | 93.94 | 1.52 |
| Administrative | 2.33 | 2.33 | 2.33 | 93.02 | 0.00 |
| Accommodation | 7.14 | 14.29 | 0.00 | 57.14 | 21.43 |
| Transportation | 12.50 | 0.00 | 0.00 | 79.17 | 8.33 |

Applications, accounting for a significant 83.85% of incidents, with a minor percentage involving backdoors (4.62%) and physical access (2.31%) as shown in Table 6.14. The *Administration* sees a substantial use of backdoors at 36.93%, while Web Application hacks also remain high at 59.66%. The *Finance* and *Education* sectors predominantly experience breaches through Web Applications, with 88.68% and 90.00%, respectively.

The *Information* sector similarly heavily relies on Web Application hacking, constituting 93.03% of breaches. In the *Professional* sector, while 80.61% of incidents occur via Web Applications, backdoor attacks are also notable at 10.20%. *Retail* and *Manufacturing* industries predominantly face breaches through Web Applications, at 93.02% and 90.24%, respectively. *Other Services* have an overwhelming majority of Web Application hacks (93.94%), and the *Administrative* sector is nearly as reliant on this method (93.02%). Lastly, the *Transportation* sector experiences a mix, with Web Application hacks leading at 79.17% and a notable 12.50% in backdoor attacks. These results highlight a significant industry-wide reliance on Web Application vulnerabilities, with varying degrees of backdoor and physical access breaches across sectors.

**Takeaway:** In various industries, Web Application vulnerabilities are the main unauthorized entry point, with a prevalence of 79.17% to 93.94%, highlighting the critical need for secure web systems. Despite this commonality, the variation in backdoor and physical access breaches across sectors underscores the need for industry-specific cybersecurity strategies.

**Social Variety.** Social actions include diverse tactics used in data breaches across various industries, underlining the critical importance of the human element in cybersecurity. These tactics include deception, manipulation, intimidation, and other forms of exploitation of users of information assets, which are categorized into seven types for this analysis: bribery, extortion, forgery, scam, phishing, pretexting, and others. These tactics range from pretexting and phishing to blackmail, threats, and scams.

Our findings reveal a significant prevalence of *Phishing*, in the *Administration* (82.16%), *Information* (63.41%), and *Education* (77.08%) sectors, highlighting the widespread use of deceptive communication to extract sensitive information as shown in Table 6.15. *Bribery* also stands out, in the *Accommodation* (60.00%) and *Administrative* (27.27%) sectors, indicating a susceptibility to financial inducements. While *Extortion* and *Pretexting* are less common, they are notable in specific sectors, with *Extortion* reaching 20.00% in *Other Services* and *Pretexting* 15.56% in *Finance*. The *Retail* sector shows a balanced use of various tactics, including *Forgery* (14.29%) and *Pretexting* (35.71%). The absence of specific social tactics such as *Scams* and *Forgery* in sectors such as *Healthcare* suggests specific vulnerabilities or protective measures unique to those industries.

Our analysis highlights the critical need for industry-specific cybersecurity strategies that account for the diverse array of social tactics employed by attackers. Understanding these varied approaches and their prevalence in different sectors is crucial for developing more effective defenses against the multifaceted nature of social engineering threats, which exploit the human aspect of information security.

**Table 6.15:** The distribution (%) of social attack variety in data breach incidents across industries. Abbreviations: Bribery (B), Extortion (E), Forgery (F), Scam (S), Phishing (P), Pretexting (PT), Other (Oth).

| | B | E | F | S | P | PT | Oth |
|---|---|---|---|---|---|---|---|
| Healthcare | 14.96 | 3.94 | 4.72 | 0.00 | 66.14 | 6.30 | 3.94 |
| Administration | 3.78 | 2.70 | 1.08 | 0.00 | 82.16 | 4.86 | 5.41 |
| Finance | 17.78 | 6.67 | 4.44 | 0.00 | 51.11 | 15.56 | 4.44 |
| Information | 4.88 | 12.20 | 0.00 | 4.88 | 63.41 | 9.76 | 4.88 |
| Education | 4.17 | 2.08 | 2.08 | 0.00 | 77.08 | 10.42 | 4.17 |
| Professional | 9.09 | 0.00 | 0.00 | 4.55 | 77.27 | 4.55 | 4.55 |
| Retail | 7.14 | 7.14 | 14.29 | 0.00 | 35.71 | 35.71 | 0.00 |
| Manufacturing | 14.81 | 7.41 | 0.00 | 0.00 | 55.56 | 11.11 | 11.11 |
| Other Services | 10.00 | 20.00 | 0.00 | 0.00 | 50.00 | 20.00 | 0.00 |
| Administrative | 27.27 | 18.18 | 0.00 | 0.00 | 45.45 | 9.09 | 0.00 |
| Accommodation | 60.00 | 0.00 | 0.00 | 0.00 | 40.00 | 0.00 | 0.00 |
| Transportation | 0.00 | 0.00 | 0.00 | 0.00 | 71.43 | 28.57 | 0.00 |

**Social Vectors.** We investigated the various social vectors or communication channels employed in data breaches across industries, shedding light on the modalities of executing social tactics such as pretexting, phishing, and other manipulative strategies. We categorized the vectors into documents, emails, in-person interactions, phone calls, software, and others, providing a nuanced understanding crucial for formulating sector-specific cybersecurity policies and employee training.

Our analysis shows an overwhelming dominance of *Email* as the preferred vector across most sectors, with its usage notably high in *Education* (97.67%), *Administration* (90.45%), and *Information* (79.31%). This trend indicates the vulnerability of *Email* to deceptive and manipulative attacks, owing to its extensive reach and ease of disseminating misleading information. Meanwhile, sectors such as *Accommodation* and *Finance* exhibit a significant reliance on more direct communication channels, with *In-person* interactions being predominant in *Accommodation* (60.00%) and both *In-person* (21.05%) and *Phone* (10.53%) vectors being substantial in *Finance*, suggesting a customized approach in these sectors, leveraging personal communication for executing attacks.

**Table 6.16:** The distribution (%) of social attack vectors used in data breach incidents across sectors. Abbreviations: Documents (D), Email (E), In-person (IP), Phone (P), Software (S), Other (Oth).

| | D | E | IP | P | S | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 3.39 | 77.12 | 14.41 | 2.54 | 0.85 | 1.69 |
| Administration | 1.12 | 90.45 | 5.06 | 1.12 | 0.56 | 1.69 |
| Finance | 7.89 | 55.26 | 21.05 | 10.53 | 2.63 | 2.63 |
| Information | 0.00 | 79.31 | 6.90 | 3.45 | 6.90 | 3.45 |
| Education | 0.00 | 97.67 | 0.00 | 0.00 | 0.00 | 2.33 |
| Professional | 0.00 | 80.00 | 0.00 | 15.00 | 5.00 | 0.00 |
| Retail | 8.33 | 50.00 | 25.00 | 16.67 | 0.00 | 0.00 |
| Manufacturing | 0.00 | 68.42 | 10.53 | 5.26 | 5.26 | 10.53 |
| Other Services | 0.00 | 77.78 | 11.11 | 0.00 | 11.11 | 0.00 |
| Administrative | 0.00 | 62.50 | 37.50 | 0.00 | 0.00 | 0.00 |
| Accommodation | 0.00 | 40.00 | 60.00 | 0.00 | 0.00 | 0.00 |
| Transportation | 0.00 | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 |

Additionally, the use of *Documents* and *Software* as vectors in sectors such as *Finance* and *Retail* (7.89% and 8.33% for *Documents*, respectively) highlights the involvement of more sophisticated attack methods involving falsified documents or malicious software. The sector-specific patterns, such as the exclusive reliance on *Email* in *Accommodation* (100.00%) and a considerable focus on *In-person* methods in administrative services (37.50%), further underline the need for customized security strategies.

Analyzing social vectors in data breaches reveals a complex landscape of communication channels with varying preferences and usages across different sectors. The pervasive use of Email and significant reliance on direct communication methods in certain industries emphasize the multifaceted nature of social engineering threats. These insights are vital for developing comprehensive, industry-specific defenses and employee awareness programs to mitigate the risks posed by these varied social attack vectors.

**Takeaway:** Social engineering tactics vary widely across industries; for instance, phishing is notably prevalent in *Administration* (82.16%), *Information* (63.41%), and *Education* (77.08%). The sector-specific percentages highlight the need for tailored cybersecurity strategies. Understanding communication channels, such as email (universal) and sector-specific methods, is crucial for building effective, customized defenses.

**Malware Variety.** We investigated the various types of malware detected in data breaches, providing a comprehensive understanding of the malware threat landscape. We categorized malware into applications, stored data, client-side, RAM scraping, ransomware, spyware/keylogging, and others, allowing for a sector-specific analysis of malware threats.

Our analysis revealed a predominant role of Ransomware in several sectors, mainly *Healthcare* (91.43%), *Administration* (86.84%), and *Education* (88.37%), suggesting that attackers frequently target these sectors with Ransomware due to the sensitive nature of their data and the potential for substantial impact. On the other hand, sectors such as *Retail* and *Manufacturing* exhibit a notable occurrence of Application-related malware (30.00% and 26.32%, respectively) and RAM Scraping (30.00% in *Retail* and 10.53% in *Manufacturing*), indicating vulnerabilities in point-of-sale systems and application-based platforms within these industries.

Furthermore, the *Accommodation* and *Administrative* sectors show a diverse range of malware types, with *Accommodation* showing a significant prevalence of RAM Scraping (44.00%) and Spyware/Keylogging (16.00%), and the *Administrative* sector reporting a notable presence of Stored Data related malware (16.67%). Additionally, *Finance* records a considerable presence of Client-side malware and Other types (8.33% each), suggesting unique threats in these areas.

Unique patterns are also observed in sectors such as *Transportation*, heavily impacted by Ransomware (88.89%), and *Information*, with a significant percentage of Other malware types (13.64%).

**Table 6.17:** The distribution (%) of malware variety detected in data breaches across sectors.

| | App data | Stored Data | Client-side | Ram Scrap. | Ransom. | Spy./Keylog. | Other |
|---|---|---|---|---|---|---|---|
| Healthcare | 0.00 | 3.57 | 0.00 | 0.00 | 91.43 | 2.14 | 2.86 |
| Administration | 0.00 | 2.63 | 0.00 | 0.00 | 86.84 | 5.26 | 5.26 |
| Finance | 8.33 | 0.00 | 8.33 | 0.00 | 66.67 | 8.33 | 8.33 |
| Information | 0.00 | 4.55 | 0.00 | 0.00 | 81.82 | 0.00 | 13.64 |
| Education | 0.00 | 0.00 | 0.00 | 0.00 | 88.37 | 9.30 | 2.33 |
| Professional | 0.00 | 3.70 | 0.00 | 0.00 | 88.89 | 7.41 | 0.00 |
| Retail | 30.00 | 5.00 | 0.00 | 30.00 | 10.00 | 5.00 | 20.00 |
| Manufacturing | 26.32 | 0.00 | 0.00 | 10.53 | 63.16 | 0.00 | 0.00 |
| Other Services | 11.11 | 0.00 | 11.11 | 0.00 | 66.67 | 11.11 | 0.00 |
| Administrative | 0.00 | 16.67 | 0.00 | 0.00 | 66.67 | 16.67 | 0.00 |
| Accommodation | 20.00 | 0.00 | 0.00 | 44.00 | 12.00 | 16.00 | 8.00 |
| Transportation | 11.11 | 0.00 | 0.00 | 0.00 | 88.89 | 0.00 | 0.00 |

These findings highlight the varied nature of malware attacks across different sectors, emphasizing the necessity of sector-specific cybersecurity strategies. Understanding these diverse malware patterns is crucial for developing effective defenses against each sector's specific threats, reinforcing the importance of tailored cybersecurity measures in the increasingly complex digital landscape.

**Malware vectors.** We examined the distribution vectors of malware in data breaches across various sectors. We categorize the vectors into Direct Installation, Downloaded by Other Malware, Email Link, Remote Injection, Removable Media, and Others. A key observation from the analysis is the overwhelming prevalence of Direct Installation as the primary vector across several sectors. This method is particularly dominant in *Healthcare* (65.38%), *Administration* (80.00%), and *Finance* (80.00%), with *Information*, *Education*, *Other Services*, and *Accommodation* sectors showing an exclusive reliance on it (100.00% each), suggesting that attackers frequently gain direct access to systems or exploit vulnerabilities for malware installation.

Additionally, certain sectors such as *Transportation* and *Retail* display a notable use of other vectors such as Removable Media (25.00% in *Transportation*) and Remote Injection (16.67% in *Retail*), indicating diverse approaches in these industries. The use of Email Links for malware distri-

**Table 6.18:** The distribution (%) of malware vectors in data breaches across various sectors. Abbreviations: Direct Install (DI), Downloaded by Malware (DM), Email Link (EL), Remote Injection (RI), Removable Media (RM), Other (Oth).

| | DI | DM | EL | RI | RM | Oth |
|---|---|---|---|---|---|---|
| Healthcare | 65.38 | 3.85 | 19.23 | 3.85 | 0.00 | 7.69 |
| Administration | 80.00 | 0.00 | 10.00 | 0.00 | 0.00 | 10.00 |
| Finance | 80.00 | 0.00 | 0.00 | 0.00 | 10.00 | 10.00 |
| Information | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Education | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Professional | 80.00 | 0.00 | 13.33 | 0.00 | 0.00 | 6.67 |
| Retail | 72.22 | 11.11 | 0.00 | 16.67 | 0.00 | 0.00 |
| Manufacturing | 87.50 | 0.00 | 0.00 | 0.00 | 0.00 | 12.50 |
| Other Services | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Administrative | 80.00 | 0.00 | 0.00 | 20.00 | 0.00 | 0.00 |
| Accommodation | 100.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| Transportation | 75.00 | 0.00 | 0.00 | 0.00 | 25.00 | 0.00 |

bution, while less common, is still significant in sectors such as *Healthcare* (19.23%). Similarly, the presence of malware being Downloaded by Other Malware is observed in *Retail* (11.11

Each sector exhibits unique patterns of malware distribution. For example, *Professional Services* demonstrate a mix with a majority of Direct Installation (80.00%) and a notable portion through Email Links (13.33%). This sector-specific distribution of malware vectors underscores the complexity of the threat landscape and highlights the necessity for tailored cybersecurity strategies. Understanding these varied distribution methods is crucial in developing robust defenses against the wide range of malware distribution tactics employed in cyberattacks, emphasizing the need for comprehensive and sector-specific cybersecurity measures.

**Takeaway:** Malware threats vary across sectors, with Ransomware dominating sensitive sectors such as *Healthcare*, *Administration*, and *Education*. *Retail* and *Manufacturing* face application related challenges, while *Finance* encounters unique threats. Direct installation is a common vector, but *Retail* and *Transportation* show diversity.

### Victims Types

Table 6.19 presents the primary targets of security breaches within various sectors. Our analysis revealed that patients are the primary victims within the *Healthcare* industry, accounting for 87.84% of breaches, indicating a significant risk to patient data. In contrast, the Customer category is the most frequent victim in many industries, with *Finance* (75.66%), *Information* (75.60%), *Retail* (85.16%), and *Accommodation* (84.62%) demonstrating high percentages, highlighting the broad risk to customer information across sectors. Students are primary victims within the *Education* sector (54.64%), reflecting the sector's responsibility for a significant amount of student data. Notably, in the *Manufacturing* and *Other Services* sectors, no single type of victim dominates, suggesting a more distributed risk profile across different stakeholder groups.

**Takeaway:** Patient data is a significant target in *Healthcare*, while customer information is consistently at risk across sectors such as *Finance*, *Information*, *Retail*, and *Accommodation*. In *Education*, students are the main victims.

### Actor Motives

Understanding the motives behind data breaches is crucial in comprehending the dynamics of cybersecurity threats and formulating effective countermeasures. This section delves into the multifaceted motivations that drive both internal and external actors to compromise data security across

**Table 6.19:** The distribution (%) of victims types by industry. Abbreviations: Patient (Pt), Customer (Cust), Employee (Emp), Other (Oth), Partner (Part), Student (Stud), Victim Organization (VO).

| | Pt | Cust | Emp | Oth | Part | Stud | VO |
|---|---|---|---|---|---|---|---|
| Healthcare | 87.84 | 4.95 | 4.19 | 3.01 | 0.00 | 0.00 | 0.00 |
| Administration | 19.81 | 33.23 | 16.34 | 30.63 | 0.00 | 0.00 | 0.00 |
| Finance | 14.04 | 75.66 | 4.99 | 5.30 | 0.00 | 0.00 | 0.00 |
| Information | 0.00 | 75.60 | 3.30 | 17.98 | 3.12 | 0.00 | 0.00 |
| Education | 9.81 | 0.00 | 13.79 | 21.75 | 0.00 | 54.64 | 0.00 |
| Professional | 13.44 | 58.69 | 7.54 | 20.33 | 0.00 | 0.00 | 0.00 |
| Retail | 8.03 | 85.16 | 0.00 | 4.62 | 2.19 | 0.00 | 0.00 |
| Manufacturing | 0.00 | 49.68 | 7.10 | 32.26 | 10.97 | 0.00 | 0.00 |
| Other Services | 0.00 | 52.82 | 11.27 | 26.06 | 9.86 | 0.00 | 0.00 |
| Administrative | 14.38 | 56.16 | 0.00 | 21.92 | 7.53 | 0.00 | 0.00 |
| Accommodation | 0.00 | 84.62 | 10.77 | 4.62 | 0.00 | 0.00 | 0.00 |
| Transportation | 0.00 | 58.44 | 19.48 | 15.58 | 0.00 | 0.00 | 6.49 |

various industries. By dissecting these motives, we can gain invaluable insights into the nature of these breaches and tailor more effective strategies to mitigate them.

**Internal Motives.** Table 6.20 shows the internal motives for cybersecurity incidents reveal that unintentional actions are the leading cause of breaches in *Healthcare* (51.63%), *Administration* (78.04%), and *Education* (76.58%), suggesting a need for better training and security practices. Financial motives are predominant in the *Accommodation* (74.19%), *Finance* (29.49%), and *Manufacturing* (32.31%) industries, indicating a need for strict internal controls to counteract insider financial threats. Espionage is notably significant in *Manufacturing* (27.69%), highlighting the importance of protecting intellectual property. The *Professional* sector shows a considerable concern with espionage at 8.22%, suggesting the need for enhanced data confidentiality measures. These trends underscore the necessity for industry-specific security strategies to combat the diverse insider threat landscape.

**Table 6.20:** The distribution (%) of internal motives for cybersecurity incidents in various industries.

| | Unintentional | Espionage | Fear | Financial | Fun | Grudge | Ideology | Convenience | Other |
|---|---|---|---|---|---|---|---|---|---|
| Healthcare | 51.63 | 1.36 | 0.00 | 24.03 | 14.69 | 2.31 | 0.21 | 3.99 | 1.78 |
| Administration | 78.04 | 0.82 | 0.00 | 8.46 | 6.68 | 1.78 | 0.82 | 2.30 | 1.11 |
| Finance | 61.11 | 2.14 | 0.43 | 29.49 | 1.28 | 0.00 | 0.00 | 2.14 | 3.42 |
| Information | 73.33 | 3.03 | 0.00 | 13.94 | 2.42 | 1.82 | 1.21 | 1.21 | 3.03 |
| Education | 76.58 | 0.63 | 0.00 | 14.56 | 3.80 | 1.27 | 0.00 | 3.16 | 0.00 |
| Professional | 58.22 | 8.22 | 0.00 | 17.81 | 0.00 | 4.79 | 0.00 | 0.68 | 0.00 |
| Retail | 46.46 | 0.00 | 1.01 | 42.42 | 4.04 | 1.01 | 0.00 | 3.03 | 2.02 |
| Manufacturing | 38.46 | 27.69 | 0.00 | 32.31 | 0.00 | 1.54 | 0.00 | 0.00 | 0.00 |
| Other Services | 57.45 | 2.13 | 0.00 | 27.66 | 6.38 | 0.00 | 0.00 | 2.13 | 4.26 |
| Administrative | 47.92 | 4.17 | 0.00 | 43.75 | 0.00 | 2.08 | 0.00 | 2.08 | 0.00 |
| Accommodation | 20.97 | 3.23 | 0.00 | 74.19 | 0.00 | 0.00 | 0.00 | 1.61 | 0.00 |
| Transportation | 44.12 | 5.88 | 0.00 | 44.12 | 5.88 | 0.00 | 0.00 | 0.00 | 0.00 |

**External Motives.** Table 6.21 shows the different external motives behind data breaches across industries. Our analysis reveals that *financial* gain emerges as the most common motive for attackers in *Healthcare* (96.24%), *Finance* (89.56%), and *Accommodation* (92.65%). While *unintentional* breaches remain relatively rare, *espionage* and other malicious motivations account for a significant portion of data breaches. *Healthcare*, *Finance*, and *Information* industries face the highest risks of *financial* gain-driven breaches, while *espionage* poses a more substantial threat to the *Administration* (27.92%) and *Manufacturing* (12.00%) sectors.

*Educationa*l and *Professional* industries are particularly susceptible to financially motivated attacks, while *Retail* and *Accommodation* industries are primarily targeted for financial gain. *Transportation* is the most likely industry to experience breaches motivated by *fun*, while *Other Services* and *Administrative* industries are susceptible to a broader range of motivations. Moreover, in sectors such as *Information* and *Retail*, motives such as *Fun* (17.66% in *Information*) and *Grudge* (0.99% in *Retail*) indicate that some breaches are driven by personal amusement or revenge. *Ideological* reasons are particularly significant in *Administration* (26.09%) and *Other Services* (41.38%), implying that some breaches in these sectors are motivated by beliefs or causes. Notably, breaches are classified as *Unintentional* in sectors such as *Healthcare* and *Administration*,

**Table 6.21:** The distribution (%) of external motives of attackers during data breaches across industries.

| | Unintentional | Espionage | Fear | Financial | Fun | Grudge | Ideology | Other |
|---|---|---|---|---|---|---|---|---|
| Healthcare | 0.60 | 0.30 | 0.15 | 96.24 | 1.20 | 0.75 | 0.45 | 0.30 |
| Administration | 1.46 | 27.92 | 0.18 | 34.12 | 5.84 | 3.47 | 26.09 | 0.91 |
| Finance | 0.00 | 2.20 | 0.27 | 89.56 | 1.65 | 0.82 | 5.22 | 0.27 |
| Information | 2.17 | 7.34 | 0.00 | 43.75 | 17.66 | 5.43 | 22.01 | 1.63 |
| Education | 0.00 | 1.10 | 0.00 | 76.92 | 9.89 | 3.30 | 7.69 | 1.10 |
| Professional | 0.00 | 9.49 | 0.00 | 68.99 | 12.66 | 2.53 | 6.33 | 0.00 |
| Retail | 0.33 | 0.00 | 0.00 | 95.38 | 1.98 | 0.99 | 1.32 | 0.00 |
| Manufacturing | 0.00 | 12.00 | 0.00 | 65.00 | 7.00 | 1.00 | 15.00 | 0.00 |
| Other Services | 0.00 | 3.45 | 1.15 | 43.68 | 3.45 | 5.75 | 41.38 | 1.15 |
| Administrative | 2.86 | 4.29 | 0.00 | 80.00 | 5.71 | 0.00 | 7.14 | 0.00 |
| Accommodation | 0.00 | 2.94 | 0.00 | 92.65 | 1.47 | 0.00 | 2.94 | 0.00 |
| Transportation | 6.52 | 4.35 | 0.00 | 67.39 | 15.22 | 4.35 | 2.17 | 0.00 |

highlighting that external attacks are not always maliciously intended.

Analyzing the motives behind data breaches is critical in understanding the ever-evolving cybersecurity landscape. By identifying the primary motivations driving these attacks, organizations gain valuable insights into the threat environment and prioritize their cybersecurity efforts accordingly.

**Takeaway:** Cybersecurity incident motives vary by sector: Unintentional actions in *Healthcare*, *Administration*, and *Education* call for improved training. Financial motives in *Accommodation*, *Finance*, and *Manufacturing* necessitate internal threat controls. Espionage is significant in *Manufacturing* and *Professional Services*, requiring enhanced confidentiality. Financial gain is the predominant external motive in *Healthcare*, *Finance*, and *Accommodation*.

### Actor Internal Variety

We analyzed the distribution of internal actors involved in cybersecurity incidents across various industries. Analyzing internal actors in cybersecurity incidents helps in targeted resource allocation, role-specific training, improved risk management, policy-making, proactive security mea-

sures, better incident response, and ensuring regulatory compliance. Our analysis revealed that end-users are the most frequently involved actors, with their percentages often being the highest.

For instance, in the *Healthcare* sector, end-users are implicated in 59.34% of incidents. System administrators stand out in the *Information* and *Professional* sectors, involved in 54.48% of incidents, indicating the crucial role of IT systems in these industries. The *Accommodation* industry sees cashiers at high risk, with a 59.57% involvement rate in incidents. In the *Finance* industry, while the '*Finance*' role is not the most involved, system administrators still face a substantial risk at 23.48%, and cashiers account for 10.61% of incidents.

Moreover, the *Retail* shows point-of-sale and IT *Administration* vulnerabilities, with cashier roles implicated in 20.31% and system admins in 21.88% of incidents. The variability of executive involvement in incidents is evident, with *Retail* notably at 10.94%, suggesting more significant consequences when their accounts are compromised. Roles such as auditors, guards, and helpdesk staff have minimal involvement, with low percentages across the board, which may indicate lower associated risks or less frequent detection of their involvement. The 'Other' category shows a wide variance, highlighting industry-specific roles not captured in standard classifications but still significant in cybersecurity events.

**Takeaway:** Internal actors in cybersecurity incidents differ across industries. End-users frequently contribute, highlighting targeted training needs. System administrators are key in IT sectors. In *Accommodation*, cashiers; in *Finance*, finance roles, and system admins; and in *Retail*, point-of-sale and IT admin are vulnerable. Executive involvement varies, signaling major potential consequences. Roles such as auditors and guards show minimal impact.

**Table 6.22:** The distribution (%) of internal variety for cybersecurity incidents in various industries.

| | Cashier | End-user | Executive | Finance | HR | Manager | Guard | Developer | System admin | Other |
|---|---|---|---|---|---|---|---|---|---|---|
| Healthcare | 0.00 | 59.34 | 5.30 | 6.06 | 0.51 | 2.78 | 0.51 | 2.02 | 5.81 | 17.68 |
| Administration | 0.30 | 41.67 | 5.95 | 2.08 | 3.27 | 6.25 | 3.57 | 5.95 | 11.90 | 19.05 |
| Finance | 10.61 | 25.76 | 2.27 | 6.82 | 0.00 | 5.30 | 0.00 | 7.58 | 23.48 | 18.18 |
| Information | 0.75 | 13.43 | 2.24 | 2.24 | 0.75 | 0.75 | 0.00 | 18.66 | 54.48 | 6.72 |
| Education | 0.00 | 36.63 | 3.96 | 3.96 | 5.94 | 6.93 | 0.00 | 2.97 | 23.76 | 15.84 |
| Professional | 0.75 | 13.43 | 2.24 | 2.24 | 0.75 | 0.75 | 0.00 | 18.66 | 54.48 | 6.72 |
| Retail | 20.31 | 18.75 | 10.94 | 3.13 | 3.13 | 4.69 | 0.00 | 3.13 | 21.88 | 14.06 |
| Manufacturing | 2.17 | 26.09 | 10.87 | 0.00 | 2.17 | 2.17 | 0.00 | 6.52 | 26.09 | 23.91 |
| Other Services | 8.33 | 37.50 | 8.33 | 0.00 | 0.00 | 4.17 | 4.17 | 4.17 | 12.50 | 20.83 |
| Administrative | 8.33 | 37.50 | 8.33 | 0.00 | 0.00 | 4.17 | 4.17 | 4.17 | 12.50 | 20.83 |
| Accommodation | 59.57 | 4.26 | 4.26 | 2.13 | 2.13 | 8.51 | 0.00 | 4.26 | 4.26 | 10.64 |
| Transportation | 3.85 | 46.15 | 7.69 | 0.00 | 3.85 | 0.00 | 0.00 | 3.85 | 7.69 | 26.92 |

## Types of Attackers

In the evolving landscape of cybersecurity, understanding the profiles of attackers involved in data breaches is crucial for developing effective defense strategies. To this end, we conducted an analysis to identify the most prevalent types of attackers perpetrating data breaches across various industries. We aim to provide a clearer understanding of the threat actors essential for tailoring security measures and policies.

Table 6.23 shed light on the varying nature of attackers in data breaches across different industries, with distinct patterns emerging in each sector. For instance, in the *Healthcare* sector, Organized Crime stands out as the primary type of attacker, accounting for 42.86%, suggesting a significant threat from well-structured and sophisticated cybercriminal groups. This is closely followed by Unaffiliated individuals, who represent 26.25% of the attackers. However, The *Administration* sector faces a different challenge, with Activists being the leading threat at 31.49%, indicating a prevalence of ideologically driven cyber threats, complemented by a diverse array of attackers categorized as 'Other' (46.38%).

Organized Crime poses the most significant risk in the *Finance* sector, such as *Healthcare*, con-

**Table 6.23:** The distribution (%) of the different types of attackers during data breach incidents across different industries.

| | Org. Crime | Unaffiliated | Activist | Ex-Employee | Customer | Other |
|---|---|---|---|---|---|---|
| Healthcare | 42.86 | 26.25 | 1.00 | 5.32 | 0.66 | 23.92 |
| Administration | 9.57 | 11.49 | 31.49 | 0.43 | 0.64 | 46.38 |
| Finance | 44.10 | 15.72 | 8.30 | 3.93 | 0.87 | 27.07 |
| Information | 15.99 | 25.00 | 29.94 | 2.33 | 1.45 | 25.29 |
| Education | 36.11 | 13.19 | 15.28 | 1.39 | 0.69 | 33.33 |
| Professional | 30.08 | 14.63 | 9.76 | 4.07 | 1.63 | 39.84 |
| Retail | 67.57 | 12.16 | 3.38 | 0.68 | 2.03 | 14.19 |
| Manufacturing | 23.08 | 13.85 | 27.69 | 4.62 | 0.00 | 30.77 |
| Other Services | 10.67 | 14.67 | 42.67 | 2.67 | 1.33 | 28.00 |
| Administrative | 45.45 | 25.00 | 6.82 | 4.55 | 0.00 | 18.18 |
| Accommodation | 47.37 | 13.16 | 5.26 | 5.26 | 0.00 | 28.95 |
| Transportation | 34.15 | 26.83 | 4.88 | 4.88 | 0.00 | 29.27 |

stituting 44.10% of the breaches, highlighting a solid financial motive behind these attacks. The *Information* sector presents a varied landscape, with Activists (29.94%) and Unaffiliated individuals (25.00%) being the most common attackers, necessitating broad security strategies. Organized Crime predominantly targets *Retail*, evident in its 67.57% share, underscoring the sector's vulnerability to financially motivated attacks. *Manufacturing* and *Other Services* sectors reveal a notable prevalence of attacks from Activists, at 27.69% and 42.67%, respectively, suggesting these sectors are more prone to breaches motivated by protest or ideological reasons. Each industry's unique pattern of attackers underscores the need for tailored cybersecurity strategies, addressing the specific risks and attacker profiles predominant in their respective fields.

**Takeaway:** *Healthcare* faces a significant threat from *organized crime*, while the *Administration* sector is prone to *ideologically-driven attacks*, primarily by activists. *Finance* is at risk from financially motivated *organized crime*, and *Retail* is highly vulnerable to such attacks. *Manufacturing* and *Other Services* sectors show a prevalence of attacks from Activists, indicating exposure to breaches driven by ideological motives.

Timeline Discovery

In examining the timeline for the discovery of data breaches across various industries, as presented in Table 6.24, distinct patterns emerge, underscoring the critical importance of timely detection. The *Healthcare* sector predominantly identifies breaches within months (46.71%) and years (17.16%), suggesting a delay possibly due to complex data systems. *Administration* displays a varied detection timeline with significant discoveries over the years (21.10%), hours (20.45%), and days (18.18%), indicating diverse capabilities. In *Finance*, breaches are often detected within months (32.16%) and days (24.31%), reflecting quicker response mechanisms likely necessitated by the immediacy of financial transactions.

The *Information* sector shows a balanced detection approach, leaning slightly towards hours (22.78%) and days (21.67%), indicative of proactive monitoring. *Education* tends to discover breaches mostly in months (34.56%), with a fair number detected within hours (22.79%) and days (11.76%), pointing to a mix of detection timelines. *Retail*, *Manufacturing*, and *Other Services* predominantly identify breaches within days to months, suggesting varying levels of monitoring effectiveness. *Accommodation* and *Transportation*, with the highest breach discovery rates in months (50.00% and 22.73%, respectively) and years (11.67% and 22.73%), highlight concerning delays in detection. Our analysis across sectors reveals a need for tailored data breach detection and response strategies, especially in the *Healthcare* and *Accommodation* sectors where delays are prominent. The

**Table 6.24:** The distribution (%) of the timeline discovery periods for data breaches across industries. Abbreviations: Seconds (Sec), Minutes (Min), Hours (Hr), Days (Day), Weeks (Wk), Months (Mo), Years (Yr).

| | Sec | Min | Hr | Day | Wk | Mo | Yr |
|---|---|---|---|---|---|---|---|
| Healthcare | 0.00 | 4.26 | 9.81 | 16.13 | 5.94 | 46.71 | 17.16 |
| Administration | 2.92 | 8.44 | 20.45 | 18.18 | 4.55 | 24.35 | 21.10 |
| Finance | 0.78 | 6.27 | 15.69 | 24.31 | 8.24 | 32.16 | 12.55 |
| Information | 5.00 | 9.44 | 22.78 | 21.67 | 10.00 | 18.33 | 12.78 |
| Education | 2.21 | 12.50 | 22.79 | 11.76 | 11.03 | 34.56 | 5.15 |
| Professional | 1.02 | 9.18 | 15.31 | 20.41 | 6.12 | 33.67 | 14.29 |
| Retail | 1.44 | 5.04 | 15.83 | 19.42 | 10.79 | 37.41 | 10.07 |
| Manufacturing | 2.27 | 2.27 | 20.45 | 27.27 | 6.82 | 29.55 | 11.36 |
| Other Services | 2.63 | 10.53 | 26.32 | 18.42 | 7.89 | 23.68 | 10.53 |
| Administrative | 0.00 | 4.35 | 8.70 | 13.04 | 10.87 | 47.83 | 15.22 |
| Accommodation | 0.00 | 3.33 | 10.00 | 16.67 | 8.33 | 50.00 | 11.67 |
| Transportation | 13.64 | 13.64 | 4.55 | 0.00 | 22.73 | 22.73 | 22.73 |

findings emphasize the necessity of continual monitoring and rapid response systems to minimize the impact of data breaches.

**Takeaway:** Data breach discovery times vary by industry. *Healthcare* typically takes months (46.71%), whereas *Administration* ranges from years (21.10%) to hours (20.45%). *Finance* often detects breaches within months (32.16%). The *Information* sector balances detection between hours (22.78%) and days (21.67%). This highlights the necessity of industry-specific, continuous monitoring strategies, particularly in *Healthcare* and *Accommodation*.

## Summary and Concluding Remark

Our examination of data breaches across critical sectors reveals concerning insights into digital vulnerability. Notably, the *Finance* and *Healthcare* sectors experience severe consequences, including quantifiable financial losses and reputational damage. Addressing these issues requires

rigorous security protocols, enhanced regulations, and specialized cybersecurity strategies. Future research should prioritize prevention through advanced detection and proactive risk management to ensure a secure digital future. Ignoring these complexities exposes us to ongoing threats in the digital age, emphasizing the importance of acknowledging the lasting impact of data breaches and working towards a secure and trustworthy future.

# CHAPTER 7: DISCUSSION AND RECOMMENDATION

This dissertation presents a thorough investigation into the rising prevalence of data breaches in the healthcare sector, exploring the issue from various angles through a series of comprehensive studies. Today, we live in a world increasingly dependent on digital technology, and protecting sensitive healthcare data has never been more critical. Toward enhancing the protection of sensitive data, we shed light on the multifaceted nature of healthcare data breaches, their causes, impacts, and the inherent threats they pose. Each work within this dissertation contributes a unique perspective, collectively constructing a holistic understanding of the issue. The first study delves into the surge of data breaches during the COVID-19 pandemic, highlighting their increasing severity and the prolonged duration before they are discovered. The second research study focuses on the geographical variance in data breaches, noting an exceptionally high incidence in the United States. The third study introduces a novel approach by correlating the security attributes of hospital websites with their susceptibility to breaches, uncovering significant security oversights. In the fourth work, we conduct a comparative study across 12 main sectors, examining the patterns, similarities, and differences in data breaches. Our goal is not only to present a detailed analysis of the current state of healthcare data security but also to understand the broader implications these breaches have on the healthcare system, patient trust, and overall data privacy. Our ultimate goal of these findings is to contribute to the discourse on healthcare cybersecurity, offering critical insights for stakeholders across the healthcare and technology sectors.

## Recommendations

In light of this dissertation's findings, it becomes imperative to offer targeted recommendations that address the identified challenges in healthcare data security. Below, we list recommendations in-

tended to guide healthcare providers, policymakers, technology developers, and the general public toward a more secure and trustworthy healthcare system.

**Enhanced Detection and Reporting Systems.** One main observation is that the decrease in incident reports post-2013 may not reflect a reduction in cybersecurity threats but rather indicate potential gaps in incident detection and reporting mechanisms. This underscores the urgent need for enhanced systems in these areas. It is crucial to implement more sophisticated detection and reporting systems. These systems must accurately identify and document security incidents, providing a realistic view of an organization's cybersecurity landscape. By improving detection capabilities, organizations identify breaches promptly, allowing for quicker response and mitigation efforts. Simultaneously, enhancing reporting mechanisms will ensure that all incidents are accurately captured and reported, facilitating better analysis and understanding of cyber threats. Implementing these improved systems is crucial for providing a clear and accurate representation of an organization's cybersecurity health, enabling more effective management of cyber risks.

**Comprehensive Security Strategy for Healthcare.** Given the critical findings that medical and personal data are prime targets in 91% of breaches and that incident discovery can span several years, there is a pressing need for a comprehensive security strategy in healthcare. This strategy must prioritize protecting sensitive information against identity theft and privacy violations by implementing advanced data protection protocols. The prolonged periods before incident discovery underscore a fundamental weakness in the current security measures within healthcare organizations. To address this, it is crucial to adopt sophisticated security solutions that safeguard data and enhance the detection of breaches. Additionally, providing cybersecurity training for all healthcare personnel and developing effective incident response protocols are essential steps. These efforts will expedite the identification of security incidents and ensure robust protection of patient information, significantly mitigating the risk of breaches. This comprehensive approach to cyber-

security in healthcare is crucial for safeguarding against the evolving landscape of cyber threats and vulnerabilities.

**Encryption of Sensitive Data.** The significant finding that a substantial portion of breached data was either stored or transmitted without encryption, coupled with over 20% of incidents involving attacks on integrity and authenticity, highlights a crucial vulnerability in data security practices. We strongly recommend mandating encryption for all sensitive data in storage and during transmission to counteract this vulnerability. This measure will be a critical defense mechanism against unauthorized access and exploitation. Furthermore, it is essential to strengthen security protocols to address the substantial risk posed by attacks on data integrity and authenticity. Enhancing these protocols will ensure that data remains protected from external breaches and internal threats that compromise data without it leaving the organization. Implementing these recommendations will significantly reduce the risk of exploitation and uphold the integrity and confidentiality of sensitive information, thereby fortifying the organization's cybersecurity framework.

**Web Presence and Security for Hospitals.** We found a growth in proprietary hospitals, and their comparatively higher adoption of DNSSEC underscores an opportunity for enhancing web security practices across all types of hospitals. This difference reveals a more comprehensive need for improved web security measures, including adopting secure domain name practices and transitioning away from plain HTTP protocol. Moreover, the finding that many hospital websites depend on unsecured HTTP protocol and maintain invalid SSL certificates further emphasizes the critical need for standardizing HTTPS and ensuring SSL certificates' validity across hospital websites. Additionally, the presence of malicious features and vulnerabilities in outdated hospital websites highlights the urgent requirement for regular website maintenance and comprehensive security audits. The aforementioned measures are essential for mitigating data leakage risks and safeguarding patient information against cyber threats. Implementing these recommendations will

significantly elevate the security posture of hospital websites, providing a safer and more reliable online environment for patients and healthcare providers.

**Cybersecurity Tailored to Industry Specifics.** The variety in data breach characteristics across different industries, notably the healthcare sector's vulnerability to opportunistic breaches and the primary role of internal actors in cybersecurity incidents, highlights the critical need for tailored cybersecurity strategies. This need is further emphasized by the risks faced by sectors like *Healthcare* and *Accommodation*, which are often targets for cyber threats due to the sensitive nature of the information they maintain. To address these sector-specific vulnerabilities effectively, developing and implementing cybersecurity strategies customized to each industry's unique requirements and challenges is essential. These strategies should account for external threats and significantly focus on mitigating risks internal actors pose through targeted training, access controls, and monitoring. Furthermore, implementing continuous monitoring systems is crucial for the early detection of potential security breaches, enabling timely responses to mitigate damages. By adopting an approach that is both tailored to the specific needs of each industry and vigilant in its monitoring for threats, organizations can significantly improve their cybersecurity posture, protecting against both opportunistic and sophisticated cyber attacks.

# CHAPTER 8: CONCLUSION

In this dissertation, we explored four key areas related to data breaches and security in healthcare organizations. First, we conducted a long-term analysis of data breaches, revealing a significant increase in incidents from 2010 to 2020. Our study identified various attack types, including internal, external, and partner attacks, with financial motives being the primary driving factor. The high cost associated with data breaches, along with the classification of threat actor actions, provides valuable insights into the evolving threat landscape. Moreover, the victims include different types, but the primary victims are the patients, followed by the customers and the employees. Ransomware motivated 82% of malware threat actors, and 45% of malware threat actions are directly installed. Second, we investigated the influence of population factors on data breach exposure, highlighting the correlation between the number of adults, state population, and incident rates. The study also uncovers the most breached assets and provides a timeline analysis, emphasizing the importance of prompt incident discovery. Third, we focused on the hospitals' online presence and website security. We examined SSL certificate validity, associated issues, and malicious behaviors. The findings shed light on the vulnerabilities and lax security practices observed in many hospital websites, emphasizing the need for improved security measures. Fourth, we conducted a comparative sector-based study investigating data breaches across various critical sectors. This broader view provides a contextual understanding of the healthcare sector's unique vulnerabilities in comparison to other sectors. Overall, this dissertation provides a comprehensive understanding of data breaches and security in healthcare organizations. It offers insights into trends, risks, and mitigation strategies, guiding future research to enhance cybersecurity in the healthcare sector.

# APPENDIX A: PUBLICATIONS COPYRIGHT

# Consent to Publish

## Lecture Notes in Computer Science

**Title of the Book or Conference Name:** WISA 2021

**Volume Editor(s) Name(s):** Hyoungshick Kim

**Title of the Contribution:** Measuring Healthcare Data Breaches

**Author(s) Full Name(s):** Mohammed Alkinoon, Sung J. Choi, David Mohaisen

**Corresponding Author's Name, Affiliation Address, and Email:**

David Mohaisen, University of Central Florida, Orlando, FL, mohaisen@ucf.edu

When Author is more than one person the expression "Author" as used in this agreement will apply collectively unless otherwise indicated.

The Publisher intends to publish the Work under the imprint **Springer**. The Work may be published in the book series **Lecture Notes in Computer Science (LNCS, LNAI or LNBI)**.

### § 1 Rights Granted

Author hereby grants and assigns to **Springer Nature Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland** (hereinafter called **Publisher**) the exclusive, sole, permanent, world-wide, transferable, sub-licensable and unlimited right to reproduce, publish, distribute, transmit, make available or otherwise communicate to the public, translate, publicly perform, archive, store, lease or lend and sell the Contribution or parts thereof individually or together with other works in any language, in all revisions and versions (including soft cover, book club and collected editions, anthologies, advance printing, reprints or print to order, microfilm editions, audiograms and videograms), in all forms and media of expression including in electronic form (including offline and online use, push or pull technologies, use in databases and data networks (e.g. the Internet) for display, print and storing on any and all stationary or portable end-user devices, e.g. text readers, audio, video or interactive devices, and for use in multimedia or interactive versions as well as for the display or transmission of the Contribution or parts thereof in data networks or search engines, and posting the Contribution on social media accounts closely related to the Work), in whole, in part or in abridged form, in each case as now known or developed in the future, including the right to grant further time-limited or permanent rights. Publisher especially has the right to permit others to use individual illustrations, tables or text quotations and may use the Contribution for advertising purposes. For the purposes of use in electronic forms, Publisher may adjust the Contribution to the respective form of use and include links (e.g. frames or inline-links) or otherwise combine it with other works and/or remove links or combinations with other works provided in the Contribution. For the avoidance of doubt, all provisions of this contract apply regardless of whether the Contribution and/or the Work itself constitutes a database under applicable copyright laws or not.

The copyright in the Contribution shall be vested in the name of Publisher. Author has asserted his/her right(s) to be identified as the originator of this Contribution in all editions and versions of the Work and parts thereof, published in all forms and media. Publisher may take, either in its own name or in that of Author, any necessary steps to protect the rights granted under this Agreement against infringement by third parties. It will have a copyright notice inserted into all editions of the Work and on the Contribution according to the provisions of the Universal Copyright Convention (UCC).

The parties acknowledge that there may be no basis for claim of copyright in the United States to a Contribution prepared by an officer or employee of the United States government as part of that person's official duties. If the Contribution was performed under a United States government contract, but Author is not a United States government employee, Publisher grants the United States government royalty-free permission to reproduce all or part of the Contribution and to authorise others to do so for United States government purposes. If the Contribution was prepared or published by or under the direction or control of the Crown (i.e., the constitutional monarch of the Commonwealth realm) or any Crown government department, the copyright in the Contribution shall, subject to any

agreement with Author, belong to the Crown. If Author is an officer or employee of the United States government or of the Crown, reference will be made to this status on the signature page.

## § 2 Rights Retained by Author

Author retains, in addition to uses permitted by law, the right to communicate the content of the Contribution to other research colleagues, to share the Contribution with them in manuscript form, to perform or present the Contribution or to use the content for non-commercial internal and educational purposes, provided the original source of publication is cited according to the current citation standards in any printed or electronic materials. Author retains the right to republish the Contribution in any collection consisting solely of Author's own works without charge, subject to ensuring that the publication of the Publisher is properly credited and that the relevant copyright notice is repeated verbatim. Author may self-archive an author-created version of his/her Contribution on his/her own website and/or the repository of Author's department or faculty. Author may also deposit this version on his/her funder's or funder's designated repository at the funder's request or as a result of a legal obligation. He/she may not use the Publisher's PDF version, which is posted on the Publisher's platforms, for the purpose of self-archiving or deposit. Furthermore, Author may only post his/her own version, provided acknowledgment is given to the original source of publication and a link is inserted to the published article on the Publisher's website. The link must be provided by inserting the DOI number of the article in the following sentence: "The final authenticated version is available online at https://doi.org/[insert DOI]." The DOI (Digital Object Identifier) can be found at the bottom of the first page of the published paper.
Prior versions of the Contribution published on non-commercial pre-print servers like ArXiv/CoRR and HAL can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgment needs to be given to the final publication and a link must be inserted to the published Contribution on the Publisher's website, by inserting the DOI number of the article in the following sentence: "The final authenticated publication is available online at https://doi.org/[insert DOI]".
Author retains the right to use his/her Contribution for his/her further scientific career by including the final published paper in his/her dissertation or doctoral thesis provided acknowledgment is given to the original source of publication. Author also retains the right to use, without having to pay a fee and without having to inform the Publisher, parts of the Contribution (e.g. illustrations) for inclusion in future work. Authors may publish an extended version of their proceedings paper as a journal article provided the following principles are adhered to: a) the extended version includes at least 30% new material, b) the original publication is cited, and c) it includes an explicit statement about the increment (e.g., new results, better description of materials, etc.).

## § 3 Warranties

Author agrees, at the request of Publisher, to execute all documents and do all things reasonably required by Publisher in order to confer to Publisher all rights intended to be granted under this Agreement. Author warrants that the Contribution is original except for such excerpts from copyrighted works (including illustrations, tables, animations and text quotations) as may be included with the permission of the copyright holder thereof, in which case(s) Author is required to obtain written permission to the extent necessary and to indicate the precise sources of the excerpts in the manuscript. Author is also requested to store the signed permission forms and to make them available to Publisher if required.
Author warrants that Author is entitled to grant the rights in accordance with Clause 1 "Rights Granted", that Author has not assigned such rights to third parties, that the Contribution has not heretofore been published in whole or in part, that the Contribution contains no libellous or defamatory statements and does not infringe on any copyright, trademark, patent, statutory right or proprietary right of others, including rights obtained through licences. Author agrees to amend the Contribution to remove any potential obscenity, defamation, libel, malicious falsehood or otherwise unlawful part(s) identified at any time. Any such removal or alteration shall not affect the warranty given by Author in this Agreement.

## § 4 Delivery of Contribution and Publication

Author agrees to deliver to the responsible Volume Editor (for conferences, usually one of the Program Chairs), on a date to be agreed upon, the manuscript created according to the Publisher's Instructions for Authors. Publisher will undertake the reproduction and distribution of the Contribution at its own expense and risk. After submission of the Consent to Publish form signed by the Corresponding Author, changes of authorship, or in the order of the authors listed, will not be accepted by the Publisher.

### § 5 Author's Discount for Books

Author is entitled to purchase for his/her personal use (if ordered directly from Publisher) the Work or other books published by Publisher at a discount of 40% off the list price for as long as there is a contractual arrangement between Author and Publisher and subject to applicable book price regulation.
Resale of such copies is not permitted.

### § 6 Governing Law and Jurisdiction

If any difference shall arise between Author and Publisher concerning the meaning of this Agreement or the rights and liabilities of the parties, the parties shall engage in good faith discussions to attempt to seek a mutually satisfactory resolution of the dispute. This agreement shall be governed by, and shall be construed in accordance with, the laws of Switzerland. The courts of Zug, Switzerland shall have the exclusive jurisdiction.

Corresponding Author signs for and accepts responsibility for releasing this material on behalf of any and all Co-Authors.

**Signature of Corresponding Author:** **Date:**

8/25/2021

☐ I'm an employee of the US Government and transfer the rights to the extent transferable (Title 17 §105 U.S.C. applies)

☐ I'm an employee of the Crown and copyright on the Contribution belongs to the Crown

*For internal use only:*
Legal Entity Number: 1128 Springer Nature Switzerland AG
Springer-C-CTP-07/2018

# Consent to Publish

## Lecture Notes in Computer Science

**Title of the Book or Conference Name:** The 10th International Conference on Computational Data & Social Networks

**Volume Editor(s) Name(s):** David Mohaisen and Ruoming Jin

**Title of the Contribution:** Security Breaches in the Healthcare Domain: A Spatiotemporal Analysis

**Author(s) Full Name(s):** Mohammed Al Kinoon, Marwan Omar, Manar Mohaisen, David Mohaisen

**Corresponding Author's Name, Affiliation Address, and Email:**

Mohammed Al Kinoon, University of Central Florida, Orlando, FL, malkinoon@knights.ucf.edu

When Author is more than one person the expression "Author" as used in this agreement will apply collectively unless otherwise indicated.

The Publisher intends to publish the Work under the imprint **Springer**. The Work may be published in the book series **Lecture Notes in Computer Science (LNCS, LNAI or LNBI)**.

### § 1 Rights Granted

Author hereby grants and assigns to **Springer Nature Switzerland AG, Gewerbestrasse 11, 6330 Cham, Switzerland** (hereinafter called **Publisher**) the exclusive, sole, permanent, world-wide, transferable, sub-licensable and unlimited right to reproduce, publish, distribute, transmit, make available or otherwise communicate to the public, translate, publicly perform, archive, store, lease or lend and sell the Contribution or parts thereof individually or together with other works in any language, in all revisions and versions (including soft cover, book club and collected editions, anthologies, advance printing, reprints or print to order, microfilm editions, audiograms and videograms), in all forms and media of expression including in electronic form (including offline and online use, push or pull technologies, use in databases and data networks (e.g. the Internet) for display, print and storing on any and all stationary or portable end-user devices, e.g. text readers, audio, video or interactive devices, and for use in multimedia or interactive versions as well as for the display or transmission of the Contribution or parts thereof in data networks or search engines, and posting the Contribution on social media accounts closely related to the Work), in whole, in part or in abridged form, in each case as now known or developed in the future, including the right to grant further time-limited or permanent rights. Publisher especially has the right to permit others to use individual illustrations, tables or text quotations and may use the Contribution for advertising purposes. For the purposes of use in electronic forms, Publisher may adjust the Contribution to the respective form of use and include links (e.g. frames or inline-links) or otherwise combine it with other works and/or remove links or combinations with other works provided in the Contribution. For the avoidance of doubt, all provisions of this contract apply regardless of whether the Contribution and/or the Work itself constitutes a database under applicable copyright laws or not.

The copyright in the Contribution shall be vested in the name of Publisher. Author has asserted his/her right(s) to be identified as the originator of this Contribution in all editions and versions of the Work and parts thereof, published in all forms and media. Publisher may take, either in its own name or in that of Author, any necessary steps to protect the rights granted under this Agreement against infringement by third parties. It will have a copyright notice inserted into all editions of the Work and on the Contribution according to the provisions of the Universal Copyright Convention (UCC).

The parties acknowledge that there may be no basis for claim of copyright in the United States to a Contribution prepared by an officer or employee of the United States government as part of that person's official duties. If the Contribution was performed under a United States government contract, but Author is not a United States government employee, Publisher grants the United States government royalty-free permission to reproduce all or part of the Contribution and to authorise others to do so for United States government purposes. If the Contribution was prepared or published by or under the direction or control of the Crown (i.e., the constitutional monarch of the Commonwealth realm) or any Crown government department, the copyright in the Contribution shall, subject to any

agreement with Author, belong to the Crown. If Author is an officer or employee of the United States government or of the Crown, reference will be made to this status on the signature page.

## § 2 Rights Retained by Author

Author retains, in addition to uses permitted by law, the right to communicate the content of the Contribution to other research colleagues, to share the Contribution with them in manuscript form, to perform or present the Contribution or to use the content for non-commercial internal and educational purposes, provided the original source of publication is cited according to the current citation standards in any printed or electronic materials. Author retains the right to republish the Contribution in any collection consisting solely of Author's own works without charge, subject to ensuring that the publication of the Publisher is properly credited and that the relevant copyright notice is repeated verbatim. Author may self-archive an author-created version of his/her Contribution on his/her own website and/or the repository of Author's department or faculty. Author may also deposit this version on his/her funder's or funder's designated repository at the funder's request or as a result of a legal obligation. He/she may not use the Publisher's PDF version, which is posted on the Publisher's platforms, for the purpose of self-archiving or deposit. Furthermore, Author may only post his/her own version, provided acknowledgment is given to the original source of publication and a link is inserted to the published article on the Publisher's website. The link must be provided by inserting the DOI number of the article in the following sentence: "The final authenticated version is available online at https://doi.org/[insert DOI]." The DOI (Digital Object Identifier) can be found at the bottom of the first page of the published paper.

Prior versions of the Contribution published on non-commercial pre-print servers like ArXiv/CoRR and HAL can remain on these servers and/or can be updated with Author's accepted version. The final published version (in pdf or html/xml format) cannot be used for this purpose. Acknowledgment needs to be given to the final publication and a link must be inserted to the published Contribution on the Publisher's website, by inserting the DOI number of the article in the following sentence: "The final authenticated publication is available online at https://doi.org/[insert DOI]".

Author retains the right to use his/her Contribution for his/her further scientific career by including the final published paper in his/her dissertation or doctoral thesis provided acknowledgment is given to the original source of publication. Author also retains the right to use, without having to pay a fee and without having to inform the Publisher, parts of the Contribution (e.g. illustrations) for inclusion in future work. Authors may publish an extended version of their proceedings paper as a journal article provided the following principles are adhered to: a) the extended version includes at least 30% new material, b) the original publication is cited, and c) it includes an explicit statement about the increment (e.g., new results, better description of materials, etc.).

## § 3 Warranties

Author agrees, at the request of Publisher, to execute all documents and do all things reasonably required by Publisher in order to confer to Publisher all rights intended to be granted under this Agreement. Author warrants that the Contribution is original except for such excerpts from copyrighted works (including illustrations, tables, animations and text quotations) as may be included with the permission of the copyright holder thereof, in which case(s) Author is required to obtain written permission to the extent necessary and to indicate the precise sources of the excerpts in the manuscript. Author is also requested to store the signed permission forms and to make them available to Publisher if required.

Author warrants that Author is entitled to grant the rights in accordance with Clause 1 "Rights Granted", that Author has not assigned such rights to third parties, that the Contribution has not heretofore been published in whole or in part, that the Contribution contains no libellous or defamatory statements and does not infringe on any copyright, trademark, patent, statutory right or proprietary right of others, including rights obtained through licences. Author agrees to amend the Contribution to remove any potential obscenity, defamation, libel, malicious falsehood or otherwise unlawful part(s) identified at any time. Any such removal or alteration shall not affect the warranty given by Author in this Agreement.

## § 4 Delivery of Contribution and Publication

Author agrees to deliver to the responsible Volume Editor (for conferences, usually one of the Program Chairs), on a date to be agreed upon, the manuscript created according to the Publisher's Instructions for Authors. Publisher will undertake the reproduction and distribution of the Contribution at its own expense and risk. After submission of the Consent to Publish form signed by the Corresponding Author, changes of authorship, or in the order of the authors listed, will not be accepted by the Publisher.

### § 5 Author's Discount for Books

Author is entitled to purchase for his/her personal use (if ordered directly from Publisher) the Work or other books published by Publisher at a discount of 40% off the list price for as long as there is a contractual arrangement between Author and Publisher and subject to applicable book price regulation.
Resale of such copies is not permitted.

### § 6 Governing Law and Jurisdiction

If any difference shall arise between Author and Publisher concerning the meaning of this Agreement or the rights and liabilities of the parties, the parties shall engage in good faith discussions to attempt to seek a mutually satisfactory resolution of the dispute. This agreement shall be governed by, and shall be construed in accordance with, the laws of Switzerland. The courts of Zug, Switzerland shall have the exclusive jurisdiction.

Corresponding Author signs for and accepts responsibility for releasing this material on behalf of any and all Co-Authors.

**Signature of Corresponding Author:** **Date:**

Mohammed Alkinoon ..... 10/01/2021

☐ I'm an employee of the US Government and transfer the rights to the extent transferable (Title 17 §105 U.S.C. applies)

☐ I'm an employee of the Crown and copyright on the Contribution belongs to the Crown

*For internal use only:*
Legal Entity Number: 1128 Springer Nature Switzerland AG
Springer-C-CTP-07/2018

# IEEE COPYRIGHT AND CONSENT FORM

To ensure uniformity of treatment among all contributors, other forms may not be substituted for this form, nor may any wording of the form be changed. This form is intended for original material submitted to the IEEE and must accompany any such material in order to be published by the IEEE. Please read the form carefully and keep a copy for your files.

**Understanding the Security and Performance of the Web Presence of Hospitals: A Measurement Study**

**Mohammed Al kinoon,Abdulrahman Alabduljabbar,Hattan Althebeiti,Rhongho Jang,DaeHun Nyang,David Mohaisen**

**2023 32nd International Conference on Computer Communications and Networks (ICCCN)**

## COPYRIGHT TRANSFER

The undersigned hereby assigns to The Institute of Electrical and Electronics Engineers, Incorporated (the "IEEE") all rights under copyright that may exist in and to: (a) the Work, including any revised or expanded derivative works submitted to the IEEE by the undersigned based on the Work; and (b) any associated written or multimedia components or other enhancements accompanying the Work.

## GENERAL TERMS

1. The undersigned represents that he/she has the power and authority to make and execute this form.
2. The undersigned agrees to indemnify and hold harmless the IEEE from any damage or expense that may arise in the event of a breach of any of the warranties set forth above.
3. The undersigned agrees that publication with IEEE is subject to the policies and procedures of the IEEE PSPB Operations Manual.
4. In the event the above work is not accepted and published by the IEEE or is withdrawn by the author(s) before acceptance by the IEEE, the foregoing copyright transfer shall be null and void. In this case, IEEE will retain a copy of the manuscript for internal administrative/record-keeping purposes.
5. For jointly authored Works, all joint authors should sign, or one of the authors should sign as authorized agent for the others.
6. The author hereby warrants that the Work and Presentation (collectively, the "Materials") are original and that he/she is the author of the Materials. To the extent the Materials incorporate text passages, figures, data or other material from the works of others, the author has obtained any necessary permissions. Where necessary, the author has obtained all third party permissions and consents to grant the license above and has provided copies of such permissions and consents to IEEE

**You have indicated that you DO wish to have video/audio recordings made of your conference presentation under terms and conditions set forth in "Consent and Release."**

## CONSENT AND RELEASE

1. In the event the author makes a presentation based upon the Work at a conference hosted or sponsored in whole or in part by the IEEE, the author, in consideration for his/her participation in the conference, hereby grants the IEEE the unlimited, worldwide, irrevocable permission to use, distribute, publish, license, exhibit, record, digitize, broadcast, reproduce and archive, in any format or medium, whether now known or hereafter developed: (a) his/her presentation and comments at the conference; (b) any written materials or multimedia files used in connection with his/her presentation; and (c) any recorded interviews of him/her (collectively, the "Presentation"). The permission granted includes the transcription and reproduction of the Presentation for inclusion in products sold or distributed by IEEE and live or recorded broadcast of the Presentation during or after the conference.
2. In connection with the permission granted in Section 1, the author hereby grants IEEE the unlimited, worldwide, irrevocable right to use his/her name, picture, likeness, voice and biographical information as part of the advertisement, distribution and sale of products incorporating the Work or Presentation, and releases IEEE from any claim based on right of privacy or publicity.

BY TYPING IN YOUR FULL NAME BELOW AND CLICKING THE SUBMIT BUTTON, YOU CERTIFY THAT SUCH ACTION CONSTITUTES

YOUR ELECTRONIC SIGNATURE TO THIS FORM IN ACCORDANCE WITH UNITED STATES LAW, WHICH AUTHORIZES ELECTRONIC SIGNATURE BY AUTHENTICATED REQUEST FROM A USER OVER THE INTERNET AS A VALID SUBSTITUTE FOR A WRITTEN SIGNATURE.

Mohammed Al kinoon

Signature

10-05-2023

Date (dd-mm-yyyy)

# Information for Authors

## AUTHOR RESPONSIBILITIES

The IEEE distributes its technical publications throughout the world and wants to ensure that the material submitted to its publications is properly available to the readership of those publications. Authors must ensure that their Work meets the requirements as stated in section 8.2.1 of the IEEE PSPB Operations Manual, including provisions covering originality, authorship, author responsibilities and author misconduct. More information on IEEE's publishing policies may be found at http://www.ieee.org/publications_standards/publications/rights/authorrightsresponsibilities.html Authors are advised especially of IEEE PSPB Operations Manual section 8.2.1.B12: "It is the responsibility of the authors, not the IEEE, to determine whether disclosure of their material requires the prior consent of other parties and, if so, to obtain it." Authors are also advised of IEEE PSPB Operations Manual section 8.1.1B: "Statements and opinions given in work published by the IEEE are the expression of the authors."

## RETAINED RIGHTS/TERMS AND CONDITIONS

- Authors/employers retain all proprietary rights in any process, procedure, or article of manufacture described in the Work.
- Authors/employers may reproduce or authorize others to reproduce the Work, material extracted verbatim from the Work, or derivative works for the author's personal use or for company use, provided that the source and the IEEE copyright notice are indicated, the copies are not used in any way that implies IEEE endorsement of a product or service of any employer, and the copies themselves are not offered for sale.
- Although authors are permitted to re-use all or portions of the Work in other works, this does not include granting third-party requests for reprinting, republishing, or other types of re-use.The IEEE Intellectual Property Rights office must handle all such third-party requests.
- Authors whose work was performed under a grant from a government funding agency are free to fulfill any deposit mandates from that funding agency.

## AUTHOR ONLINE USE

- **Personal Servers**. Authors and/or their employers shall have the right to post the accepted version of IEEE-copyrighted articles on their own personal servers or the servers of their institutions or employers without permission from IEEE, provided that the posted version includes a prominently displayed IEEE copyright notice and, when published, a full citation to the original IEEE publication, including a link to the article abstract in IEEE Xplore. Authors shall not post the final, published versions of their papers.
- **Classroom or Internal Training Use.** An author is expressly permitted to post any portion of the accepted version of his/her own IEEE-copyrighted articles on the author's personal web site or the servers of the author's institution or company in connection with the author's teaching, training, or work responsibilities, provided that the appropriate copyright, credit, and reuse notices appear prominently with the posted material. Examples of permitted uses are lecture materials, course packs, e-reserves, conference presentations, or in-house training courses.
- **Electronic Preprints.** Before submitting an article to an IEEE publication, authors frequently post their manuscripts to their own web site, their employer's site, or to another server that invites constructive comment from colleagues. Upon submission of an article to IEEE, an author is required to transfer copyright in the article to IEEE, and the author must update any previously posted version of the article with a prominently displayed IEEE copyright notice. Upon publication of an article by the IEEE, the author must replace any previously posted electronic versions of the article with either (1) the full citation to the IEEE work with a Digital Object Identifier (DOI) or link to the article abstract in IEEE

# LIST OF REFERENCES


[1] A. O. Adebayo. A foundation for breach data analysis. *Journal of Information Engineering and Applications*, 2(4):17–23, 2012.

[2] A. Alabduljabbar, R. Ma, S. Alshamrani, R. Jang, S. Chen, and D. Mohaisen. Poster: Measuring and Assessing the Risks of Free Content Websites. In *Network and Distributed System Security Symposium,(NDSS'22), San Diego, California*, 2022.

[3] A. Alabduljabbar, R. Ma, S. Choi, R. Jang, S. Chen, and D. Mohaisen. Understanding the Security of Free Content Websites by Analyzing their SSL Certificates: A Comparative Study. In *CySSS@AsiaCCS'22: Proceedings of the 1st Workshop on Cybersecurity and Social Sciences, Nagasaki, Japan, 30 May 2022*, pages 19–25. ACM, 2022.

[4] A. Alabduljabbar and D. Mohaisen. Measuring the Privacy Dimension of Free Content Websites through Automated privacy policy analysis and annotation. In *Companion of The Web Conference 2022, Virtual Event / Lyon, France, April 25 - 29, 2022*, pages 860–867. ACM, 2022.

[5] H. Alasmary, A. Anwar, A. Abusnaina, A. Alabduljabbar, M. Abuhamad, A. Wang, D. Nyang, A. Awad, and D. Mohaisen. ShellCore: Automating Malicious IoT Software Detection Using Shell Commands Representation. *IEEE Internet Things J.*, 9(4):2485–2496, 2022.

[6] H. Alasmary, A. Khormali, A. Anwar, J. Park, J. Choi, A. Abusnaina, A. Awad, D. Nyang, and A. Mohaisen. Analyzing and Detecting Emerging Internet of Things Malware: A Graph-Based Approach. *IEEE Internet Things J.*, 6(5):8977–8988, 2019.

[7] M. Alkinoon, A. Alabduljabbar, H. Althebeiti, R. Jang, D. Nyang, and D. Mohaisen. Understanding the Security and Performance of the Web Presence of Hospitals: A Measurement Study. In *32nd International Conference on Computer Communications and Networks, ICCCN 2023, Honolulu, HI, USA, July 24-27, 2023*, pages 1–10. IEEE, 2023.

[8] M. Alkinoon, S. J. Choi, and D. Mohaisen. Measuring Healthcare Data Breaches. In *Information Security Applications - 22nd International Conference, WISA 2021, Jeju Island, South Korea, August 11-13, 2021, Revised Selected Papers*, volume 13009 of *Lecture Notes in Computer Science*, pages 265–277. Springer, 2021.

[9] M. Alkinoon, M. Omar, M. Mohaisen, and D. Mohaisen. Security Breaches in the Healthcare Domain: A Spatiotemporal Analysis. In *Computational Data and Social Networks - 10th International Conference, CSoNet 2021, Virtual Event, November 15-17, 2021, Proceedings*, volume 13116 of *Lecture Notes in Computer Science*, pages 171–183. Springer, 2021.

[10] M. Alqadhi, A. Alabduljabbar, K. Thomas, S. Salem, D. Nyang, and D. Mohaisen. Do Content Management Systems Impact the Security of Free Content Websites? A Correlation Analysis. *CoRR*, abs/2210.12083, 2022.

[11] M. Alqadhi, A. Alkinoon, S. Salem, and D. Mohaisen. Understanding the Country-Level Security of Free Content Websites and their Hosting Infrastructure. In *10th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2023, Thessaloniki, Greece, October 9-13, 2023*, pages 1–10. IEEE, 2023.

[12] M. Alqadhi, M. Alkinoon, J. Lin, A. Abdalaal, and D. Mohaisen. Entangled Clouds: Measuring the Hosting Infrastructure of the Free Contents Web. In *Proceedings of the 2023 on Cloud Computing Security Workshop, CCSW 2023, Copenhagen, Denmark, 26 November 2023*, pages 75–87. ACM, 2023.

[13] M. Alqadhi and D. Mohaisen. The Infrastructure Utilization of Free Contents Websites Reveal their Security Characteristics. *CoRR*, abs/2311.15363, 2023.

[14] O. Alrawi and A. Mohaisen. Chains of Distrust: Towards Understanding Certificates Used for Signing Malicious Applications. In *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11-15, 2016, Companion Volume*, pages 451–456. ACM, 2016.

[15] A. Anwar, A. Khormali, J. Choi, H. Alasmary, S. Salem, D. Nyang, and D. Mohaisen. Measuring the Cost of Software Vulnerabilities. *EAI Endorsed Trans. Security Safety*, 7(23):e1, 2020.

[16] APIVoid. A framework provides JSON APIs useful for cyber threat analysis, threat detection and prevention, 2022.

[17] M. P. Bach, S. Seljan, B. Jakovic, A. Buljan, and J. Zoroja. Hospital Websites: From the Information Repository to Interactive. In *CENTERIS 2019 - International Conference on ENTERprise Information Systems / ProjMAN 2019 - International Conference on Project MANagement / HCist 2019 - International Conference on Health and Social Care Information Systems and Technologies 2019, Sousse, Tunisia*, 2019.

[18] A. Bates, J. Pletcher, T. Nichols, B. Hollembaek, D. Tian, K. R. B. Butler, and A. Alkhelaifi. Securing SSL Certificate Verification through Dynamic Linking. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*, pages 394–405. ACM, 2014.

[19] J. Berkowsky and T. Hayajneh. Security issues with certificate authorities. In *8th IEEE Annual Ubiquitous Computing, Electronics and Mobile Communication Conference, UEMCON 2017, New York City, NY, USA, October 19-21, 2017*, pages 449–455. IEEE, 2017.

[20] U. C. Bureau. North American Industry Classification System, 2023.

[21] Y. Chen and Z. Su. Guided differential testing of certificate validation in SSL/TLS implementations. In *Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2015, Bergamo, Italy, August 30 - September 4, 2015*, pages 793–804. ACM, 2015.

[22] M. Chernyshev, S. Zeadally, and Z. Baig. Healthcare data breaches: Implications for digital forensic readiness. *Journal of medical systems*, 43(1):1–12, 2019.

[23] S. J. Choi and M. E. Johnson. Understanding the Relationship Between Data Breaches and Hospital Advertising Expenditures. *The American Journal of Managed Care*, 25(5), January 2019.

[24] T. Chung, Y. Liu, D. R. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson. Measuring and Applying Invalid SSL Certificates: The Silent Majority. In *Proceedings of the 2016 ACM on Internet Measurement Conference, IMC 2016, Santa Monica, CA, USA, November 14-16, 2016*, pages 527–541. ACM, 2016.

[25] T. Chung, R. van Rijswijk-Deij, B. Chandrasekaran, D. R. Choffnes, D. Levin, B. M. Maggs, A. Mislove, and C. Wilson. A Longitudinal, End-to-End View of the DNSSEC Ecosystem. In *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017*, pages 1307–1322. USENIX Association, 2017.

[26] J. Clark and P. C. van Oorschot. SoK: SSL and HTTPS: Revisiting Past Challenges and Evaluating Certificate Trust Model Enhancements. In *2013 IEEE Symposium on Security and Privacy, SP 2013, Berkeley, CA, USA, May 19-22, 2013*, pages 511–525. IEEE Computer Society, 2013.

[27] CloudFlare. What is a top-level domain (TLD)?, 2022.

[28] CloudFlare. What is a domain name registrar?, 2022.

[29] Coretelligent. breach detection: Could you detect a data breach?, 2023.

[30] S. E. Coull, A. M. White, T. Yen, F. Monrose, and M. K. Reiter. Understanding domain registration abuses. *Comput. Secur.*, 31(7):806–815, 2012.

[31] M. Cova, C. Leita, O. Thonnard, A. D. Keromytis, and M. Dacier. An Analysis of Rogue AV Campaigns. In *Recent Advances in Intrusion Detection, 13th International Symposium, RAID 2010, Ottawa, Ontario, Canada, September 15-17, 2010. Proceedings*, volume 6307 of *Lecture Notes in Computer Science*, pages 442–463. Springer, 2010.

[32] L. Coventry and D. Branley. Cybersecurity in healthcare: A narrative review of trends, threats and ways forward. *PubMed*, April 2018.

[33] V. Enterprise. "Introduction to the 2020 DBIR: Verizon Enterprise Solutions". *Verizon Enterprise*, 2020.

[34] V. Enterprise. Verizon Data Breach Investigations Report, 2022.

[35] O. for Civil Rights (OCR). Breach Notification Rule, 2013.

[36] O. for Civil Rights (OCR). U.S. Department of Health and Human Services Office for Civil Rights Breach Portal, 2023.

[37] T. I. O. for Standardization (ISO). International Organization for Standardization: 3,166 COUNTRY CODES, 2021.

[38] K. Gwebu and C. W. Barrows. Data breaches in hospitality: is the industry different? *Journal of Hospitality and Tourism Technology*, 2020.

[39] H. Hammouchi, O. Cherqi, G. Mezzour, M. Ghogho, and M. E. Koutbi. Digging Deeper into Data Breaches: An Exploratory Data Analysis of Hacking Breaches Over Time. In E. M.

Shakshuki and A. Yasar, editors, *The 10th International Conference on Ambient Systems, Networks and Technologies (ANT 2019) / The 2nd International Conference on Emerging Data and Industry 4.0 (EDI40 2019) / Affiliated Workshops, April 29 - May 2, 2019, Leuven, Belgium*, volume 151 of *Procedia Computer Science*, pages 1004–1009. Elsevier, 2019.

[40] H. S. C. C. C. (HC3). A Cost Analysis of Healthcare Sector Data Breaches, 2019.

[41] D. Healthcare. What is the Difference Between Non-Profit and For-Profit Hospitals?, 2022.

[42] HIPPA. What are the Penalties for HIPAA Violations?, 2023.

[43] California Attorney General's Office. Data branch notification report, 2023.

[44] IBM. Cost of a Data Breach Report 2023, 2023.

[45] I. B. M. C. (IBM). How much does a data breach cost?, 2021.

[46] I. B. M. C. (IBM). What is a data breach?, 2023.

[47] I. S. R. G. (ISRG). Let's Encrypt, 2020.

[48] F. Kamoun and M. Nicho. Human and Organizational Factors of Healthcare Data Breaches: The Swiss Cheese Model of Data Breach Causation And Prevention. *Int. J. Heal. Inf. Syst. Informatics*, 9(1):42–60, 2014.

[49] F. Kamoun and M. Nicho. Human and Organizational Factors of Healthcare Data Breaches: The Swiss Cheese Model of Data Breach Causation And Prevention. *Int. J. Heal. Inf. Syst. Informatics*, 9(1):42–60, 2014.

[50] D. Kim, H. Cho, Y. Kwon, A. Doupé, S. Son, G. Ahn, and T. Dumitras. Security Analysis on Practices of Certificate Authorities in the HTTPS Phishing Ecosystem. In *ASIA CCS '21: ACM Asia Conference on Computer and Communications Security, Virtual Event, Hong Kong, June 7-11, 2021*, pages 407–420. ACM, 2021.

[51] D. Kim, B. J. Kwon, and T. Dumitras. Certified Malware: Measuring Breaches of Trust in the Windows Code-Signing PKI. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, CCS 2017, Dallas, TX, USA, October 30 - November 03, 2017*, pages 1435–1448. ACM, 2017.

[52] D. Kim, B. J. Kwon, K. Kozák, C. Gates, and T. Dumitras. The Broken Shield: Measuring Revocation Effectiveness in the Windows Code-Signing PKI. In *27th USENIX Security Symposium, USENIX Security 2018, Baltimore, MD, USA, August 15-17, 2018*, pages 851–868. USENIX Association, 2018.

[53] B. J. Kwon, S. Hong, Y. Jeon, and D. Kim. Certified Malware in South Korea: A Localized Study of Breaches of Trust in Code-Signing PKI Ecosystem. In *Information and Communications Security - 23rd International Conference, ICICS 2021, Chongqing, China, November 19-21, 2021, Proceedings, Part I*, volume 12918 of *Lecture Notes in Computer Science*, pages 59–77. Springer, 2021.

[54] B. . E. law firm. HIPAA Regulations: General Provisions: Definitions: Health Plan - § 160.103, 2015.

[55] J. Lee and S. J. Choi. Hospital productivity after data breaches: Difference-in-differences analysis. *Journal of medical Internet research*, 23(7), 2021.

[56] V. Liu, M. A. Musen, and T. Chou. Data Breaches of Protected Health Information in the United States. *JAMA*, 313(14):1471–1473, 2015.

[57] C. Makridis and B. Dean. Measuring the economic effects of data breaches on firm outcomes. *Journal of Economic and Social Measurement*, 43(1-2):59–83, 2018.

[58] A. McLeod and D. Dolezel. Cyber-analytics: Modeling factors associated with healthcare data breaches. *Decision Support Systems*, 108:57–68, 2018.

[59] N. Menachemi1 and T. H. Collum. Benefits and drawbacks of electronic health record systems, 2011.

[60] U. Meyer and V. Drury. Certified Phishing: Taking a Look at Public Key Certificates of Phishing Websites. In *Fifteenth Symposium on Usable Privacy and Security, SOUPS 2019, Santa Clara, CA, USA, August 11-13, 2019*. USENIX Association, 2019.

[61] M. A. Mishari, E. D. Cristofaro, K. M. E. Defrawy, and G. Tsudik. Harvesting SSL Certificate Data to Identify Web-Fraud. *Int. J. Netw. Secur.*, 14(6):324–338, 2012.

[62] A. Mohaisen. Towards Automatic and Lightweight Detection and Classification of Malicious Web Contents. In *Third IEEE Workshop on Hot Topics in Web Systems and Technologies, HotWeb 2015, Washington, DC, USA, November 12-13, 2015*, pages 67–72. IEEE Computer Society, 2015.

[63] A. Mohaisen and O. Alrawi. Unveiling Zeus: automated classification of malware samples. In *22nd International World Wide Web Conference, WWW '13, Rio de Janeiro, Brazil, May 13-17, 2013, Companion Volume*, pages 829–832. International World Wide Web Conferences Steering Committee / ACM, 2013.

[64] A. Mohaisen, O. Alrawi, and M. Mohaisen. AMAL: High-fidelity, behavior-based automated malware analysis and classification. *Comput. Secur.*, 52:251–266, 2015.

[65] O. Networks. Largest Healthcare Data Breaches Reported in February 2022 Confirms Need for Network Security Based on Zero Trust Microsegmentation, 2022.

[66] NLTK. Natural Language Toolkit, 2022.

[67] U. D. of Health and H. S. (HHS). Business Associates, 2019.

[68] U. D. of Health and H. S. (HHS). Health Insurance Portability and Accountability Act, 2023.

[69] U. D. of Homeland Security. Homeland Infrastructure Foundation-Level Data (HIFLD), 2019.

[70] T. O. of the National Coordinator for Health Information Technology (ONC). Benefits of EHRs, 2017.

[71] Office for Civil Rights. Breach Notification Rule, 2013.

[72] OpenAI. March 20 ChatGPT Outage, 2023.

[73] D. Oregon Department of Transportation and M. V. Services. Data Breach - Oregon DMV, 2023.

[74] Pingdom. Website Performance and Availability Monitoring, 2022.

[75] W. Raghupathi, V. Raghupathi, and A. Saharia. Analyzing Health Data Breaches: A Visual Analytics Approach. *AppliedMath*, pages 175–199, 2023.

[76] N. P. D. Rank. The National Practitioner Data Bank (NPDB), 2021.

[77] M. Roetteler, M. Naehrig, K. M. Svore, and K. E. Lauter. Quantum Resource Estimates for Computing Elliptic Curve Discrete Logarithms. In *Advances in Cryptology - ASIACRYPT 2017 - 23rd International Conference on the Theory and Applications of Cryptology and Information Security, Hong Kong, China, December 3-7, 2017, Proceedings, Part II*, volume 10625 of *Lecture Notes in Computer Science*, pages 241–270. Springer, 2017.

[78] A. Sarabi, P. Naghizadeh, Y. Liu, and M. Liu. Risky business: Fine-grained data breach prediction using business profiles. *Journal of Cybersecurity*, 2(1):15–28, 2016.

[79] SecurityWeek. Yum Brands Discloses Data Breach Following Ransomware Attack, 2023.

[80] A. H. Seh, M. Zarour, M. Alenezi, A. Sarkar, A. Agrawal, R. Kumar, and P. R. Khan. Healthcare Data Breaches: Insights and Implications. *Healthcare*, 8:133, 05 2020.

[81] B. K. Siddartha and G. K. Ravikumar. Analysis of Masking Techniques to Find out Security and other Efficiency Issues in Healthcare Domain. In *Third International conference on I-SMAC*, pages 660–666, 2019.

[82] T. Smith. Examining Data Privacy Breaches in Healthcare. Technical report, Walden University, 2016.

[83] T. T. Smith. Examining Data Privacy Breaches in Healthcare. Technical report, Walden University, 2016.

[84] B. Steinwald, , and D. Neuhauser. The Role of the Proprietary Hospital, 1970.

[85] Sucuri. Website security check and malware scanner, 2023.

[86] T-Mobile. T-Mobile Informing Impacted Customers about Unauthorized Activity, 2023.

[87] United States Department of Justice. alleged international hacker indicted for massive attack on u.s. retail and banking networks, 2009.

[88] U.S. HHS. Business Associate Contracts, 2013.

[89] M. J. Van Eeten and J. M. Bauer. Economics of malware: Security decisions, incentives and externalities. *OECD Science*, 2008.

[90] L. Vargas, L. Blue, V. Frost, C. Patton, N. Scaife, K. R. B. Butler, and P. Traynor. Digital Healthcare-Associated Infection: A Case Study on the Security of a Major Multi-Campus Hospital System. In *26th Annual Network and Distributed System Security Symposium, NDSS 2019, San Diego, California, USA, February 24-27, 2019*. The Internet Society, 2019.

[91] Veris community. The Vocabulary for Event Recording and Incident Sharing (VERIS), 2021.

[92] VirusTotal. Analyze suspicious files and URLs to detect types of malware, automatically, 2023.

[93] W3Techs. Usage statistics of Default protocol https for websites, 2022.

[94] S. Walker-Roberts, M. Hammoudeh, O. Aldabbas, M. Aydin, and A. Dehghantanha. Threats on the horizon: Understanding security threats in the era of cyber-physical systems. *The Journal of Supercomputing*, 76(4):2643–2664, 2020.

[95] S. Walker-Roberts, M. Hammoudeh, and A. Dehghantanha. A Systematic Review of the Availability and Efficacy of Countermeasures to Internal Threats in Healthcare Critical Infrastructure. *IEEE Access*, 6:25167–25177, 2018.

[96] WHOIS. Registration data lookup tool, 2023.

[97] S. Wikina. What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches. *Perspectives in health information management*, page 1h, 2014.

[98] S. B. Wikina. What Caused the Breach? An Examination of Use of Information Technology and Health Data Breaches. *Perspectives in health information management*, 11(Fall), 2014.

[99] X. Yu, N. Samarasinghe, M. Mannan, and A. M. Youssef. Got Sick and Tracked: Privacy Analysis of Hospital Websites. In *IEEE European Symposium on Security and Privacy, EuroS&P 2022 - Workshops, Genoa, Italy, June 6-10, 2022*. IEEE, 2022.

[100] L. Zhang, D. R. Choffnes, T. Dumitras, D. Levin, A. Mislove, A. Schulman, and C. Wilson. Analysis of SSL certificate reissues and revocations in the wake of heartbleed. *Commun. ACM*, 61(3):109–116, 2018.