

Information Assurance Technique for Mitigation of Data Breaches in the Human Service Sector

Chevroen Washington, Phillip Yarbrough, Shavon Parker, Rafia Islam, Vishnu Vardhan Patamsetti, Olatunde Abiona

Department of Computer Information Systems, Indiana University Northwest, Gary, USA

Email: chswash@indiana.edu, pmyarbrough2631@gmail.com, shavpark@iu.edu, rislam@iun.edu, vpatamse@iu.edu, oabiona@iun.edu

How to cite this paper: Washington, C., Yarbrough, P., Parker, S., Islam, R., Patamsetti, V.V. and Abiona, O. (2022) Information Assurance Technique for Mitigation of Data Breaches in the Human Service Sector. *Int. J. Communications, Network and System Sciences*, 15, 15-30.

<https://doi.org/10.4236/ijcns.2022.152002>

Received: October 29, 2021

Accepted: February 14, 2022

Published: February 17, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

This research paper analyzes data breaches in the human service sector. The hypothesis for the solution to this problem is that there will be a significant reduction in data breaches in the human service sector due to an increase in information assurance. The hypothesis is tested using data from the United States Department of Health and Human Services data breach notification repository during January 2018-December 2020. Our result shows that without the increased mitigation of information assurance, data breaches in the human service sector will continue to increase.

Keywords

Information Assurance, Ransomware, Data Breach, Hacker, HIPPA, Phishing, Department of Health and Human Services

1. Introduction

In March 2021, a major hacking event occurred when hackers infiltrated the Microsoft Exchange Server. The hacking party stole emails from over 250,000 customers (about half the population of Wyoming) worldwide. This is only the latest major hacking scheme in 2021 and there are countless events that are happening as we speak. There are two major questions that need to be answered every time one of these events occurs. How can we stop these attacks? How can we find out what data has been affected and the damage that can occur now and in the future? The answer is information assurance. According to the National Institute of Standards and Technology, information assurance is “the measures

that protect and defend information and information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation. These measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities” [1].

To understand that there will be a significant reduction in data breaches in the human service sector due to an increase in information assurance, we must understand 2020 data breaches and their linear regression towards 2021. Dan Lohrmann recently authored an article titled *2020 Data Breaches Point to Cybersecurity Trends in 2021*. A shocking statistic from this article is “despite 1923 breaches (49%) without a confirmed number of records exposed, the total number of records compromised in 2020 exceeded 37 billion, a 141% increase compared to 2019 and by far the most records exposed in a single year since we have been reporting on data breach activity” [2]. It was also noted that 676 breaches that included some types of ransomware were a 100% increase compared to 2019 breaches. Another shocking statistic from this article was that there were five data breaches that had the most impact in the year 2020. These data breaches exposed one billion or more records. Underneath these five major data breaches, 18 breaches exposed between 100 million and one million records [2].

Healthcare was the most victimized industry in 2020 when it came to data breaches. Healthcare data breaches accounted for 12.3% of reported breaches in the year 2020. It is speculated that this number will rise dramatically, again, in 2021. It is no surprise that the human service sector is under attack with data breaches occurring daily throughout the world. Implementing, practicing, and maintaining information assurance throughout the human service sector will drastically reduce data breaches [2]. **Figure 1** below shows the top 5 biggest ransomware attack payouts in 2020.

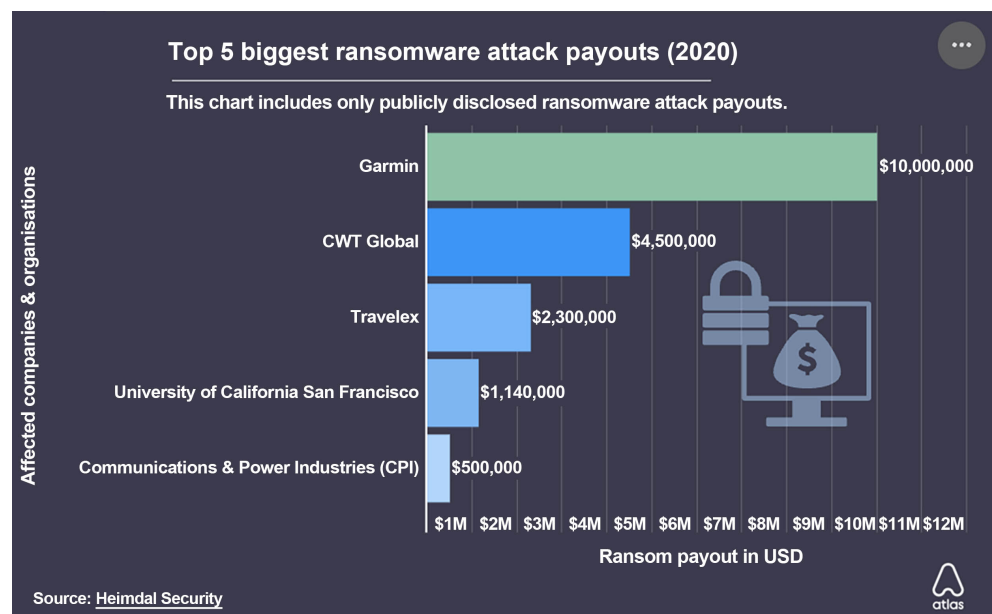


Figure 1. Top 5 biggest ransomware attack payouts (2020).

The goal of this paper is to explore the current state of data breaches in the human service sector, the impact of data breaches, the benefits of increased information assurance, and the outcome if no changes are made. Section 2 will discuss the literature review of multiple resources and summarizations of the current state of data breaches within the human service sector, data breaches in detail, and the laws created to protect health information. In Section 3, we detail our proposed design that moves information assurance to the front of policies and procedures in the human service sector on the organizational level. In Section 4, we perform an analysis of our design using actual breach notification data to run predictive statistics. Section 5 concludes our paper by exploring the next steps of increasing information assurance will decrease data breaches.

2. Literature Review

Alina Pecteu authored an article on the top five ransomware payouts where a company could save millions of dollars. The second largest ransomware attack was on Colonial Pipeline. The cybercriminal group called DarkSide was involved in the Colonial Pipeline attack in 2021. The hackers' main goal was to target the company's business network and infiltrate its billing system. Colonial Pipeline is the largest fuel pipeline operator in the United States of America and carries refined gasoline and jet fuel across a long route spanning from Texas to New York. In exchange for data decryption, operators asked for the same sum as in the Brenttag case, \$4.4 million [3].

The first largest ransomware attack was on CWT Global. According to Alina, "according to a record of ransom negotiations seen by Reuters, the US travel services company CWT paid \$4.5 million to malicious hackers who stole vast amounts of their confidential business files and said they had taken 30,000 computers down" [3]. These hackers stole two terabytes of data which included financial records, security documents, and details on all employees.

If information assurance were implemented, practiced, and maintained in these two companies, there would have been less of a drastic measure when it comes to leaked data, ransom payments, and future data corrupted. Information assurance could have had proper steps involved to deviate from going down these drastic routes, thus causing a user's data to be protected. According to PurleSec: Cyber security services, **Figure 2** shows the average ransomware payout cost per incident.

The cost per incident per year could have been saved with information assurance. There is a steady rise in the average payout per year for ransomware payout costs and it is expected that these costs will rise dramatically in 2021 and 2022. Healthcare organizations are experiencing data breaches by frequently using business associates and covered entities to provide necessary care for patients [4]. According to the US Department of Health and Human Services (DHHS), "individuals, organizations, and agencies that fall within the guidelines of a covered entity under the Health Insurance Portability and Accountability Act of

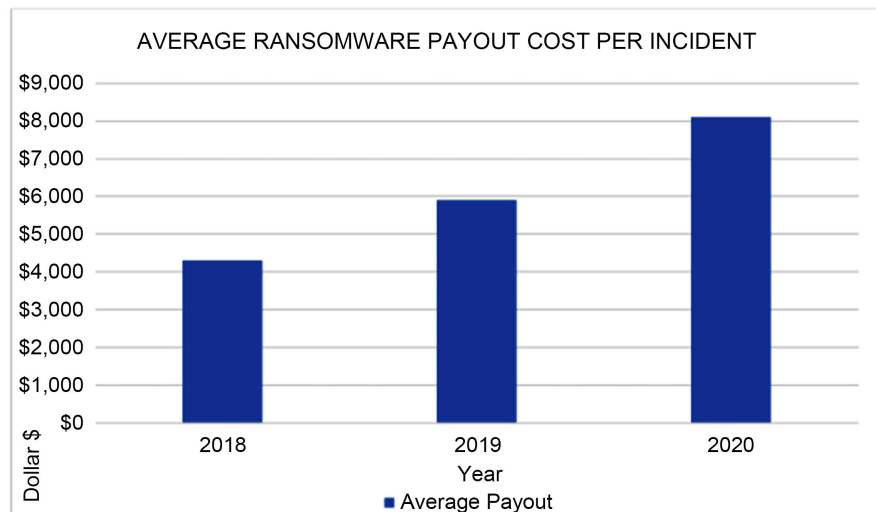


Figure 2. Average ransomware payout cost per incident.

1996 (HIPAA) must comply with the Privacy Rules' to shield the security and privacy of health information and must provide individuals with certain rights concerning their health information" [5].

2.1. Information Assurance

Information assurance or IA is essential to prevent data breaches because of the idea of making sure data is correctly stored to protect the user's data. AI (Artificial Intelligence) stands by these simple principles: integrity, availability, authenticity, confidentiality, and non-repudiation [6]. These risk assessments help organizations identify vulnerabilities capable of allowing threats to impact an entire infrastructure, individual systems, or business processes; information assurance risk evaluation provides knowledge about the probability of a threat exploiting an asset's vulnerability as well as the potential impact it could have from a cost, business operation, compliance, or technology perspective [6].

Information assurance identifies ways to control and safeguard critical information in a more effective manner, stressing organizational risk management and overall information quality [7]. IA is typically a broader strategic initiative comprised of a wide range of information protection and management processes; some examples of this can include security audits, network architecture, compliance audits, database management and development, implementation, and enforcement of organizational information management policies [7]. The goal is to maintain data integrity, reliability, and accessibility, including taking precautions against unauthorized destruction or alteration of information and ensuring non-repudiation and the authenticity of data [7]. The main goals of information assurance will make it an ideal method for helping prevent future data breaches.

To protect the healthcare sector's data system, we must understand that HIPAA and EMRs (Electronic Medical Records) give assurance to protect the privacy

and security of PHI (Protected Health Information) that must be managed in a technologically driven environment. A company under any healthcare service can acquire technological tools that can aid in monitoring security and privacy compliance to assure security. Trish Markus (2004) questioned the establishment of a “culture of compliance,” that indicates management involvement and commitment issues through employee communication and training procedures about information assurance. Mercuri (2004) quoted a chief information officer as stating that HIPAA “compliance is not sold in a bottle,” where, “providing employees with policies and procedures for their job classification and requiring them to read and sign off on them” is not adequate.

2.2. Process of Information Assurance

By the US Government’s definition information assurance is a measure that protects and defends information systems by ensuring their availability, integrity, authentication, confidentiality, and non-repudiation; these measures include providing for restoration of information systems by incorporating protection, detection, and reaction capabilities.

To build an effective information assurance we often debate between technology-related sources and theory building and testing sources. Both TRA and TAM utilize factors including individual patients’ beliefs, attitudes, and intention to adopt technology such as EMRs that assure the information assurance policy associated with those EMRs. In the study of Hu, Chau, and Tulu *et al.* (1999, 2002, 2003), they mentioned TAM as adequate, with exception to TAM’s explanation of attitude and intention. This theory builds the perception of telemedicine for physicians useful, Hu *et al.* (1999) suggested, “proper user training is essential. Attitude also significantly influenced physician behavioral intention”. **Figure 3** below shows the theory of reasoned action.

The widely accepted theory of TRA has often been used to support normal relationships between external factors, beliefs, attitudes, intentions, and behavior. According to the article, “TAM is a TRA spin-off and has supported the intention to use, perceived usefulness, and behavior when adopting modern technology.” For information assurance, these theoretical techniques can examine the acceptance and compliance behavior of a system user who wants to adopt new organizational policies. In the context of information assurance, the TRA model will have the capability to capture the compliance of healthcare security and protect privacy policies.

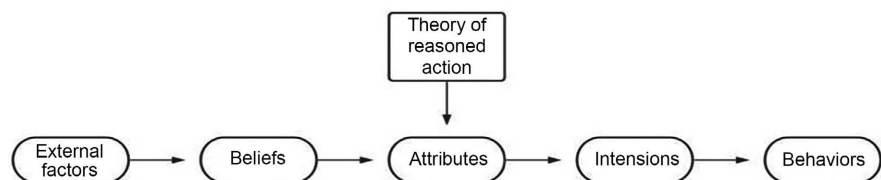


Figure 3. Theory of reasoned action.

2.3. Health Insurance Portability and Accountability Act

The Health Insurance Portability and Accountability Act of 1996 was signed into law by President William “Bill” Clinton. This new law created the national standards to protect patient health information that was deemed sensitive from being exposed with the consent of the patient. During the time of its creation the requirement of electronic medical records was nowhere on the public radar. “As a part of the American Recovery and Reinvestment Act, all public and private health-care providers and other eligible professionals were required to adopt and demonstrate “meaningful use” of electronic medical records (EMR) by January 1, 2014 in order to maintain their existing Medicaid and Medicare reimbursement levels”. The ARRA sent the entire medical industry into a tailspin requiring the change from paper filing too digital in five years’ time. While this was occurring, HIPAA was also updating with the new information by enacting the HIPAA Breach Notification Rule in 2009. “The HIPAA Breach Notification Rule, 45 CFR §§ 164.400-414, requires HIPAA covered entities and their business associates to provide notification following a breach of unsecured protected health information”. There is also a rule under the Federal Trade Commission, however for this paper we will focus on the Department of Health and Human Services (HHS). In HHA a data breach is defined as follows:

“An impermissible use or disclosure of protected health information is presumed to be a breach unless the covered entity or business assessment of at least the following factors:

- 1) The nature and extent of the protected health information involved, including the types of identifiers and likelihood of re-identification;
- 2) The unauthorized person who used the protected health information or to whom the disclosure was made;
- 3) Whether the protected health information or to whom the disclosure was made;
- 4) The extent to which the risk to the protected health information has been mitigated” [8].

HHA has also provided information on what is deemed as protected health information that is breached. “Unsecured protected health information is protected health information that has not been rendered unusable, unreadable, or indecipherable to unauthorized persons through the use of a technology or methodology specified by the Secretary in guidance” [8]. If an applicable covered health-care provider or eligible entity were to encounter a data breach HHA has provided the steps to notify them. “Following a breach of unsecured protected health information, covered entities must provide notification of the breach to affected individuals, the Secretary, and in certain circumstances, to the media” [8]. The media must be notified by the protected entity if the data breach impacts over 500 residents of a state or jurisdiction. “Like individual notice, this media notification must be provided without unreasonable delay and in no case later than 60 days following the discovery of a breach...” [8].

In Section 3, we detail our proposed design that moves information assurance to the front of policies and procedures in the human service sector on the organizational level.

3. Design and Proposed Work

3.1. Ideas for Reducing the Data Breaches that Can Be Implemented

The healthcare sector has the highest cost record of data breaches compared to other industries in the past decade. The reason to get attracted to healthcare data is that the attacker can get insurance information, tax information, and social security number easily. This information helps the criminal to conduct insurance and tax fraud which is profitable. The following are some ideas that we can prevent data breaches in the healthcare sector.

3.2. Analyze the Existing Security System

According to HIPAA, it is necessary to analyze the existing system, keep it upgraded and have a backup plan when any major threat appears. Keeping the running system up to date prevents a lot of malicious activity and reduces the annual maintenance cost [9].

3.3. Conduct a Security Risk Analysis

The first step the health care providers can take that submit their system to a security evaluation according to HIPAA and HITECH (Health Information Technology for Economic and Clinical Health) (Health Information Technology for Economic and Clinical Health) (Health Information Technology for Economic and Clinical Health). When a system goes through constant analysis, it is easier to detect threats and prevent them [9].

3.4. Keep the Staff Educated

The staff need to be well educated and trained to prevent data breaches and the training process needs to be constant. When any new case or technology update comes by the employers need to set up a meeting and discuss the topics with staff. The staff should educate themselves with the upgrade and adopt changes of technology [9].

3.5. Encrypt Data

Encryption technologies can help migrate the components of cyberattacks. As we know, encrypted data is secured in case of lost data without breaches. Encryption can also save the company from government petitions [9].

3.6. Develop an Exit Strategy

In exit strategy the employees should leave the work premises with a proper logout log and the time-to-time records. Every exit action should be maintained

with restrictions and time maintenance. The records need to be well maintained to prevent data leakage and should be done by employees and authorities. All the login credentials and passwords should automatically update and should be generated at every instance of time. The employees should be aware of the company security terms & conditions, and the legal actions if someone breaks them [9].

3.7. Update Software with All Patches and Updates

Software companies are constantly inventing and updating new products and technologies to protect data and healthcare companies should stay up to date to protect their data. Sometimes the protection can be costly, but it prevents vulnerabilities and increases security. Right patches and updates allow for organizations to control threats and let them not affect their businesses [9].

3.8. Identify and Attack the Source

Many companies focus on the training of staff to reduce the number of data breaches they encounter. Physical human error has been a major impact of the safety of protected medical information, however that was prior to the mandate of electronic medical records. All entities that collect or have access to protected medical information must submit notification to the U.S. Department of Human and Health Services of data breaches. **Figure 4** below shows the percentage of data breaches reported to HHS and business associates present during 2018-2020. As shown in **Figure 4**, upon review of the last three years of data breaches notifications collected there were 31% that occurred in the presence of an employee [10]. Agencies experienced more than 50% of their overall breaches due to hacking.

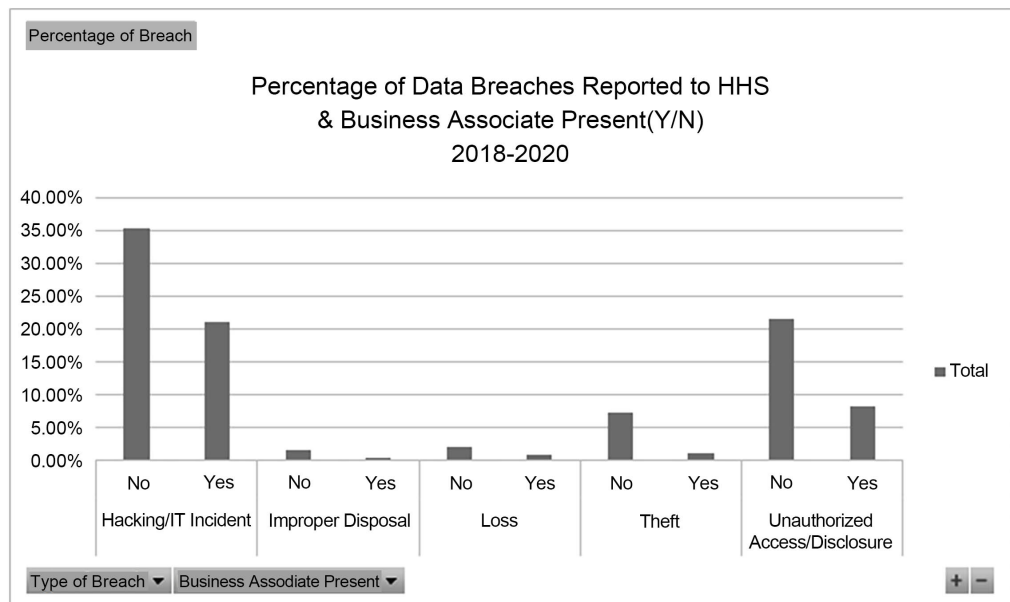


Figure 4. Percentage of data breaches reported to HHS with business associate present.

Figure 5 below shows the percentage of hacking incident office locations reported to HHS. Figure 5 takes a closer look into the hacking reported by the agencies it shows that top percentage of hacking location was email at 45% and network server at 38% [10].

The error in both scenarios can be related to not having a strong information assurance program at the agency. During the push to electronic medical records health providers focused on improving their equipment and left information assurance to the back burner as something their employees needed to learn. Figure 6 below shows the disconnection of IA from human service sector.

A hacker breached health care providers over 50 percent of the notifications over a three-year period with over 40 percent of the access location being email [10]. Typically hacking attacks via email occur with phishing emails sent to employees that have links that let the hackers into the company. To reduce the amount of data breaches caused by email phishing healthcare providers need to pour more into information assurance. Figure 7 shows a graphical relationship between EMR, medical equipment and IA.

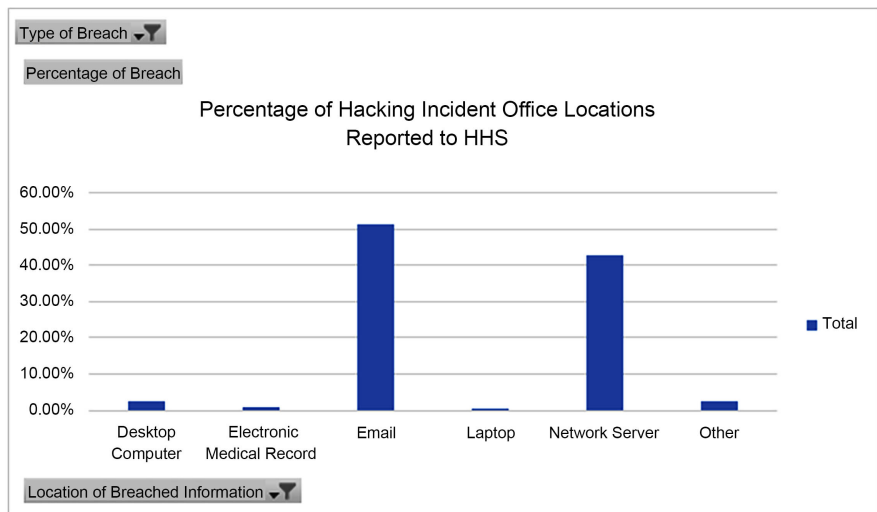


Figure 5. Percentage of hacking incident office locations reported to HHS

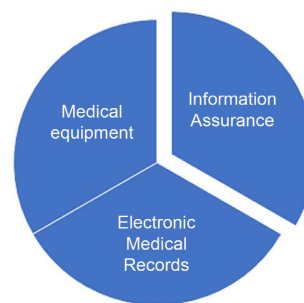


Figure 6. Pie chart of disconnection of IA from human service sector.

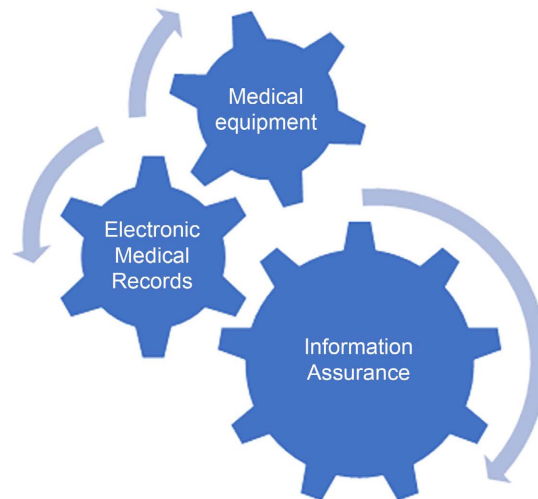


Figure 7. Graphics of relationship with IA.

Section 4 will discuss the analysis of design based on descriptive statistics, inferential statistics, and tree classifiers for our hypothesis using data reported to the U.S. Dept. Of Health and Human Services (HHS) breach notification repository.

4. Analysis of Design

The focus of this paper is the importance of increasing information assurance in the human service sector and by doing so will decrease the amount of data breaches. **Figure 8** shows a screenshot of the data retrieved HHS. For the analysis of the design the data used are breach notifications to HHS from 1/1/2018-12/31/2020.

The following descriptive and inferential statistics will show the basis for this paper's push for increased information assurance.

4.1. Descriptive Statistics

The data collected from HHS data breach notification database offers quite an insight into the human service sector breaches. **Figure 9** shows data breaches reported to HHS by state. The following graph displays notification based on which state has seen the most breaches in their human service sectors. The state of California had 109 reported breaches over the three-year period with only three entities with a repeat report. The state of Texas had 97 reported breaches over the three-year period with only four entities with a repeat report. There are some states that reported less than ten breaches over the three-year period which are Arkansas, Delaware, North & South Dakota, and Vermont.

Figure 10 shows data breaches reported to HHS by business associate present. The following graph shows a business associate was present at the time of the breach. For most of the breaches no one was present, therefore showing a preview into why information assurance is needed.

Name of State	Covered Individual Breach	Type of Location	Business	Web Description	
Nissan MO TN	Health Pla	50410	Hacking/Email	Yes	Magellan Rx Pharmacy, the covered entity (CE), reported that its business associate (BA) experienced a ransomware attack affecting the electronic protected health information (ePHI) of 50,410 individuals. This case is du...
Northwes IL	Healthcan	682	Unauthori	Electronic No	The covered entity (CE), Northwestern Memorial Hospital, reported that a workforce member impermissibly accessed the electronic protected health information (ePHI) of 682 individuals. The ePHI involved included name...
Home Sta MO	Health Pla	1020	Unauthori	Paper/Fili Yes	The covered entity (CE), Home State Health Plan, Inc., reported that its business associate (BA) sent letters containing the protected health information (PHI) of 1,020 individuals, to the wrong recipients. The PHI involved in...
Beebe Me DE	Healthcan	56953	Hacking/Network	Yes	The covered entity (CE), Beebe Medical Foundation, reported that its business associate experienced a ransomware attack affecting the electronic protected health information (ePHI) of approximately 56,953 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Communi CT	Healthcan	1634	Unauthori	Email No	The covered entity (CE), Community Health Resources, Inc., reported that an employee inadvertently sent a group email to 1,634 individuals without using the blind copy function. The electronic protected health information (ePHI) of 1,634 individuals was disclosed to the wrong recipients.
Allcare H OR	Health Pla	5707	Hacking/Network	Yes	The covered entity (CE), AllCare Health, Inc., reported that its business associate experienced a ransomware attack that affected the electronic protected health information (ePHI) of 5,707 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Magnolia MS	Health Pla	759	Unauthori	Paper/Fili Yes	The covered entity (CE), Magnolia Health Plan, reported that an employee of its business associate (BA) inadvertently mailed documents containing the protected health information (PHI) of 759 individuals to the wrong recipients. The PHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Five Point GA	Healthcan	1223	Hacking/Network	No	The covered entity (CE), Five Points Optometrists, P.C., dba Five Points Eye Care, reported that it was the victim of a cyber-attack that affected the electronic protected health information (ePHI) of 1,223 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Tom Woo IN	Health Pla	912	Hacking/Network	No	The covered entity (CE), Tom Wood East, Inc., reported that it was the victim of a ransomware attack that affected the electronic protected health information (ePHI) of 912 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
GenRx Ph AZ	Healthcan	137110	Hacking/Network	No	The covered entity (CE), GenRx Pharmacy, reported that it was the victim of a ransomware attack that affected the electronic protected health information (ePHI) of 137,110 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Nebraska NE	Healthcan	39912	Hacking/Network	Yes	The covered entity (CE), Nebraska Methodist Health System, reported that its business associate experienced a ransomware attack affecting the electronic protected health information (ePHI) of approximately 39,912 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Holy Rede PA	Healthcan	1295	Hacking/Email	No	The covered entity (CE), Holy Redeemer Ambulatory Surgical Center, reported that an employee was the victim of an email phishing attack that affected the electronic protected health information (ePHI) of 1,295 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
HMC Heal FL	Business A	985	Hacking/Email	Yes	HMC Healthworks, Inc., a business associate (BA), reported that several employees were the victims of an email phishing scheme that affected the electronic protected health information (ePHI) of 985 individuals. The ePHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.
Wellness PA	Healthcan	545	Theft	Paper/Fili No	Wellness Pharmacy, the covered entity (CE), reported that documents containing the protected health information (PHI) of 545 individuals was stolen during rioting and looting. The PHI involved included names, dates of birth, and other personally identifiable information (PII) of the affected individuals.

Figure 8. Screenshot of data retrieved from HHS.

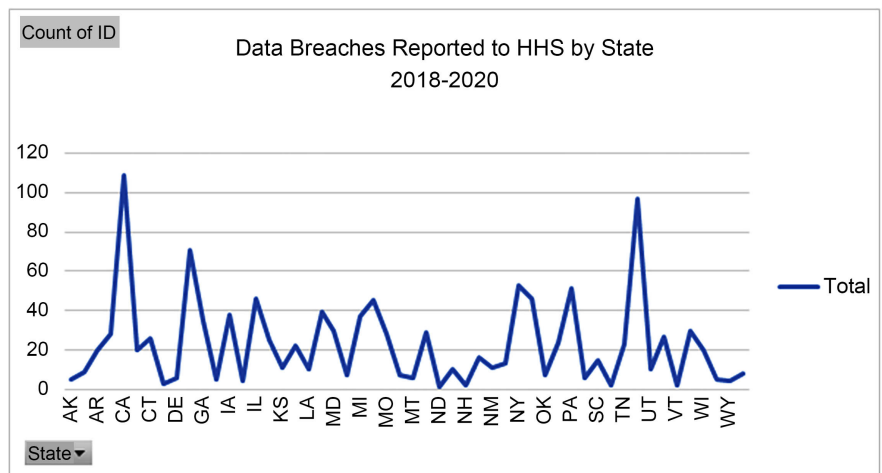


Figure 9. Data breaches reported to HHS by state.

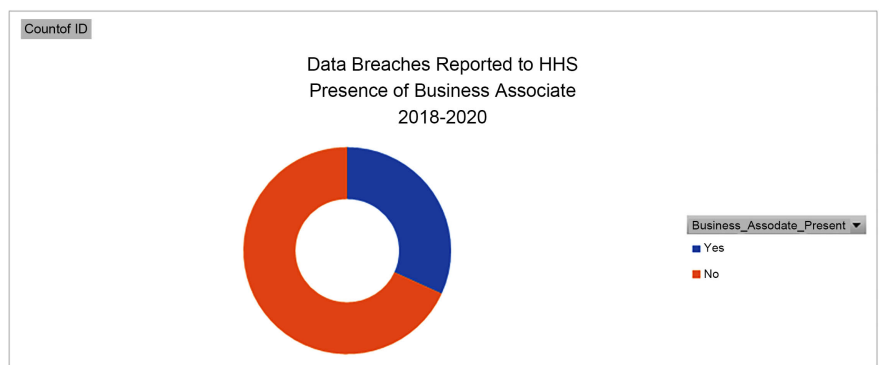


Figure 10. Data breaches reported to HHS by presence of business associate.

Figure 11 shows data breaches reported to HHS during 2018-2020. The following graph offers a look at the breakdown of breaches over the three-year period. In 2018 there were 369, in 2019 there were 477, and in 2020 there were 358 reported breaches, respectively.

The following set of tables is the summary statistics for each attribute in the dataset retrieved from HHS over the three-year period.

The top covered entity that reported breaches was Walmart Inc. with a frequency of six reports. The top covered entity type was Healthcare Provider with a frequency of 919. The top type of breach was Hacking/IT Incident with a frequency of 681. The top location of a breach was email with a frequency of 394.

Figure 12 displays the summary statistics for the attributes “Name of Covered Entity”, “State”, and “Covered Entity Type”. Both are categorical attributes, therefore the only information computed was the total count of records, the count of unique records, and the most frequent record. The top record for each attribute was Walmart Inc., California, and Healthcare Provider, respectively.

Figure 13 displays the summary statistics for the attributes “Individual Affected”, “Breach Submission Date”, and “Type of Breach”. The average number of people affected by the breaches over the three-year period was 58,948. Out of the 1204 total records 553 of them had unique submission dates. The top or most frequent type of breach was Hacking/IT Incident over the three-year period.

Figure 14 displays the summary statistics for the attributes “Location of Breached Information” and “Business Associate Present”. The top location of breaches was “email,” and the top response of business associate present was “no”.

Figure 15 shows a summary statistics of web description. In the figure below, it displays the last attribute which is “Web Description” which is an open text field for reporters to describe the incident.

4.2. Inferential Statistics

Based on the data collected from HHS hacking is the leading way the data has been breached in the human service sector. We believe that the type of breach has a higher importance on the location of the breach than the presence of a business associate. Using the python model “ExtraTreesClassifier” to find the feature importance of the data we were able to test HHS data based on if no changes are to the human services sector, which proves our hypothesis. “ExtraTreesClassifier” is imported into Python from the “sklearn” program which using an estimator to fit several randomized decision trees on various samples of data then using it to improve the predictive accuracy and control over-fitting to the data. Using this model to determine the likelihood that type of breach or presence of business associate will appear again based on decisions made prior.

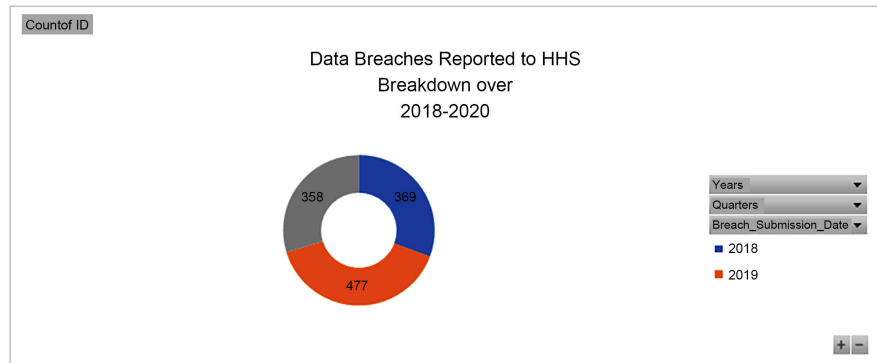


Figure 11. Data breaches reported to HHS Breakdown over 2018-2020.

```

Name_of_Covered_Entity State Covered_Entity_Type \
count 1204 1196 1203
unique 1128 51 4
top Walmart Inc. CA Healthcare Provider
freq 6 109 919
mean NaN NaN NaN
std NaN NaN NaN
min NaN NaN NaN
25% NaN NaN NaN
50% NaN NaN NaN
75% NaN NaN NaN
max NaN NaN NaN
    
```

Figure 12. Summary statistics for name of covered entity, state, and covered entity type.

```

Individuals_Affected Breach_Submission_Date Type_of_Breach \
count 1.204000e+03 1204 1204
unique NaN 553 5
top NaN 9/14/2020 Hacking/IT Incident
freq NaN 20 681
mean 5.894810e+04 NaN NaN
std 4.813961e+05 NaN NaN
min 5.000000e+02 NaN NaN
25% 1.148500e+03 NaN NaN
50% 3.307000e+03 NaN NaN
75% 1.500250e+04 NaN NaN
max 1.150000e+07 NaN NaN
    
```

Figure 13. Summary statistics individuals affected, breach submission date, and type of breach.

```

Location_of_Breached_Information Business_Associate_Present \
count 1204 1204
unique 50 2
top Email No
freq 394 821
mean NaN NaN
std NaN NaN
min NaN NaN
25% NaN NaN
50% NaN NaN
75% NaN NaN
max NaN NaN
    
```

Figure 14. Summary statistic for location of breached information and business associate present.

Figure 16 below displays the modeling results of the top two attributes from the dataset “Business Associate Presence” and “Type of Breach” after the data was normalized to the same scale. On a scale of zero to one the presence of a business associate (BAP) was 0.15 of importance based on the location of the breach. On a scale of zero to one the type of breach (ToB) was 0.8 of importance based on the location of the breach. It can be inferred from Figure 15 that ToB has a higher importance than BAP on the location of the breach, therefore the focus should be on using information assurance to prevent hacking.

	Web_Description
count	1056
unique	1054
top	Northwood, Inc., a business associate (BA) of ...
freq	2
mean	NaN
std	NaN
min	NaN
25%	NaN
50%	NaN
75%	NaN
max	NaN

Figure 15. Summary statistics of web description.

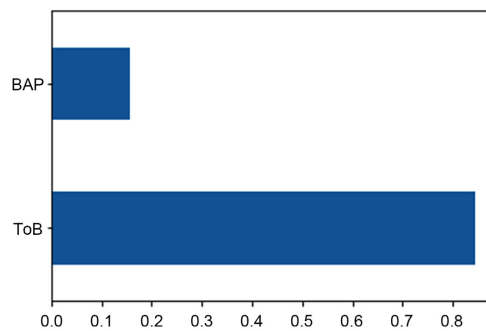


Figure 16. Results of ExtraTreesClassifier for business associate present and type of breach.

```

predicts_y = model.predict(X_test)
print(predicts_y)
✓ 0.3s

[4 4 4 2 4 2 2 5 2 2 6 4 6 4 4 6 2 6 2 4 2 2 2 2 4 6 6 6 6 2 4 6 6 6 4 6 2
 2 6 2 2 6 6 6 4 6 6 4 6 2 4 6 2 6 4 4 2 2 6 6 2 5 6 4 2 2 4 6 6 6 6 6 6 2
 2 2 6 6 4 2 6 2 6 2 6 2 4 6 6 2 2 4 6 6 2 2 4 6 6 6 6 2 6 2 2 2 6 2 6 6 4
 2 6 6 6 6 4 6 6 3 4 5 2 6 4 6 2 6 2 5 4 2 4 6 6 2 2 4 2 2 5 6 2 6 6 6 6 2
 2 6 2 4 6 2 6 6 2 4 6 2 6 5 2 6 6 6 2 4 6 6 6 4 2 2 2 4 4 6 4 4 4 6 4 6
 4 6 4 2 6 2 6 6 2 4 4 4 2 2 2 6 6 2 6 4 2 2 2 6 2 2 2 2 4 6 6 2 6 2 2 2 6
 6 6 6 2 2 2 6 2 2 2 2 2 2 2 2 6 4 2 2]

print(predicts_y.mean())
✓ 0.4s

4.053941908713693
    
```

Figure 17. Prediction results of naïve bayes classifier.

Based on the results of the “ExtraTreesClassifier” we decided to explore further the implications of no increase of information assurance in the human service sector. Naïve Bayes Classifier is a set of supervised learning algorithms based on applying Bayes’ theorem with the “naive” assumption of conditional independence between every pair of features given the value of the class variable. Using python for Naïve Bayes Classifier we were able to predict the location of data breaches. The locations were coded as follows:

- 1 = Desktop Computers;
- 2 = Electronic Medical Record;
- 3 = Email;
- 4 = Laptop;
- 5 = Network Server;
- 6 = Other;
- 7 = Paper/Films.

The classifier predicted next several locations of breaches that will occur to be notified to HHS and the average location of the breach will be “Laptops” based on the HHS breach notification data. **Figure 17** shows the prediction result.

5. Conclusions

Data breaches are a huge problem in the information systems sector but are even a bigger problem in the health services sector, where data is more sensitive and protected by many laws. This paper proposes a solution to this problem by implementing an information assurance approach to stop data breaches in their tracks. Information assurance provides a way to protect the data and assure that only authorized persons have access to it, which is a key idea when dealing with health records and data. Proven by the hacking reports mentioned above, the hacker breached over 50 percent of the providers in a three-year period. This attack could have been avoided if the right measures were taken and applied to accurately protect the data.

The steady rise of ransomware attacks makes information assurance worth the trouble of implementing something new in hopes of reducing the increase in attacks on data. The costs associated with the implementation will benefit overall from not having information assurance and having payout costs from the damage of the continuous data breaches. The Health Insurance Portability and Accountability Act being signed into law had the right idea but did not have the appropriate tools to ensure that it could be properly implemented, and this is where information assurance can protect what the bill set out to achieve. The act could have never predicted that medical records would take a digital turn but implementing information assurance can help bridge the gap between hackers and them not being able to penetrate data that should be protected.

Information assurance is not a complete solution to the problem of data breaches in the health services sector, but it can be said that information assurance is a step in the right direction. The inferential statistics included in the paper show the

difference and the continuing struggle and issue that would exist if it were not applied to the health services sector. As information assurance advances, hackers' ability to find new ways to breach the system and gain access to sensitive data reduces. With the majority of the hacking instances happening with email and then network servers, there is an urgent need to improve data security and protection in emails and network servers. This calls for an improvement in data security and protection. It can be said considering that most people in the world use emails to connect and communicate, this is especially true within healthcare and an increase in information assurance can be a great solution.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Nist (n.d.) Information Assurance (IA)-Glossary. CSRC. https://csrc.nist.gov/glossary/term/information_assurance
- [2] Lohrmann, D. (2021) 2020 Data Breaches Point to Cybersecurity Trends for 2021. GovTech. <https://www.govtech.com/blogs/lohrmann-on-cybersecurity/2020-data-breaches-point-to-cybersecurity-trends-for-2021.html>
- [3] Petcu, A.G. (2021) Ransomware Payouts in Review. Highest Payments, Trends & Stats. Heimdal Security Blog. <https://heimdalsecurity.com/blog/ransomware-payouts/>
- [4] Johnson, S. (2019) Safeguarding against Data Breaches. UTHSC Digital Commons. <https://dc.uthsc.edu/cgi/viewcontent.cgi?article=1065&context=hiimappliedresearch>
- [5] Ronquillo, J.G., Winterholler, J.E., Cwikla, K., Szymanski, R. and Levy, C. (2018) Health IT, Hacking, and Cybersecurity: National Trends in Data Breaches of Protected Health Information. *JAMIA Open*, 1, 15-19. <https://doi.org/10.1093/jamiaopen/ooy019>
- [6] CAST Publications (n.d.) Information Assurance. Software Intelligence for Digital Leaders. <https://www.castsoftware.com/glossary/information-assurance>
- [7] Lord, N. (2018) Information Protection vs. Information Assurance: Differentiating Between Two Critical IT Functions. DataInsider. <https://digitalguardian.com/blog/information-protection-vs-information-assurance-differentiating-between-two-critical-it>
- [8] HHS.gov (n.d.) Breach Notification Rule. <https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html>
- [9] Wabo, B. (2021) 14 Ways to Prevent Data Breaches in Your Organization. <https://www.vigilant.us/news/2017/2/21/14-ways-to-prevent-data-breaches-in-your-organization-credit-a-lign>
- [10] U.S. Department of Health and Human Services (1.1.2018-12.31.2020) U.S. Department of Health and Human Services Office for Civil Rights. https://ocrportal.hhs.gov/ocr/breach/breach_report.jsf