THE ROLE OF PROTOCOL ANALYSIS IN CYBERSECURITY: CLOSING THE GAP ON UNDETECTED DATA BREACHES

by

James Garringer

A Capstone Project Submitted to the Faculty of

Utica College

December 2018

in Partial Fulfillment of the Requirements for the Degree of

Master of Science in
Cybersecurity

Abstract

Organizations of all sizes are targets for a cyberattack. Undetected data breaches result in the catastrophic loss of personally identifiable information (PII) causing considerable financial and reputation harm to organizations, while also imposing a risk of identity fraud to consumers. The purpose of this study was to consider the impact that undetected data breaches have on organizations with an additional focus on shortening the gap between the time of data breach and the time of detection through manual protocol analysis and intrusion detection system (IDS) solutions. This research reviewed the available literature detailing the effects of undetected data breaches on organizations as well as the advanced exploitation of protocols and anomaly detection through manual protocol analysis and IDS.

Manual protocol analysis provides situational anomaly detection when compared to baseline network traffic, but implies privacy concerns and does not allow timely detection of most cyberattacks. Automated IDS stream-based flows allow quicker detection of cyberattacks. Network flow-based IDS misses hidden attacks due to lack of a data payload requiring manual analysis instead, while host-based IDS adversely affects the performance of the host computer, but successfully identifies anomalies based on known signatures. This study recommended a complementary defense-in-depth solution which employs manual protocol analysis and both host-based and network-based IDS solutions as a viable strategy for reducing the time between data breach and time of detection. This study additionally recommended that security operation center personnel and IT departments should receive protocol analysis training to support manual detection against a known network traffic baseline.

Keywords: Cybersecurity, Professor Donnie Wendt, Wireshark, self-similarity, mosaic effect, machine learning, incident response.

Acknowledgements

I wish to express my gratitude to the faculty and staff of Utica College for their professionalism, rigor, and desire to impart knowledge unto others. I would especially like to thank Professor Donnie Wendt who was instrumental in my conducting and presenting this research. Doctor Austen Givens has provided great encouragement for this study and my learning overall throughout the program. Dr. Givens' expert insight into critical national infrastructures has opened a new world to me. I also wish to express my deepest gratitude to Mark Low for his expertise and many late-night conference calls. I am grateful to my cohorts, especially Robert Hicks and Tristan Ross for their collegiality and thought-provoking conversations.

I am grateful to my mentor, Rick Murphy, who many years ago exposed me to a world of protocol analysis. I have anchored my career and academic pursuits in this area, which all began with Rick's encouragement and many theoretical and practical conversations. I am forever in his debt.

Finally, I wish to thank my family and friends. I thank my friends for their encouragement, and for not forgetting that I exist after being unavailable for quite a long time. I am grateful to my parents, John, and Deborah for their constant encouragement throughout my academic and professional careers. My son, James, has always been a great conversationalist and curious inquisitor; my sincerest thanks to him for countless coffee-fueled discussions which have always sought to push our knowledge.

Table of Contents

List of Illustrative Materials

Introduction

Standalone computers are useful for many purposes, but interconnection with other computers dramatically expands that usefulness. Network security problems began when communication and information sharing between computer systems became necessary. The earliest computer networks communicated by way of dedicated pathways referred to as circuits. An example of a circuit-switched network is the telephone switched network (Leiner et al., 2009).

Circuit-switched networks had several limitations, which precluded growth to the density and worldwide sprawl that is typified by today's Internet. Because circuit-switched networks lacked redundant links, a failure anywhere in a connection would cause the system to fail. When faced with noise or full channel capacity, circuit-switched networks would fail to transmit the original message reliably (Kleinrock, 1961).

Claude Shannon (1948) first defined a channel as the medium used to transmit a message; and suggested a radio frequency band, wire, or light beam as examples of potential communication channels. Shannon's definition of a communication system laid the groundwork for all methods of transmitting and receiving messages from a source to a destination. In Shannon's communication system model, a transmitter sends a signal to a receiver. In the midst of the channel is a potential noise source that may affect the reliability of the transmitted and received signals being equal. Further, the maximum number of bits that can be successfully sent and received by the system during a given timeframe describe the channel's capacity.

Leonard Kleinrock (1961) first envisioned a solution to the limitations of circuit-switched networks and addressed both singular links and available channel capacity in the system. In Kleinrock's packet-switched network, transmitted packets may take one of many possible paths between the source and destination rather than relying on dedicated circuits. When necessary, packets may briefly be stored at the source and then forwarded through the system to allow for channel capacity to become available. This dynamism provides a framework for interconnected networks, which may have computer nodes with different characteristics, such as link speed.

J.C.R. Licklider was the first director of the Defense Advanced Research Project Agency (Leiner et al., 2009). Licklider (1963) composed a now seminal memorandum addressed to Members and Affiliates of the Intergalactic Computer Network on April 23, 1963. The goal of the memo was to collect the commonalities of many individuals and projects who were interested in furthering the benefit of information processing and advancement in human intellect. Licklider (1963) further commented that lacking a formal name for the idea of a globally interconnected collection of computers that could be used to share information and resources he devised the name as entitled.

While working for the National Physical Laboratory in the United Kingdom, Donald Davies (1966) first referred to packaging a message for transmission onto a network. A packet, as defined by Davies, includes small chunks of user data along with source and destination addressing information. In his proposal for a national communication network based on packet-switching, Davies (1966) described organizing data into packets to accommodate differing network speeds by storing the outbound packet in memory long enough to determine available network capacity before transmitting. Paul Baran (1964) of RAND Corporation also developed a store-and-forward mechanism for his distributed communications network concept, which would have applicability for military voice networks. Both Davies and Baran stated the redundancy of paths added to the security of the network. Davies' (1966) proposal hinted at the possibility of link diversity being useful in cryptography.

According to Leiner et al. (2009), the first message sent through a connected network consisted of two connected computers on the Advanced Research Project Agency Network (ARPANET). The ARPANET was a United States (US) Department of Defense research project that was influenced in concept by the early work of Kleinrock. The precursor to the modern-day packet switch was the interface message processor (IMP). The IMP was the packet switch used by the first two, and subsequent, ARPANET hosts. The first installation of an IMP was at the University of California, Los Angeles, and the second was at Stanford Research Institute. These two packet switches and their attached host computers exchanged the first computer networked message in September 1969 (Leiner et al., 2009).

An increase in usage for the sake of application and information sharing began as early as the third and fourth computer systems attached to ARPANET. The University of California, Santa Barbara, and the University of Utah were the next two nodes added to the ARPANET with a focus on displaying mathematical functions and rendering of three dimensions over the network (Leiner et al., 2009). The use of the ARPANET grew to more than fifty host computers and forty packet-switching computers by May 1974 after implementing a new internetworking protocol known as the Network Control Program (NCP) (Kleinrock & Naylor, 1974).

Computers communicate with each other by adhering to a defined set of rules, known as a protocol, when sending and receiving messages (Cerf & Kahn, 1974). Vinton Cerf and Robert Kahn laid the groundwork for allowing remote packet-switched networks to communicate with one another in 1974 by defining a protocol framework for a Transmission Control Program. A limitation of the early NCP that the work by Cerf and Kahn (1974) sought to address was that it could only communicate with the destination interface message processor. The internetworking strategy defined by Cerf and Kahn (1974) introduced the concept of an internetwork gateway that would allow communication between dissimilar hardware and software. This strategy gave birth to what would eventually become the Internet as it is known today (Cerf & Kahn, 1974; Leiner et al., 2009).

Handling variable size data payloads and allowing data fragmentation into multiple datagram transmissions is a fundamental requirement of packet-switched networks. Reassembly of datagram fragments occurs at the destination network (Cerf & Kahn, 1974). Cerf and Kahn modeled the design requirements for the Transmission Control Protocol / Internet Protocol (TCP/IP) protocol suite after the original Transmission Control Program separating the functions into two protocols; the Transmission Control Protocol (TCP), and the Internet Protocol (IP). The TCP provides a connection between hosts initiated by a process on one computer and connecting to a process on another computer (Postel, 1981b). The Internet Protocol (IP) breaks application data into smaller chunks suitable for transmission on packet-switched networks; these smaller chunks are called datagrams (Postel, 1981a). The combination of TCP and IP provides a reliable network connection and addressing for the transmission of data through interconnected networks on the Internet or a local packet-switched network occurs (Postel, 1981a, 1981b). A 16-bit security field designed to classify an IP packet may be used to identify a US Department of Defense security classification. Apart from this security field, there is no built-in addressing or connection-level security inherent within the TCP/IP protocol (Postel, 1981a, 1981b).

Protocol headers for both the TCP and IP protocols include source and destination addresses among other required fields that help to identify and deliver a packet on the network (Postel, 1981a, 1981b). Protocol headers, which are interpreted and acted upon by a computer, switch, or Internet gateway, may also be captured and analyzed using protocol analyzer software on the path that packets will take. Protocol headers have numerous fields, which may contain arbitrary values that may be exploited by a cyberattacker. Craig Rowland (1997) demonstrated the inherent weakness in TCP/IP headers when used for covert channel data exfiltration.

According to Kleinrock and Naylor (1974) protocol analysis was first used as a tool to measure the success of the ARPANET. A data packet was traced through the ARPANET network to calculate, based on timestamps, network performance as a function of time. Myriad statistics were calculated based on the data packets passing through the ARPANET network, allowing for the creation of baseline performance indicators (Kleinrock & Naylor, 1974). Modern-day analysts leverage protocol analyzers in much the same way, using them to perform a review of packets for anomalies in header fields or payload contents. Significant differences in network traffic patterns from the baseline network operation are considered anomalies (Ahmed, Naser Mahmood, & Hu, 2016). Protocol analyzers allow network administrators a facility to capture network packets from the medium they are traversing. Once captured, the packets are decoded by the analyzer for review by a protocol analyst (Goyal & Goyal, 2017). Protocol analysts learn how the protocols are designed by reading the relevant protocol specifications and viewing baseline and unusual network traffic. The Request For Comments (RFC) Editor aggregates protocol specifications, such as TCP/IP, from the Internet Engineering Task Force, the Internet Architecture Board, the Internet Research Task Force, and other individual contributors (Internet Society, 2018).

A cyberattack is an unwanted activity, which affects the confidentiality, integrity, and accessibility of the victim's computer system, and is often illegal in the country or locale where the host computer resides. Cyberattacks which begin outside of an organization likely use the Internet, and therefore the TCP/IP protocol suite, to connect to and access a computer with malintent. Detecting the occurrence of a cyberattack or related network breach may be accomplished by manual protocol analysis or through an intrusion detection system, which automates detection based on packet inspection for anomalies or network traffic flow analysis (Golling & Koch, 2014).

Background

Cyberattackers perform various attacks against victim networks, which cause damage regarding any combination of confidentiality, integrity, and availability of victim networks and their data. The final event tied to a breach will compromise the confidentiality of an organization's data assets more than 80 percent of the time when a breach exists for the full term of the attacker's intended duration (Verizon, 2018). Data confidentiality is compromised only up to 32 percent of the time when the breach is not carried to term (Verizon, 2018).

Organizations must respond to incidents quickly to lower the costs of remediating a breach or other cyberattack. Successful remediation also requires training for responders and the ability to analyze incidents quickly (Ponemon Institute, 2018). In the US, it is unlawful to access a computer or network without authorization. Further, it is illegal to purposefully transmit a program or command which will cause damage to the remote computer (18 U.S.C. § 1030, 2009). Attackers who successfully breach a network may attempt to maintain access with remote command and control software in the form of installed malware. Once installed, the malware may operate undetected by the victim organization. There is considerable literature for review relating to the role protocol analysis has in detecting cyberattack breaches yet the time to detection of breach remains high and nearly unchanged for the last three years (Ponemon Institute, 2016, 2017, 2018).

Statement of the Problem

Two-hundred days pass, on average, before victimized organizations detect the presence of malicious software installed as part of a cyberattack (McConnell, 2017). Lengthy periods of time between initial breach and mitigation provide attackers ample time to move laterally within target networks, exfiltrate and possibly destroy data, or disrupt normal network operation. China's Advanced Persistent Threat 1 (APT1), as named by Mandiant, exfiltrated hundreds of terabytes of data while maintaining network access for an average of 356 days. In an extreme case, APT1 maintained access without detection within the victim network for nearly five years (Mandiant, 2013).

According to Verizon (2018), there were 53,000 cyberattack related incidents compromising the confidentiality, integrity, and availability of data in organizations represented by the 2018 Verizon Data Breach Investigations Report, a report which analyzed incident and breach data for 2017 as reported by 67 contributing entities. Cyberattack incidents in 2017 resulted in 2,216 confirmed data breaches requiring a response from affected organizations. Sixty-eight percent of reported network breaches took months or longer to be detected. It is important to note that the Verizon report does not consider the use of pilfered credentials via malware-captured login details as a breach; instead, these numbers reflect only network access vectors that progress from incident to breach through non-credential exploitation. Pilfered credentials account for an additional 43,000 breaches in 2017 (Verizon, 2018).

Each data breach comes at a cost to affected organizations of between $6 million and $10 million to remediate according to Admiral Mike McConnell, former director of the National Security Agency (McConnell, 2017). The Ponemon Institute (2018) asserted that the cost of data breaches occurring in 2017 was an average of $3.86 million worldwide and $7.91 million in the

United States where data breaches are the most costly. Organizations of all sizes are targets for cyberattack. Further, the amount of personal data stored by an organization increases in volume and value as the size of the organization, by market capitalization, increases making larger organizations frequent targets of cyberattack (Wheatley, Maillart, & Sornette, 2016). Small businesses are also targets for a cyberattack. According to Verizon's Data Breach Investigations Report, more than 1,285 reported data breaches targeted small businesses, representing 58 percent of all reported breaches for 2017 (Verizon, 2018).

Undetected cyberattacks contribute to the rising cost of incident response. The public often perceives data exposure due to breach as a negligent phenomenon resulting in damage to an organization's reputation and loss of current and future customers further contributing to the cost of an undetected breach (Ponemon Institute, 2018). Organizations may mitigate their exposure to damage and reduce mitigation costs by responding to an incident quickly. This ability is developed by detecting and analyzing incidents as they occur (Ruefle et al., 2014).

When approaching the problem of security in an organization, cybersecurity analysts assume that breach of organization networks has already occurred. Organizations must assume that a breach has already occurred, or is imminent, when developing a cyber incident response plan (Densham, 2015). This research study examines the role of protocol analysis in cybersecurity and how it may contribute to closing the gap between the time of an attack and time of detection. This research will be of interest to organizations of all sizes and will reduce the frequency of undetected cyberattacks.

Purpose of the Study

The purpose of this study is to identify the techniques required for organizations to detect cyberattacks faster. This research will discuss the ramifications of undetected cyberattacks as they pertain to organizations under attack with a focus on available supporting literature. An examination into methodologies used to manually detect cyberattacks through protocol analysis and their viability when considered as part of incident response will be included. Finally, this research will discuss the application of intrusion detection and prevention systems to the timely detection of cyberattacks, lessening the gap between the time of the incident and the time of detection of cyberattacks.

Research Questions

This research will seek to address the following questions:

Q1. How are organizations affected by cyberattacks that are not detected before damage occurs?

Q2. How are cyberattacks and network anomalies manually detected?

Q3. How do intrusion detection and prevention systems contribute to the detection of cyberattacks?

Literature Review

In the past several years, various authors have contributed to the discussion around cyberattacks and the ramifications of undetected data breaches. This research analyzed sources of scholarly writing, journal articles, and news sources to determine the effects of undetected cyberattacks on organizations. This research further examined extant research in the areas of anomaly detection and automated intrusion detection.

Undetected Cyberattacks

The Ponemon Institute (2018) defined a data breach as an event that potentially compromises a record of information which may identify a person and their financial or medical records. A compromised record, according to Ponemon (2018), is a record which has been exfiltrated by an attacker during a data breach. Rid and Buchanan (2015) noted that cyberattacks might include a data breach or reconnaissance activities performed by an attacker.

In recent years, there have been significant data breaches, which have included the personally identifiable information (PII) of millions of consumers. Personally identifiable information may include name, address, social security number, driver's license, and financial and medical records (Gupta, 2018). The Privacy Rights Clearinghouse (2018) compiled a database of categorized breach data taken from the Office of Civil Rights within the US Department of Health and Human Services and various media sources. According to the Privacy Rights Clearinghouse (2018), between 2014 and October 2018, there were 5.58 billion compromised records as a result of cyberattacks and hacking. While there were 796 data breach incidents which involved one or more compromised records spanning the US, the top 10 data breaches involving hacking accounted for more than 89 percent of all exposed records during the five-year span (Privacy Rights Clearinghouse, 2018). The top 100 data breaches (see Appendix

A for a complete listing) during the same period account for 5.57 billion exposed records leaving 8.3 million records for all other breaches combined (Privacy Rights Clearinghouse, 2018).

Table 1

Top 10 Data Breaches 2014 – 2018

Company	Records	State	Date Made Public
Yahoo!	3,000,000,000	California	December 14, 2016
Yahoo!	500,000,000	California	September 22, 2016
FriendFinder	412,000,000	California	November 16, 2016
MySpace	360,000,000	California	May 31, 2016
Under Armour	150,000,000	California	March 30, 2018
Equifax Corporation	145,500,000	Georgia	September 7, 2017
Ebay	145,000,000	California	May 21, 2014
LinkedIn	117,000,000	California	May 17, 2016
Anthem	80,000,000	Indiana	February 5, 2015
J.P Morgan Chase	76,000,000	New York	August 28, 2014
Total	4,985,500,000		

Note. Results of a search for exposed records related to hacking-only data breaches occurring between 2014 and October 2018. Adapted from “Data breaches,” by the Privacy Rights Clearinghouse, 2018.

According to Edwards, Hofmeyr, and Forrest (2016), the dataset provided by the Privacy Rights Clearinghouse for the period from 2005 through 2014 suggests that overall annual breaches remain relatively flat year over year. Furthermore, their research showed that both the frequency of data breach and the number of data records compromised continues to remain the same as of September 2015. Edwards et al. (2016) additionally theorized that the relative year over year consistency in both data breach size and frequency is likely due to a simultaneous improvement in security practice and attacker technique.

Undetected data breaches yielding the most compromised records per year as maintained by the Privacy Rights Clearinghouse (Privacy Rights Clearinghouse, 2018) are recounted in

Table 2. The Yahoo! data breach in 2016 was the largest reported compromise, which accounts for 53.7 percent of all compromised records during the period 2014 through 2018. Table 2 lists the top data breach for each year during the same period and accounts for 63 percent of all compromised records (Privacy Rights Clearinghouse, 2018).

Table 2

Top Annual Data Breach 2014 – 2018

Company	Records	State	Year
Under Armour	150,000,000	California	2018
Equifax Corporation	145,500,000	Georgia	2017
Yahoo!	3,000,000,000	California	2016
Anthem	80,000,000	Indiana	2015
Ebay	145,000,000	California	2014
Total	3,520,500,000		

Note. Results of a search for exposed records related to hacking-only data breaches occurring between 2014 and October 2018. Adapted from "Data breaches," by the Privacy Rights Clearinghouse, 2018.

Use of stolen data. Abhishek Gupta (2018) used the moniker *Identity Theft 2.0* to describe the evolution of identity theft, which leverages stolen records to learn more about an individual. Customers of data brokers collect consumer PII and share the data with data brokers. Data brokers merge the collected consumer PII with the data collected by other data broker customers. This merged PII data is, in turn, sold back to companies seeking to bolster the type of data they have available about their customers. Large-scale data breach data is used to fill in gaps in consumer profiles (Gupta, 2018).

Ed Felten, Professor of Computer Science at Princeton University, participated in the Privacy and Civil Liberties Oversight Board's Defining Privacy Forum in 2014. Felten framed the activities surrounding managing information about consumers as a trio beginning with data collection followed by merging of data, and finally by inferring consumer behavior by modeling with predictive analysis (Medine et al., 2014). Organizations routinely record data about their customers. Data collection includes items that are overtly disclosed by consumers as well as implicit items based on consumer behavior both online and offline (Medine et al., 2014). Merging existing databases with newly available datasets allow linking of user behavior to identity wherever possible to determine a common identifying record such as a social security number (SSN). Felten referred to the process of linking records about an individual and subsequently inferring new meaning based on the new information as the mosaic effect. Gupta (2018) further cautioned that advances in artificial intelligence create a mosaic effect, allowing for the merging of multiple records for the same individual, which is especially worrisome. Credit scores may be negatively affected by fraudulent credit activity, hate groups may target individuals for persecution, and unwanted release of negative information are potential concerns for those included in a breach dataset (Gupta, 2018).

Cost of a data breach. The Ponemon Institute (2018) has released its *Cost of Data Breach Study* analysis annually since 2005. The report began with US-based organizations, but the breadth of the report has grown to represent organizations in fifteen countries or regions. Organizations who have experienced a breach incident participated in the Ponemon Institute's benchmark study. Organizations who participated in the 2018 benchmark survey are from Australia, Brazil, Canada, France, Germany, India, Indonesia, Italy, Japan, Philippines, Malaysia, Saudi Arabia, Singapore, South Africa, South Korea, The Middle East, Turkey, United Arab Emirates, United Kingdom, and the US. Participating countries increased gradually from 10 to 15 between 2014 and 2018 (Ponemon Institute, 2014a, 2015, 2016, 2017, 2018). The average cost among study participants for the five-year period from 2014 to 2018 was

$3,758,000. Table 3 shows each year's average cost (Ponemon Institute, 2014a, 2015, 2016, 2017, 2018).

Table 3

Average Cost of Data Breach

Year	Average Cost
2018	$3,860,000
2017	$3,620,000
2016	$4,000,000
2015	$3,790,000
2014	$3,520,000

Note. Average cost of data breaches per year for the period 2014 through 2018. Adapted from "2014 Cost of Data Breach Study: Global Analysis," by the Ponemon Institute, 2014, "2015 Cost of Data Breach Study: Global Analysis," by the Ponemon Institute, 2015, "2016 Cost of Data Breach Study: Global Analysis," by the Ponemon Institute, "2017 Cost of Data Breach Study: Global Analysis," by the Ponemon Institute, 2017, "2018 Cost of Data Breach Study: Global Analysis," by the Ponemon Institute, 2018.

According to Experian (2018), failure to comply with regulations introduces additional costs to organizations in the form of fines. New rules imposed by the General Data Protection Regulation (GDPR) in May 2018 provide requirements for the processing, storage, and security of data belonging to European Union citizens. Fines imposed for those who fail to comply with the GDPR may be up to four percent of the preceding year's annual revenue or $23 million, whichever is greater. Rules imposed by the GDPR apply to all companies who serve customers in the European Union (Experian, 2018).

According to the GDPR, the use of personal data requires documentation which does not change (European Commission, 2016). The collection of personal data must be for a specific, explicit, and legitimate purpose. The purpose of data collection may not change except for archival activities that may benefit the public interest, scientific or historical research, or statistical analysis (European Commission, 2016). Processing of personal data may not be used to benefit the controlling company if processing would violate the fundamental rights of the person who provided the data; this is especially true if the person is a child (European Commission, 2016). Certain protected data types require permission of the person to which the data applies. A person who divulges personal information which would disclose race, political leanings, religion, genetic or biometric data with the purpose of identifying that person, their sexual orientation or behaviors is prohibited unless that person provides express permission (European Commission, 2016).

Cost models. The Ponemon Institute (2018) has implemented an activity-based cost model where certain types of activities are assigned a value based on the collected responses from survey participants. The discovery, response, and post-breach recovery activities reflect the cost of data breaches to organizations. Direct and indirect costs along with opportunity cost are considerations asserted by Ponemon when calculating total cost (Ponemon Institute, 2018). Direct costs are costs which are incurred with cash, whereas indirect costs are costs which reflect lost employee time and effort while performing breach remediation activities. The cost of an investigation, identifying victims of exposed data, communicating with, and notifying, victims and regulatory entities, and internal training and incident handling preparation are costs associated with initial incident discovery and response. Post-breach discovery costs may include legal fees for defense and regulatory compliance, consulting services, identity-theft protection and monitoring services for victims, and loss of customer business or loyalty (Ponemon Institute, 2018). The Ponemon model seeks to attribute cost for data breaches up to 100,000 compromised records; this model does not apply directly to mega breaches with one million to fifty million compromised records or more. The Ponemon Institute (2018) suggested the cost of a mega-breach to be $40 million for one million compromised records, and $350 million for fifty million records.

Jay Jacobs of Data Driven Security (2014) concluded that the Ponemon Institute's per capita cost model is simplistic as it accounts for total losses divided by the number of compromised records. Jacobs stated the Ponemon model might be used to apply a simple estimate of potential loss by multiplying a fixed dollar value from the latest report by the number of records at risk of potential loss. Actual losses, however, are more accurately described by log-log linear regression analysis of the Ponemon supplied data where the log of the number of records lost and the log of financial loss are used to determine a predicted loss that is more accurate than that of the Ponemon Institute model (Jacobs, 2014). Jacobs' formula is log(financial loss) = 7.68 + 0.76*log(records compromised). While the results are more accurate than the Ponemon model, Jacobs (2014) further concluded that there are more factors to be considered beyond the number of records lost in a data breach. Natural fluctuations and uneven variances are factors that preclude the use of a linear or linear regression model to represent the change in records lost year over year.

Public opinion. The Ponemon Institute (2018) reported that loss of customer trust and subsequent customer retention is dependent on the industry in which the breach occurred. Referring to the metric as abnormal customer churn, or loss of business, Ponemon cautioned that healthcare and financial industries are the most susceptible to customer churn followed by pharmaceuticals, services, and technology industries. Higher customer churn rates contribute negatively to the overall cost of a data breach. As a result of the benchmark survey, Ponemon suggested that organizations with less than a one percent abnormal churn rate had an average breach cost of $2.7 million versus a $4 million cost for a four percent abnormal churn rate (Ponemon Institute, 2018).

In May of 2015, Lillian Ablon, Paul Heaton, Diana Lavery, and Sasha Romanosky (2016) of the RAND Corporation surveyed 6,000 adults who participated in the American Life Panel, a panel designed to represent adults in the US. The focus of the survey was on the frequency of breach notification, the type of data compromised, and subsequent consumer response. The representative panel allowed estimation at the scale of the full US. Twenty-six percent of respondents received a breach notification in the prior year. As a result, Ablon et al. estimated that 64 million US adults, or one-quarter of the US population, received a breach notification in the same period. Ablon et al. estimated that 36 million US adults received two or more breach notifications in the same year. Respondents reported an average personal cost of $500 in recovering from a breach. Seventy-seven percent of respondents were happy with the way the reporting company handled the breach, and 89 percent of respondents continued to do business with the company (Ablon et al., 2016).

The Ponemon Institute (2014b) surveyed 797 individuals in 2014 to measure consumer sentiment toward organizations who lost their data. Of those surveyed, approximately 400 were victims of a data breach. While 32 percent of respondents ignored the breach notification, 29 percent accepted offers of free credit monitoring services, and 48 percent of respondents felt that they are at risk of identity theft. Consumer sentiment toward how organizations handle a breach included feeling that identity theft protection and credit monitoring services should be available to victims (Ponemon Institute, 2014b).

Romanosky, Hoffman, and Acquisti (2014) performed an analysis of litigation rulings over cases related to consumer data compromise, and concluded companies who notify consumers of a data breach and offer free credit monitoring services are six times less likely to be sued. In addition, the loss of financial PII increased the likelihood of litigation against the organization six-fold. When plaintiffs cite a cause for statutory damages against organizations who lose PII, defendant organizations are more likely to settle out of court. As cited by Romanosky et al. (2014), the Computer Fraud and Abuse Act allows for $5,000 in statutory damages per record lost.

Breach notification laws. According to the National Conference of State Legislatures (2018), all fifty states in the US have breach notification laws requiring individual notification when a breach has occurred. The California Information Practices Act of 1977 (California Code, 1977) requires that data breach victims whose PII has been lost or stolen be notified immediately upon discovery unless delayed by law enforcement investigation. The New York State Breach Notification Act of 2005, as cited by FindLaw (FindLaw, 2005), requires notification to New York residents as quickly as possible and without delay, but does allow notification delays for law enforcement investigation.

The 115th US Congress (2017) proposed a bill, the Data Security and Breach Notification Act, which will provide a uniform requirement in the US for consumer notification. The bill was read twice and referred to the Committee on Commerce, Science, and Transportation as of November 2017. If the bill becomes law, it will require consumer notification of a data breach affecting their PII within 30 days of discovery. The proposed legislation would protect citizens and residents of the US. According to Congress, waiver of the 30-day notification requirement is possible only when the data-controlling organization can prove that additional time is necessary to identify affected consumers accurately, prevent further unauthorized PII disclosure, or to remediate the integrity of the data system. Notification to credit reporting agencies is also required by Congress when the number of affected consumers is more than 5,000 records (115th Congress, 2017).

Incident response. Matt Ehrlich (2017) suggested that planning to manage data breaches is more realistic than preventing them. Fraud based on leaked or stolen PII is assumed imminent following a data breach. Detection of a data breach may occur after stolen data appears on the dark web or years after remediation of the breach. Ehrlich continued to suggest a three-pronged approach to strengthening data breach protection. First, a culture of security and privacy awareness through employee training will strengthen internal security practices. Second, attempt to track the attacker's activity, and the presence of data leaked on the dark web, to learn where organizational vulnerabilities exist. Third, organizations should provide fraud protection services, such as credit monitoring, to consumers who have been affected by a data breach. This three-pronged approach can help prevent reputation and financial harm to the organization (Ehrlich, 2017).

According to Paul Cichonski, Tom Millar, Tim Grance, and Karen Scarfone (2012) of the National Institute of Standards and Technology (NIST), learning and improvement after an incident are essential functions within incident response handling. Detecting the occurrence of an incident may be challenging. Methods of detecting incidents include automated systems such as intrusion detection systems (IDS) implemented at the network and host level, anti-virus, and log analysis software. The volume of reports from IDS, however, can impede quick detection. Computer user reports may also spur manual detection by response teams. Finally, a lack of training, in-depth technical knowledge, and experience may hamper incident detection and response (Cichonski et al., 2012).

The Federal Trade Commission (FTC) (2016) recommended crafting an incident response plan which includes a communication strategy for all stakeholders. Organizations should not make misleading statements about a data breach. Furthermore, the FTC recommended efforts be made to protect customers and educate them about the extent to which their exposed data could harm them, is prudent (Federal Trade Commission, 2016).

Corey and Wilsker (2015) considered the legal responsibilities of organizations in their state of New Hampshire to include assembly of an incident response team with executive leadership authority, information technology personnel capable of assessing the damage and extent of a breach, and public relations specialists to manage communication with the public and other stakeholders. Organizations are required to report a data breach to the Attorney General's office or the authority over any regulated industry such as the financial sector. The response plan should include notification of individual affected consumers, and the notification to those users should comply with any breach notification requirements in their state of residence. To that end, it is imperative that organizations plan for adequate response in all states in their customer base (Corey & Wilsker, 2015).

Anomaly Detection

Ahmed and Mahmood (2014) defined network traffic analysis as a means to infer patterns from network protocol communication. Preventing disruption to an organization's network can be facilitated through the proactive use of protocol analysis and network traffic anomaly detection. According to Ahmed and Mahmood (2014), network traffic anomalies are separated into three categories: point, contextual, and collective anomalies categorize network traffic anomalies. A point anomaly describes a dataset in which one data point is different from the remainder of the dataset, whereas a contextual anomaly occurs when a dataset appears to be out of the ordinary only when in a certain context. Collective anomalies are described as a subset of an entire dataset which behaves outside of what is expected overall, but when examined as single data point appear to be normal (Ahmed & Mahmood, 2014).

According to Sestito et al. (2018), anomaly detection must be preceded by an understanding of the normal traffic patterns on the network. Normal traffic is defined as the traffic that is representative of the communication patterns of all processes operating on the network. An event that deviates to a high degree from the normal traffic model is considered anomalous. Barford and Plonka (2001) stated that four categories classify network traffic anomalies; irregularities related to network operations, flash-crowd anomalies, measurement failures, and attacks. Misconfiguration of the network interrupted connections, and problems with network equipment characterize irregularities in network operations. Flash-crowd anomalies are described as a spike in network traffic, typically to or from one computer or device, and which eventually subsides to normal patterns. Problems gathering data may result in measurement failures causing the status of the network to be inaccurate (Sestito et al., 2018). Anomalies in traffic which affect the confidentiality and integrity of information in the system are attacks.

Protocol analyzers. According to Singh, Kumar, Singla, and Ketti (2017), network sniffers may capture network traffic. There are many packet sniffers available including the open-source sniffer and protocol analyzer Wireshark, command-line based tcpdump sniffer, Kismet wireless sniffer, Ngrep command-line capture file analyzer, and NetworkMiner sniffer. Hong, Liu, and Govindarasu (2014) examined Wireshark, ColaSoft Packet Builder, and Nmap as freely available tools used to validate anomaly detection algorithms. Hong et al. (2014) compared the performance of network-based anomaly detection algorithms with the command-line Tshark protocol analyzer. Tshark allows capture of live network traffic or manipulation of saved capture files (Hong et al., 2014).

Wireshark. Gajendra Singh and Sandeep Baliya (2015) reviewed Wireshark as a tool for analyzing malicious network traffic. Wireshark allows filtering of all traffic captured by applying display or capture filters. Protocol dissectors interpret packets, which may be selectively shown by applying display filters. Capture filters limit the packets that are captured. Singh and Baliya demonstrated the creation of a traffic flow graph using the statistics feature of Wireshark. Based on analyzed traffic, specific IP addresses, transport protocols, or URLs were selectively blocked to protect the network from malicious traffic using the iptables firewall configuration utility (Singh & Baliya, 2015). Goyal and Goyal (2017) noted that Wireshark could save network traffic captures into several files.

Tcpdump. The freely available tcpdump command line tool was compared to Wireshark by Goyal and Goyal (2017). Tcpdump was found to consume less battery power, memory, and processor resources while Wireshark captured packets quicker. When capturing packets in monitor mode, tcpdump dropped up to three percent more packets than Wireshark. While capturing Ethernet packets, Wireshark captured up to one percent more packets than tcpdump (Goyal & Goyal, 2017).

Network attacks. Myriad network attacks such as denial of service, probes, user to root escalation, and remote to local user attacks exist as categorized by Ahmed and Mahmood (2014). Yang, Wang, Zhang, and Li, (2016) categorized detection techniques as either host-based or network-based. Host-based tools detect botnets, worms, and other threats. Due to the lack of scale, adverse performance effects on the host, and signatures which are compatible with network-based anomaly detection, host-based detection schemes alone are not enough to protect an organization from cyberattack. Yang et al. (2016) discussed network-based detection as a technique to discover cyberattacks based on network traffic analysis or honeypot activity. Drawbacks to network-based detection include the unlikelihood that honeypots can detect all threats and the prevalence of hidden malicious traffic among normal network flows (Yang et al., 2016).

Probes. Ahmed and Mahmood (2014) referred to network probes as a reconnaissance technique used to determine the functionality of a remote host. While not considered to be directly damaging, Ahmed and Mahmood stated that probes are a threat that should be taken seriously. Network probes may reveal valuable information about the network to an attacker. Khamphakdee, Benjamas, and Saiyod (2014) stated that attackers use freely available probing applications to perform reconnaissance activities against target networks. Information about the target network is collected by tools such as nmap, satan, and mscan. Collected information may be used for more sophisticated attacks including DoS, privilege escalation, and unauthorized local user access (Khamphakdee et al., 2014).

El-Hajj, Al-Tamimi, and Aloul (2015) explained that a three-way handshake establishes TCP connections. A TCP connection begins with a client sending a TCP packet with the synchronization (SYN) bit set. The receiving host responds with a TCP packet with both the SYN and acknowledgment (ACK) bits set if the port is open. A TCP packet with the RST bit set signifies a closed port. Finally, upon receiving the SYN/ACK packet, the client responds with an ACK packet completing the three-way handshake and establishing a TCP connection between the client and host. Common port scanning techniques include a connect scan, SYN scan, finish (FIN) scan, and ACK scan. The connect scan identifies an open port when the TCP three-way handshake completes (El-Hajj et al., 2015). When performing an SYN scan the attacker sends an

SYN packet, and when the requisite SYN/ACK packet is received the client sends an RST packet. An attacker may scan a network by initiating a FIN scan where the client sends a TCP packet with the FIN bit set. The server does not respond if the port is open, but transmits a TCP packet with the RST bit set if the port is not open. When an attacker sends a TCP packet with the ACK bit set without the prior two steps of the three-way handshake, for unfiltered ports, the firewall responds with a TCP packet with the RST bit set. Filtered ports on the firewall do not respond to ACK port scanning attempts. The SYN, FIN, and ACK scans are stealthy and not recorded in firewall logs (El-Hajj et al., 2015).

Denial of service. Singh and Baliya (2015) stated that the goal of DoS attacks is to disrupt the legitimate use of the network or computer resource. Tripathi and Hubballi (2018) presented new slow-rate DoS attack methodologies for the Hypertext Transfer Protocol version 2 (HTTP/2). Tripathi and Hubballi concluded that the HTTP/2 protocol is vulnerable to several DoS attacks. The HTTP/2 protocol, standardized in 2015, improved upon application bandwidth usage that its predecessor HTTP/1.1 did not implement. Slow-rate DoS attacks target free available connections on the victim web server. The proposed attacks involve a malicious client sending specially crafted HTTP packets to targeted servers. When compared to HTTP/1.1 the authors determined that HTTP/2 has more attack vectors, which can be exploited including both cleartext and encrypted HTTP requests (Tripathi & Hubballi, 2018).

Tripathi and Hubballi (2018) proposed five attack scenarios targeting servers running the HTTP/2 protocol. When compared to other DoS attacks, the required number of packets to impose DoS to the target was 150 packets versus 1,000,000 or 2,000,000 packets as proposed by other authors. Detecting an attack against web servers running the HTTP/2 protocol where the traffic is encrypted first requires that the data payload is decrypted. Tripathi and Hubballi

(2018) suggested the configuration of a proxy server to intercept the traffic creating a capture point where the traffic is not encrypted.

According to Daniel Stenberg (2015) Transport Layer Security (TLS) is optionally implemented for HTTP/2. Some web browser makers may require TLS in their HTTP/2 browser implementation including Mozilla's Firefox and Google's Chrome. Stenberg (2015) commented that developers had debated the need for requiring specific ciphers for HTTP/2 or including some weaker ciphers on a blacklist.

Tripathi and Hubballi's (2018) proposed slow-rate DoS attack methodologies manipulate HTTP/2 headers, GET and POST commands, and withholding responses and acknowledgments to web servers. When a malicious client sends a complete GET header but sets a zero value for the HTTP/2 protocol's SETTINGS_INITIAL_WINDOW_SIZE parameter, the web server assumes that the client cannot accept any packets and awaits a WINDOW_UPDATE message, which the malicious client does not send (Tripathi & Hubballi, 2018).

Additional proposed attacks included coercing the web server to be in a state where it is expecting the client to send a response, additional data, or an acknowledgment to the server (Tripathi & Hubballi, 2018). A malicious client seeking to perpetrate a DoS attack does not transmit the expected response, data, or acknowledgment. Web servers included in experiments were Apache, Nginx, H2O, and Nghttp2. Each server implementation allowed for a range of open client connections, between 150 and 2060, before exhausting resources and denying legitimate users access (Tripathi & Hubballi, 2018).

Malware activity. Botnet operators leverage fast-flux Domain Name System (DNS) services to provide a complex, load balanced, and highly available network (Yang et al., 2016). Katz, Perets, & Matzliach (2017) found that botnets use fast-flux networks to avoid the discovery of malware and its command and control (C&C) infrastructure. Fast-flux networks are characterized by frequently changing DNS nameservers, domain names, and host IP addresses. Malware hosting and delivery, communication, phishing, and web proxying are malicious activities that are protected from discovery by fast-flux networks. The most common ports utilized were standard web TCP ports 80 and 443 as well as TCP port 7547, which was found to be a common exploitable router port. Katz et al. (2017) concluded that manual detection of malicious activity using fast-flux networks is fruitless as the evidence collected for investigation changes quickly. Additionally, the effort to detect fast-flux networks should focus on algorithms capable of detecting the changing characteristics of the network (Katz et al., 2017).

Covert channel. Butler Lampson (1973), of the Xerox Palo Alto Research Center, first described a covert channel, in the context of inter-process bounding, as a channel that was not intended to carry information. Craig Rowland (1997) famously introduced proof-of-concept functionality in the covert_tcp application which demonstrated three distinct covert channels due to weaknesses in the TCP/IP protocol suite. Covert_tcp facilitates the transfer of data from source to destination in various header fields of the TCP and IP protocols, which accept arbitrary values such as the TCP Initial Sequence Number (ISN). Rowland (1997) discovered that the TCP three-way handshake is vulnerable to covert channel techniques, which involve manipulating unused or optional fields in the TCP or IP headers. Arbitrary sequence numbers encoded as ASCII characters and ACK packets are used to transmit covert channel data from one host running the covert_tcp in client mode to another host running in server mode (Rowland, 1997).

Mehic, Slachta, and Voznak (2016) stated that hiding data in covert channels include two distinct categories. Data may be hidden in the unused fields of protocol headers where arbitrary values are allowed or by encoding data in the behavior characteristics of the carrying protocol. Mehic et al. (2016) suggested that current network data hiding techniques include a frequently used information carrier, and a large amount of network traffic in which small amounts of hidden data exist and are transmitted repeatedly along with normal traffic flows. The authors commented that current intrusion detection systems could not process huge datasets in real-time, making detection unlikely (Mehic et al., 2016).

Wendzel and Keller (2014) defined a micro-protocol as having a protocol header stored inside the hidden data payload of a clandestine channel communication. Micro protocols add proxy, dynamic routing, and connection management characteristics to covert channels. Wendzel and Keller (2014) hypothesized an increase in the use of micro protocols by both botnets and malware in the future. Use cases for micro protocols include botnet C&C, covert journalist communication in Internet censored states, and military and secret agency covert communication. Wendzel and Keller (2014) surveyed known micro protocols and found varying degrees of required bits are required for micro protocols and their underlying protocols. The Ping Tunnel (PT) micro-protocol uses the Internet Control Message Protocol (ICMP) Echo Request and Echo Reply messages to bypass restrictive firewalls. An additional ICMP-based micro-protocol is the RM protocol named for its creators Ray and Mishra. The RM micro-protocol requires the smallest number of bits of all protocols surveyed. The dG micro-protocol separates the 16-bit UDP destination field into two halves requiring a sequence number and a data payload. The dG micro-protocol is useful for port-knocking covert channel communication (Wendzel & Keller, 2014).

Port-knocking, defined by Khader, Hadi, & Hudaib (2016), involves connection attempts to a predetermined sequence of closed ports on a listening server. When the correct sequence of connection attempts to closed ports is complete, the server evaluates the payload of the received packets. Table 4 lists the underlying protocols for each surveyed micro-protocol along with the number of bits used by the micro-protocol. Kaur, Wendzel, Eissa, Tonejc, and Meier (2016) concluded that micro protocols have headers that use more bits than are required making detection of the covert channel more likely.

Table 4

Micro-Protocols

Protocol	Underlying Protocol	Bits
Ping Tunnel (PT)	ICMP	192
dG	UDP	16
RM	ICMP	8
Covert File Transfer Protocol (CFTP)	IP	16
HyH	IP, UDP, RTP	Variable
Smart Covert Channel Tool (SCCT)	Various	Variable

Note. Bits required for covert communication in micro protocols and their underlying protocol. Adapted from "Anomaly-based network intrusion detection: Techniques, systems, and challenges" by Garcia-Teodoro, P., Diaz-Verdejo, J., Macia-Fernandez, G., & Vazquez, E., 2009, *Computers & Security*, *28*(1), 18–28.

Intrusion Detection

When anomaly detection systems provide classification information about a dataset, the results are either given a score or labels in a binary manner such that the data is either normal or an anomaly (Ahmed & Mahmood, 2014). El-Hajj et al. (2015) concluded that port scanning for open TCP or UDP ports is a critical part of reconnaissance activities performed by malicious actors. An attacker may launch an attack through an open port found through probing. Clients and servers make connections between one another through one of an available 65,536 ports. Well-known or commonly used by applications, TCP and UDP ports include port numbers one through 1,023. Additional registered service ports, which map to applications, include ports

1,024 through 49,151. The remaining ports above port 49,151 are dynamically assigned (El-Hajj et al., 2015).

While an IDS passively monitors network traffic for either a match to a signature or significant deviation from normal behavior, an additional function is provided by intrusion prevention systems (IPS), which take an additional step in blocking detected anomalous behavior (Naik, Diao, & Shen, 2018). Kenkre, Pai, and Colaco (2015) stated that IPS deploy sensors in the path that data must take to arrive at target hosts. If a packet arrives at the sensor on the way to a target host and matches a detection signature, the packet is summarily dropped before reaching its target (Kenkre et al., 2015).

Detection methods. Currently, the self-similar model is the most common model used to detect traffic anomalies and allows for the burstable nature of network traffic (El-Hajj et al., 2015). Kaur, Saxena, and Gupta (2017) defined self-similarity in network traffic as a measure of burstiness at different time scales. As distributed denial of service (DDoS) attacks on a network are occurring, the self-similar nature of the network traffic decreases, which is a reliable indication of an attack (Kaur et al., 2017).

Marchetti, Pierazzi, Colajanni, and Guido (2016) concluded that advanced persistent threat (APT) intrusions to individual hosts are nearly impossible to detect as traffic is disguised by standard encrypted web traffic. Advanced persistent threat actors target known vulnerabilities, which mirror normal web traffic making signature-based IDS ineffective. Detecting hosts that have been infected by APT placed malware and are suspected of exfiltrating typically requires manual protocol analysis and judgement based on total bytes uploaded. Marchetti et al. (2016) devised a suspiciousness score that applied to individual hosts within a monitored network. The suspiciousness score considered the number of megabytes uploaded, number of IP flows, and the number of external IP addresses contacted by the host. Marchetti et al. concluded that standard protocol analysis could not discover a host exfiltrating 500 megabytes per day, although their proposed research and suspiciousness score will detect the host (2016).

Liu, Jin, Min, & Xu (2014) discussed the use of three statistical characteristics when detecting network traffic anomalies caused by DDoS attacks including variance, autocorrelation, and self-similarity. Liu et al. asserted self-similarity as a common characteristic of communication networks. DDoS attacks introduce anomalous traffic, which will reduce self-similarity. Liu et al. (2014) concluded that packet inspection, given the time delay required for analysis, is not a viable solution for DDoS detection.

Network flows. Jirsik, Cermak, Tovarnak, & Celeda (2017) stated that IP flow-based analysis is widely used to measure traffic in large networks and to make possible the discovery of cyber threats. An IP flow is characterized by the IP traffic that passes by a specific network point during a given time. Jirsik et al. (2017) concluded that stream-based analysis is the best solution for real-time detection in highly dense IP networks. Umer, Sher, and Bi (2017) identified that flow-based intrusion detection systems do not consider the payload of a captured packet, but instead analyze flow records to determine if traffic is unusual. Because network flows do not include the data payload, flow-based intrusion detection is faster than packet inspection techniques and do not have the privacy concerns inherent in packet inspection systems (Umer et al., 2017). The architecture of a flow-based intrusion system as described by Umer et al. (2017) includes a metering process where packets are captured, time-stamped, sampled, and filtered. Flow information is exported and stored in the flow database, which acts as input for rule processing and the detection engine. Umer et al. (2017) further found that flow-based intrusion detection systems are not as accurate as packet inspection due to the potential for hidden attacks, which are not included in the flow, as the data payload is not included.

The Internet Engineering Task Force (IETF) defined the IP Flow Information Exchange (IPFIX) protocol to move IP flow data from an exporter process to one or more collector processes over a transport protocol (Claise, Trammell, & Aitken, 2013). The IPFIX protocol supports multiple transport protocols. The preferred transport protocol is the Stream Control Transfer Protocol (SCTP) because it manages network congestion and supports a high volume of exporter traffic. Transmit control protocol also facilitates communication in congested networks. While IPFIX supports UDP as a transport protocol, it is not an ideal choice because UDP does not provide reliability and congestion management. The IETF explained that an exporter process attempts a connection using supported transport protocols on port 4739, and secure connections over the same transport protocols on port 4740 (Claise et al., 2013).

Data sets for IDS testing. El-Hajj et al. (2015) discussed the importance of valid datasets for testing the efficacy of an IDS. Testing IDS requires network traffic to examine. Data can be generated with available tools, or entire datasets in the form of network traffic capture files may be downloaded from the Internet.

Generated traffic. Simulated data sets as suggested by El-Hajj et al. (2015) do not accurately represent normal network traffic. Generated traffic allows testing of an IDS to determine the hit rate and false positive detection frequency of the system. Four categories of background traffic are possible for testing and include no background traffic, real background traffic, sanitized background traffic, and generated background traffic. El-Hajj et al. (2015) focused their attention on generated background traffic in conjunction with attack scripts to create repeatable tests with the same traffic for evaluation.

Downloadable datasets. Databases suitable for testing IDS may be downloaded from multiple sources. According to El-Hajj et al. (2015) datasets available online are mostly outdated and are missing some information suggesting that IDS evaluation accuracy may be negatively affected. Also important to a valid test is the inclusion of non-malicious background traffic. A sampling of datasets available for research includes the Knowledge Discovery and Data Mining dataset which consists of 22 different attack types and contains 743 megabytes of data (University of California Irvine, 1999). The MIT Lincoln Laboratory (1998; 1999; 2000) made available three data sets specifically for IDS evaluation each containing 27 attack types in the 1998 data set and 56 attack types in the 1999 data set. The 2000 dataset contained a DDoS attack perpetrated by a novice attacker where DDoS malware was installed on a compromised host from which a DDoS attack was perpetrated against an offsite host (MIT Lincoln Laboratory, 2000).

Snort. Snort is an open source IDS based on the *libpcap* capture library (Naik et al., 2018). Snort is a signature-based IDS with additional capability for anomaly detection and packet analysis. Leveraging a rule-based system, Snort detects DoS attacks, worm activity, and port scanning in real-time (Naik et al., 2018). The Snort Project (2018) expanded the functionality of Snort v3.0 when compared to its v2.0 predecessor. Snort v3.0 leverages a stream processor and new IPS actions to respond to events that match a signature or are significantly different from baseline traffic (The Snort Project, 2018).

Khamphakdee et al. (2014) proposed an improved collection of Snort-IDS rules, which improved detection rates with fewer false alarms when compared to the built-in Snort rules. The MIT-DARPA 1999 dataset provided by MIT Lincoln Laboratory (1999) provided an evaluation dataset containing various malicious attacks. Khamphakdee et al. (2014) evaluated new Snort-

IDS rules specific to only network probe attacks. The study concluded that 100 percent of probe attacks included in the MIT-DARPA 1999 dataset were detectable with the proposed Snort-IDS rules. There were, however, more positive detections for probe attacks than in the list of confirmed attacks due to multiple entries for attacks at the same time. Finally, Khamphakdee et al. (2014) recommended that Snort rules require frequent updates.

Naik et al. (2018) proposed an IDS integrating Snort with their solution for a dynamic fuzzy rule interpolation scheme, which leverages a smaller rule base. A baseline of network behavior was determined by calculating three characteristics. First, Snort captured packets to determine the average time between packets (ATP) and the destination host. Next, the number of packets sent (NPS) by the source host per second was logged. Finally, the number of packets received (NPR) per second by the destination host was recorded. Combined, this information formed a base for comparison against a series of port scan attacks. The baseline ATP value was determined to be 18 milliseconds with an NPR less than 1,000 packets per second, and an NPS value less than 270 packets per second. Naik et al. (2018) determined that the ATP value was the most critical value in determining the presence of a port scan attack. There was an inverse relationship between the ATP value and the amount of port scan traffic, as port scan traffic increased the ATP value decreased. Experiments conducted by Naik et al. (2018) concluded that a sparse Snort rule base combined with an inferential intelligence based on baseline values for ATP, NPR, and NPS allowed for decreased false positive and negative detection of port scan attacks.

Summary

A review of this research has explored the effects of undetected cyberattacks on organizations, the significance, and cost of data breaches, manual detection of cyberattacks and network anomalies, and the use of intrusion detection systems in improving incident response. Exposed PII because of data breaches has reached staggering numbers. Data breaches in the past five years have exceeded 5.58 billion records (Privacy Rights Clearinghouse, 2018).

The top 10 data breaches due to hacking over the past five years account for 89 percent of all exposed PII records in the US. Additionally, the top 100 data breaches account for 5.57 billion lost records (Privacy Rights Clearinghouse, 2018). Overall, data breaches remain flat year over year for the period 2005 through 2014 (Edwards et al., 2016). Simultaneous improvement in security practice and attacker technique may explain the relative consistency in both data breach size and frequency (Edwards et al., 2016).

The Ponemon Institute (2014a; 2015; 2016; 2017; 2018) reported the average cost of data breaches for the five-year period from 2014 to 2018 was $3,758,000. The Ponemon Institute (2018) suggested the cost of a mega-breach to be $40 million for one million compromised records, and $350 million for fifty million records. Seventy-seven percent of respondents were happy with the way the reporting company handled the breach, and 89 percent of respondents continued to do business with the company (Ablon et al., 2016).

All 50 states in the US have breach notification laws requiring individual notification when a breach has occurred (National Conference of State Legislatures, 2018). The Data Security and Breach Notification Act will require notification of a data breach affecting PII within 30 days if the bill becomes law (115th Congress, 2017). Organizations must report data breaches to the Attorney General’s office or the authority over any regulated industry such as the financial sector (Corey & Wilsker, 2015).

Ahmed and Mahmood (2014) defined network traffic analysis to infer patterns from network protocol communication. Preventing disruption to an organization’s network can be facilitated through the proactive use of protocol analysis and network traffic anomaly detection. According to Sestito et al. (2018), an understanding of the normal traffic patterns on the network must precede anomaly detection.

Singh and Baliya (2015) reviewed Wireshark as a tool for capturing and analyzing network traffic for malicious behavior. Display and capture filters limit the traffic shown in Wireshark. The tcpdump command line protocol analyzer dropped up to three percent more packets than Wireshark while capturing in monitor mode, and while capturing Ethernet packets tcpdump dropped up to one percent more packets than Wireshark (Goyal & Goyal, 2017).

Yang et al. (2016) stated that host-based detection schemes alone are not enough to identify a cyberattack. Network-based detection is a technique to allow discovery of cyberattacks based on network traffic analysis or honeypot activity (Yang et al., 2016). Challenges to network-based detection include the unlikelihood that honeypots can detect all threats and the prevalence of hidden malicious traffic amongst normal traffic flows (Yang et al., 2016).

Ahmed and Mahmood (2014) referred to network probes as a reconnaissance technique used to determine the functionality of a remote host. Network probes may reveal valuable information about the network to an attacker. Singh and Baliya (2015) stated that the goal of DoS attacks is to disrupt the legitimate use of the network or computer resource. Tripathi and Hubballi (2018) concluded that the HTTP/2 protocol is vulnerable to several DoS attacks.

Botnet operators leverage fast-flux DNS services to provide a complex, load balanced, and highly available network (Yang et al., 2016). Katz et al. (2017) found that botnets use fast-flux networks to avoid the discovery of malware and its command and control (C&C) infrastructure. Fast-flux networks are characterized by frequently changing DNS nameservers, domain names, and host IP addresses. Malware hosting and delivery, communication, phishing, and web proxying are malicious activities that are protected from discovery by fast-flux networks (Katz et al., 2017).

Wendzel and Keller (2014) defined a micro-protocol as having a protocol header stored inside the hidden data payload of a clandestine channel communication. Micro protocols add proxy, dynamic routing, and connection management characteristics to covert channels. Use cases for micro protocols include botnet C&C, covert journalist communication in Internet censored states, and military and secret agency covert communication Wendzel and Keller (2014).

An IDS passively monitors network traffic for a match to a signature or evidence of a significant deviation from normal behavior. Blocking detected anomalous traffic is an additional function provided by an IPS (Naik et al., 2018). Kenkre et al. (2015) stated that IPS deploy sensors in the path that data must take to arrive at target hosts.

Currently, the self-similar model is the most common model used to detect traffic anomalies and allows for the burstable nature of network traffic (El-Hajj et al., 2015). As DDoS attacks on a network are occurring, the self-similar nature of the network traffic decreases, which is a reliable indication of an attack (Kaur et al., 2017).

Marchetti et al. (2016) concluded that APT intrusions to individual hosts are nearly impossible to detect as traffic is encrypted. Manual protocol analysis and judgment based on total bytes uploaded are required to detect hosts that have been infected by APT placed malware. Marchetti et al. (2016) devised a suspiciousness score to detect the presence of data exfiltration traffic, which can detect a host exfiltrating 500 megabytes per day. Liu et al. (2014) asserted self-similarity as a common characteristic of communication networks. DDoS attacks introduce anomalous traffic, which will reduce self-similarity. Liu et al. (2014) concluded that packet inspection, given the time delay required for analysis, is not a viable solution for DDoS detection.

Jirsik et al. (2017) concluded that stream-based analysis is the best solution for real-time detection in highly dense IP networks. Umer et al. (2017) identified that flow-based intrusion detection systems do not consider the payload of a captured packet, but instead analyze flow records to determine if traffic is unusual. Because network flows do not include the data payload, flow-based intrusion detection is faster than packet inspection techniques and do not have the privacy concerns inherent in packet inspection systems (Umer et al., 2017).

El-Hajj et al. (2015) discussed the importance of valid datasets for testing the efficacy of an IDS. Testing IDS requires network traffic to examine. While tools are available to generate data, entire datasets in the form of network traffic capture files are available on the Internet.

Snort is a signature-based IDS with additional capability for anomaly detection and packet analysis. Leveraging a rule-based system, Snort detects DoS attacks, worm activity, and port scanning in real-time (Naik et al., 2018). Experiments conducted by Naik et al. (2018) concluded that a sparse Snort rule base combined with an inferential intelligence based on baseline values for ATP, NPR, and NPS allowed for decreased false positive and negative detection of port scan attacks.

Discussion of the Findings

This study examined the role of protocol analysis in cybersecurity. The research included a synthesis of scholarly journals, reports, and correlated database search results to address the research questions. Studies reflecting on the state of data breaches in the US form the basis for discussion around the effect of undetected cyberattacks on organizations. This research examined manual protocol analysis and intrusion detection systems as potential means to reduce the time of cyberattack to the time of detection.

Effect of Undetected Cyberattacks on Organizations

The first research question considered in this research was the effect that undetected cyberattacks have on organizations. A comparative study of available literature concluded that a significant cost to organizations exists along with considerable risk to consumers in the form of identity theft. Multiple reporting organizations contribute to the understanding of the cost and significance of data breaches each year. The Ponemon Institute *Cost of a Data Breach Study* is released annually to report the global cost of a data breach both collectively and individually. Verizon collates and analyzes real-world data breach reports into its annual *Data Breach Investigations Report*. The Privacy Rights Clearinghouse curates data breach information in a searchable database, which allows export of data suitable for research. Ongoing and focused research and reporting by these institutions indicated a clear need for improvement in breach detection, incident response, and cost mitigation.

While Densham (2015) asserted that organizations must assume that a data breach has already occurred, Edwards et al. (2016) noted that the Privacy Rights Clearinghouse data indicated that data breaches were flat year over year for the period 2005 through 2014. In addition to the consistent trend in the number of data breaches, the frequency of data breaches and the number of records lost continue to remain the same as of September 2015. The sophisticated nature of cyberattacks paired with a simultaneous improvement in security practice is offered by Edwards et al. (2016) as a possible explanation for the relative year over year consistency in both data breach size and frequency. Lacking a downward trend in the number of data breaches, their frequency, and the number of records lost combined with a cybersecurity industry predilection toward assuming a breach has already occurred, points to weakness in incident monitoring and response. Cichonski et al., (2012) suggested that deficiencies in training, in-depth technical knowledge, and experience may hamper incident detection and response (Cichonski et al., 2012). A culture of security and privacy awareness through employee training will improve internal security practices (Ehrlich, 2017). Training, combined with continued improvement in security practice, may contribute to improved incident detection.

The effect of data breaches significantly burdens organizations of all sizes. Small businesses were data breach victims 1,285 times in 2017 according to Verizon (2018). Wheatley, Maillart, and Sornette (2016) suggested that personal data stored by an organization increases in volume and value as the size of the organization, by market capitalization, increases making larger organizations frequent targets of cyberattack. Large corporations contributed to the overall 5.58 billion compromised records with 52 of the top 100 data breaches during the period 2014 through 2018 (Privacy Rights Clearinghouse, 2018). The top 52 data breaches meet the threshold of a mega-breach as defined by the Ponemon Institute (2018). A mega-breach is estimated to cost $40 million for one million records lost and $350 million for 50 million records lost. Jay Jacobs' log-log costing model challenges the accuracy of the Ponemon per capita cost-based model. The Jacob's model cites fluctuations in the natural variance of compromised records and financial loss related to data breaches as factors precluding the use of a linear or linear regression model to represent the year over year change in records lost.

While the top 100 data breaches account for 99.8 percent of all PII records lost, the top 10 data breaches contribute 89 percent to the overall lost records for the period 2014 to 2018 (Privacy Rights Clearinghouse, 2018). The top 10 large-scale data breaches, referred to as mega-breaches by the Ponemon Institute (2018) account for $4.98 billion for the five-year period 2014 to 2018. A significant portion of data breaches affected organizations in California. Three of the top five data breaches were reported by California organizations along with seven of the top 10 organizations. A review of the data breaches listed in Appendix A revealed 35 of the top 100 data breaches targeted California organizations. Second and third place on the list of top 100 data breaches targeted Georgia and New York with nine and seven data breaches respectively. Combined, more than half of data breaches for the period beginning 2014 through 2018 occurred in the three states.

During 2017, McConnell stated that the average cost of a data breach in the US was $6 million to $10 million, while the Ponemon Institute (2018) identified $3.86 million worldwide and $7.91 million as the average in the US for the same year. The worldwide cost of a data breach averaged $3.75 million over the period 2014 through 2018. While there is some disagreement around cost prediction models, the overall cost of data breaches has maintained a largely flat trajectory over the last five years. The average cost of a data breach in the US is double the average cost when compared to worldwide Ponemon Institute (2018) participants.

McConnell (2017) stated that cyberattacks and related data breaches continue to take an average of 200 days to detect. The Ponemon Institute (2018) defined a data breach as an event that potentially compromises a record of information, which may identify a person and their financial or medical information. Additionally, attackers exfiltrate compromised records during a cyberattack. Of the confirmed 2,216 data breaches reported by Verizon (2018), the majority took months or longer to detect. The time that a data breach goes undetected creates an identity theft risk for consumers.

Leaked PII including name, address, social security number, driver's license, and financial and medical records are frequently sold to criminals on the dark web (Gupta, 2018). Although consumers mostly continue to do business with organizations who lose their PII, there is still significant cost and risk to organizations in the form of remediation efforts, cost of identity theft monitoring services, and damage to reputation. When financial PII is leaked, organizations face a six-fold likelihood of legal action by affected consumers (Romanosky et al., 2014). About a third of surveyed consumers ignored data breach notifications, and only 29 percent accepted identity monitoring services. Ehrlich (2017) recommended organizations offer fraud protection services as part of a three-pronged approach to prevent reputation and financial harm to organizations. These statistics contradict survey respondent sentiment that organizations who have lost PII should provide identity theft monitoring services to consumers (Ponemon Institute, 2014b).

Recovering from a data breach related incident poses an average remediation cost of $500 to consumers. The mosaic effect explained by Felten (Medine et al., 2014) identified a trend toward inferring consumer behavior based on linking large datasets and using leaked PII to fill in knowledge gaps. Gupta (2018) stated that artificial intelligence contributes to the efficiency of the mosaic effect ultimately creating a risk of unwanted release of information, fraudulent credit activity, and targeting by hate groups based on discovered consumer behavior. As an extension of organizational cost and risk, the cost to consumers includes direct monetary cost as well as significant risk to identity theft and profiling which could be considered a permanent and persistent threat. With 200 days as the average time to detect a data breach, remediation efforts do not begin until PII is exposed actualizing the threat to consumers. Consumer awareness and education around the potential for identity theft and the creation of behavior-related databases should increase the likelihood that consumers will accept credit-monitoring services when offered.

Manual Detection of Network Anomalies

The second research question considered in this research was how cyberattacks and network anomalies are manually detected. Defining categories of anomalies as point, contextual, and collective, Ahmed and Mahmood (2014) noted that network traffic analysis is a means to infer patterns from network protocol communication. Ahmed and Mahmood (2014) asserted that preventing disruption to an organization's network can be facilitated through the proactive use of protocol analysis and network traffic anomaly detection. Normal traffic is defined as the traffic that is representative of the communication patterns of all processes operating on the network (Sestito et al. 2018). An event that deviates to a high degree from the normal traffic model is considered anomalous. Categorizing anomalies is a common thread among researchers. Barford and Plonka (2001) categorized network traffic into four distinct behavior classes including network traffic anomalies, irregularities related to network operations, flash-crowd anomalies, measurement failures, and attacks. Despite differences in category definition, researchers Sestito et al. (2018), Barford and Plonka (2001), and Ahmed and Mahmood (2014) agreed that anomaly detection requires an understanding of normal network traffic behavior. Anomalies in traffic which affect the confidentiality and integrity of information in the system are attacks (Sestito et al., 2018).

Data gathering is imperative to measure network behavior accurately (Sestito et al., 2018). This research identified a limited selection of available tools suitable for data gathering. Both Wireshark and tcpdump are freely available protocol analyzers, which are available to researchers and network analysts. Wireshark allows for limiting either captured or displayed traffic with capture and display filters (Singh & Baliya, 2015). While tcpdump was found to consume less power, memory, and processor resources, Wireshark captured more packets while capturing traffic on Ethernet networks and while in monitor mode (Goyal & Goyal 2017). The statistics feature of Wireshark, as demonstrated by Singh and Baliya (2015), displays the flow of network traffic graphically. When under cyberattack, security teams may block IP addresses, transport protocols and URLs discovered through traffic analysis (Singh & Baliya, 2015). Both tools are capable of successfully capturing traffic, but Wireshark is preferred, as research has shown it to capture more packets at the cost of power, memory, and processor when compared to tcpdump.

Host-based or network-based detection techniques detect network attacks including DoS, probes, user-to-root escalation, and remote-to-user attacks. Host-based tools can detect botnets, worms, and other threats, but due to factors such as lack of scale, signature compatibility, and performance effects, host-based detection software alone is not enough to protect an organization from cyberattack (Yang et al., 2016). It is similarly unlikely that network-based traffic analysis can detect all threats. In addition, hidden malicious traffic amongst normal traffic flows makes manual detection more difficult. Network attacks are increasingly sophisticated. While manual threat detection is possible, it is unlikely that all threats can be detected and remediated quickly (Yang et al., 2016).

Khamphakdee et al. (2014) identified nmap, satan, and mscan as freely available probing applications. Reconnaissance against target networks is the purpose of probe attacks. The information gathered from a probe attack is the basis of more sophisticated cyberattacks such as DoS, privilege escalation, and unauthorized local user access.

Denial of service attacks disrupt the legitimate use of network or computer resources. Unlike high traffic volume attacks, slow-rate DoS attack methods proposed by Tripathi and Hubballi (2018) exploit the HTTP/2 protocol. Specially crafted HTTP/2 packets consume available web server connections in five scenarios designed by Tripathi and Hubballi (2018). As few as 150 packets can deplete available connections against a target HTTP/2 server. Both plaintext and encrypted HTTP requests achieve successful attacks (Tripathi & Hubballi, 2018). Attacks with this level of precision suggest a required review of the protocol or its implementation in server software.

Weaknesses in the TCP protocol allow attackers to probe networks without being detected. Transmission control protocol sessions start with a predictable three-way handshake, which is exploited by attackers. Open TCP network ports are identified by responding hosts with the SYN and ACK bits set in the TCP header; likewise, a predictable host response with the RST bit set in the TCP header signifies a closed TCP port. Attackers leverage this basic protocol knowledge to conduct reconnaissance against a target network. El-Hajj et al. (2015) described stealthy port scanning techniques, which leverage the use of the SYN, FIN, and ACK bits to perform probe scans without detection by the remote firewall.

As shown in the Literature Review, attackers have significant protocol weaknesses and known behaviors to leverage when performing reconnaissance and attacks. Malware and covert channels take advantage of header fields, which allow arbitrary values. Botnets demonstrate the use of stealthy C&C networks with fast-flux DNS systems, which provide frequently changing DNS nameservers, domain names, and host IP addresses (Katz et al., 2017). Common TCP ports combined with fast-flux DNS are used to disguise malware hosting and delivery, malware control, phishing, and web proxy traffic among normal network traffic making manual detection ineffective. Katz et al. (2017) instead recommended intrusion detection systems, which focus on algorithms which can detect the fast-changing nature of fast-flux networks.

Attackers may exfiltrate PII covertly by leveraging TCP weaknesses reported by Rowland (1997). Both Rowland (1997) and Mehic et al. (2016) discussed covert channels which exist due to the use of unused fields of protocol headers where arbitrary values are allowed or by encoding data in the behavior characteristics of the carrying protocol. Wendzel and Keller (2014) reviewed the use of micro-protocols, which are carried by ICMP, IP, UDP, or RTP protocols and vary in range from eight to 192 bits in hidden communication. Rowland (1997) demonstrated covert channel data flow by exploiting TCP and IP headers and protocol behaviors. When included in a busy network, covert channel communication may exist along with normal network traffic making detection through manual protocol analysis or intrusion detection systems unlikely due to the lack of real-time processing over large datasets (Mehic et al., 2016).

Contribution of IDS and IPS to Detection of Cyberattacks

The third research question considered in this research asked how IDS and IPS contribute to the detection of cyberattacks. Naik et al. (2018) defined IDS as a passive monitoring system which attempts to match network traffic to a known signature or significant deviation from a normal traffic baseline. An IPS sensor additionally takes an active role in defense by dropping packets which match a signature along the path to a target host (Kenkre et al., 2015). Intrusion detection literature reflects heavily on statistical analysis of network traffic seeking to determine anomalies, and network attacks, by finding patterns in traffic. The most prevalent model found in this research was the self-similar model. El-Hajj et al. (2015) suggested that the self-similar model allows for the burstable nature of network traffic. Measured over time and in different time scales, the burstiness of network traffic is measured. In the presence of a network attack, such as a DDoS, the self-similar traffic decreases identifying the traffic as anomalous and likely a cyberattack (Kaur et al., 2017).

Marchetti et al. (2016) concluded that signature-based intrusion detection is ineffective against APT network intrusions as standard encrypted web traffic hides the presence of APT activity. To determine the likelihood of compromise, host generated traffic receives a suspiciousness score based on total bytes transmitted, number of IP flows, and the number of external IP addresses contacted. Mirroring El-Hajj et al. (2015) and Kaur et al. (2017), Liu et al. (2014) asserted traffic self-similarity as a cornerstone of intrusion detection technique. Additionally, manual packet inspection was considered not viable as a DDoS detection technique due to the time required to perform analysis in the face of an ongoing cyberattack. Jirsik et al. (2017) concluded that stream-based analysis is the best solution for real-time detection in highly dense IP networks because flows do not include the data payload. In agreement, Umer et al. (2017) discussed flow-based IDS as a successful technique for detecting unusual network traffic.

Testing IDS requires an analysis of the hit rate and false positive detection frequency. Generated traffic to mirror normal flows in combination with attack scripts creates predictable datasets for hit rate and false positive analysis (El-Hajj et al., 2015). Downloadable datasets including the Knowledge Discovery and Data Mining from the University of California, Irvine (1999) and the MIT Lincoln Laboratory (1998; 1999; 2000) are well known for research efficacy. Khamphakdee et al. (2014) concluded that their Snort-IDS network signatures were

100 percent effective in detecting probe attacks included in the MIT-DARPA 1999 dataset, but returned a higher number of probe attacks than those contained in the dataset. Research conducted by Naik et al. (2018) determined that a sparse Snort rule base combined with inferential intelligence based on network baseline values for ATP, NPR, and NPS decrease the prevalence of false positive and negative detection for port scan attacks. Manual protocol analysis tools can arguably calculate the inferential intelligence baseline values for the average time between packets, the number of packets sent, and the number of packets received, but not in real time.

Limitations of the Study

While conducting this research, some limitations exist in the chosen literature for review. While Wireshark and tcpdump are popular and freely available network capture tools, other network capture software packages exist and which also work with the libpcap library. Additional tools, such as NetworkMiner, Kismet, Ngrep, and ColaSoft Packet Builder are capable tools for network capture and analysis. These other capture tools could be equally useful tools for manual protocol analysis and anomaly detection.

Cyberattacks identified in the Literature Review included probes, denial of service, malware command and control activity, and several covert channel techniques. While these cyberattacks represent the considerable threat to organizations, they are not complete as DDoS attacks, which attempt to consume network bandwidth, are characteristically different from the HTTP/2 DoS attack described in this research. Covert channel techniques, while highlighting the seminal work of Rowland (1997) as well as a discussion of several micro-protocols (Mehic et al., 2016), did not include an in-depth detailing of multiple protocol headers and fields which allow arbitrary values. Likewise, probe attacks common in network reconnaissance were discussed, especially, with the nmap utility. Each possible nmap attack was not reviewed in detail and left to the reader for future research.

The Snort IDS is a viable tool for detecting cyberattacks. As there are myriad other tools available for IDS and IPS, Snort is one tool which acts as a flow-based IPS. Honeypot and decoy techniques were not reviewed as part of this research but may add to the efficacy of a defense-in-depth strategy.

Summary

In conclusion, the significant findings of this research include answers to the research questions posed in the Statement of the Problem. Organizations of all sizes are affected by undetected cyberattacks with 5.58 billion compromised records at an average cost of more than $7.91 million in the US. More than 200 days pass, on average, between the time of breach and the time of detection causing organizations to lose valuable PII at the hands of criminals. As an extension of adverse organizational effects brought about by undetected cyberattacks, consumers are at risk of identity theft and the mosaic effect inherent in multiple available consumer profile datasets. Consumers are further vulnerable to identity theft due to a propensity for turning down identity theft monitoring services or ignoring breach notifications completely.

Manual detection of network anomalies and data breaches are possible when comparing anomalous network traffic to a known baseline of traffic, which reflects all expected network processes. While some attacks are detectable through protocol analysis, the time to detect anomalies manually is not practical. Likewise, encrypted network flows make manual analysis for fast remediation of network data breaches impractical. The self-similarity nature of network traffic, the average time between packets, and the number of packets sent and received are statistical indicators of baseline and anomalous network traffic. Manual analysis of fast-flux

C&C techniques, covert channel and micro-protocols are ineffective due to their complexity. While manual protocol analysis is not a single solution providing for fast detection of cyberattacks and resulting data breaches, manual protocol analysis does play a role in the situational evaluation of network traffic where the total number of bytes sent provides a base for inferential intelligence.

Intrusion detection and prevention systems contribute to the detection of cyberattacks in their ability to analyze network traffic more quickly. Host-based and network-based intrusion detection systems employ the principles of network traffic self-similarity to identify the presence of a cyberattack. The self-similarity of network traffic decreases in the face of a cyberattack. Flow-based IDS were found to be most adept at detecting cyberattack due to the exclusion of the data payload and the resultant real-time analysis enjoyed by stream-based systems. Flow-based IDS likewise do not cause privacy concerns inherent with packet inspection techniques. Flow-based IDS do not include data payloads making these systems less accurate than packet inspection.

Recommendations

This research examined the role of protocol analysis in cybersecurity. The impact of undetected cyberattacks and their resulting data breaches carry a significant financial burden for organizations even though consumer sentiment favors continued business with the organization. Organizations suffer a loss of customers and a higher than average consumer churn as a result but ultimately survive after a reported data breach. After examining available literature focusing on protocol analysis and automated intrusion detection techniques, this research concluded that neither protocol analysis or intrusion detection systems alone are completely effective against current and emerging attack techniques.

Security Operation Centers and Information Technology Departments

Security operation centers and information technology departments should consider approaching network anomaly detection through a multi-layered defense-in-depth strategy, which includes both manual and automated intrusion detection and prevention systems. Comparisons between a network traffic baseline, including all services in use on each network segment, and network traffic, which includes a suspected anomaly or ongoing cyberattack, form the basis for manual anomaly detection through protocol analysis.

Network anomaly detection through manual protocol analysis is impractical when traffic is encrypted, or contains complex flows including C&C communication channels with fast-flux DNS, covert channel, or micro-protocols. While manual detection through protocol analysis is time-consuming and impractical for most attacks, the phenomenon of network traffic self-similarity, the average time between packets, and the number of packets sent and received are statistically sound data points that may help to discover network anomalies when compared to a known baseline. Training programs, which focus on individual protocol knowledge and protocol analysis tools, such as Wireshark and tcpdump, will prepare security analysts for the situational need for manual protocol analysis.

Host-based and network-based IDS, which leverage self-similarity and known signature algorithms, are the second component to anomaly detection complementing manual detection through protocol analysis. Flow-based network IDS systems are known to be more efficient than full-payload protocol analysis techniques and do not impose the same privacy concerns as manual protocol analysis techniques. While it is not possible to detect attacks hidden in the data payload, stream-based solutions are required to decrease the 200-day time-to-detection rate to a rate that precludes financial damage to organizations and identity-theft risk to consumers.

Organizational Limits to Stored PII

Organizations should not keep more consumer PII than is necessary to execute their business serving the consumer. Reducing data kept will reduce the overall effect of data breaches on consumers and organizations. When a data breach has occurred, an audit should be conducted by an independent third party to evaluate the appropriateness of the data stored by the organization. Personally identifiable information should be considered the property of the individual to whom the information applies and organizations should be deemed to be acting as an opt-in curator. A convenient mechanism to delete all PII from an organization should be available to consumers.

Consumer Behavior and Advocacy

Consumers are not adequately prepared for the risk that they face because of a data breach. Consumers should proactively subscribe to identity theft monitoring services, especially after having received a breach notification. When offered identity theft monitoring services by organizations who have experienced a data breach, the offered subscription should be carefully considered for acceptance.

To better prepare consumers, a consumer advocacy organization should be created and funded, in part, by fines levied against ill-prepared organizations who lose PII of consumers. The program should be partially government funded. The goal of the advocacy organization is to provide aggressive advertisements to consumers through all forms of current consumer channels such as social media, streaming video services, Internet web banners, television, and radio.

Recommendations for Future Research

Outlined in this section, and submitted for consideration, are future research recommendations. Exigency for improvement in how organizations handle PII is an impetus for continued research. In addition, research to enumerate the available detection systems fully as well as improvements in communication protocols may be undertaken.

Organization size and location study. There was considerable evidence exposed in this research, which suggested a corollary between organization size and location to data breach frequency and size. Specifically, California was the dominant state within the US where data breaches occurred the most. A study should be conducted to identify the degree to which this phenomenon exists and if there are any identifiable organizational traits, which may be addressed to reduce the frequency and size of data breaches in California and other high-frequency states in the US. Additionally, mega-breaches appear to trend upward in size as the size of the organization increases. An investigation into the size of a mega-breach as compared to organization size should be conducted.

Monitoring PII on the dark web. The dark web has been identified as a primary channel for the sale of exfiltrated consumer PII. Monitoring the dark web for the presence of PII datasets, which match the PII curated by an organization, was suggested as an actionable incident response tactic. A study may reveal a direct correlation to remediation time and cost savings because of dark web monitoring. Additionally, a review of dark web marketplaces may reveal insight into the value of various types of PII including the worth of individual records and any emerging trends in sought after records. Armed with marketplace value, organizations may be able to partition stored data into less attractive datasets. Research should be conducted into methods of introducing noise into stored datasets with algorithms, which make directly exfiltrated datasets worthless.

Cost of a data breach in the US. This research revealed the cost of a data breach in the US is more than double the worldwide average due to notification costs. The Ponemon Institute has conducted significant research surrounding the cost of a data breach, including identifying notification costs as the significant causation to US data breach remediation costs. While notification laws exist for all states in the US, it is unclear if notification requirements alone are the only contributors to US costs. A study should be conducted to identify significant causation into US costs and opportunities to reduce those costs.

A comparative study of IDS, IPS, and decoy systems. There are many solutions available for IDS, IPS, and decoy systems. Decoy systems aim to provide a safe environment in which attackers may be monitored to learn about their intent. Decoy systems were not covered as part of this research and should be considered in future research. Intrusion detection systems, such as Snort, play a significant role in detecting a cyberattack. An expanded and comparative study of available IDS and IPS solutions may uncover trends in stream-based analysis and novel statistical approaches to near real-time detection of reconnaissance and remediation of unauthorized accesses to organization networks. As uncovered in this research, IDS and IPS are critical components to a defense-in-depth strategy. Future research, which evaluates all available solutions, including decoy systems, will help security operation centers and information technology departments to identify the best combination of products.

Securing the TCP/IP suite. Researchers have exposed significant weaknesses in the TCP/IP suite. The presence of TCP and IP headers, which accept arbitrary values, are exploitable for use in covert channel and clandestine micro-protocol communication. Some values, such as the payload of ICMP packets, may carry arbitrary values leading to use as a covert channel, which is used to exfiltrate data. A study into protocol weaknesses and potential improvements to protocols, such as ICMP, should be conducted to identify ways to lessen the number of covert channels available for data exfiltration.

Beyond State of the Art. The current state of the art includes a defense-in-depth strategy, as discussed in this research, consisting of situational manual protocol analysis, IDS and IPS. Researchers must consider the relationship between data volume and processing speed. Specifically, researchers should conduct a study into computing requirements as data generation increases. As researchers seek to close the gap on undetected data breaches, the role of manual protocol analysis, as well as IDS and IPS, must evolve to address increases in the volume of data to be processed.

This study considered several techniques for anomaly detection. The self-similar nature of network traffic, for example, provides a characteristic that may allow deep learning neural network techniques applicability to the prediction of anomalous network behavior. A machine learning based solution might route detected anomalous network traffic to a designated network for further study or inclusion in suitable training datasets or drop the traffic. Sophisticated network attacks are often not detected by current flow-based IDS systems requiring research into more sophisticated detection methods. As cyberwar is already a domain of war, cyberattack should immediately become a domain for machine learning and deep learning research and application.

Inward-focused defense-in-depth strategies are required today because malicious traffic reaches the perimeter of an organization's network. To overcome shortcomings of an inward facing defense-in-depth strategy where organizations deploy perimeter security, host and network-based IDS, and situational manual protocol analysis, solutions based on machine learning algorithms should function at all public layers of the Internet. With an intelligent machine learning system in place at all Internet service providers, content delivery networks, and cloud services the defense-in-depth strategy becomes complete and more likely to prevent cyberattack and undetected data breaches. This strategy prevents malicious traffic from reaching organization networks. Investment and placement of these systems should be compulsory critical infrastructure and considered a significant means of protecting organizations of all kinds.

Future research into anomaly detection through quantum computing characteristics of entanglement and superposition may allow for processing enormous network traffic datasets at unprecedented speed. Current public key cryptography technology, however, leverages the multiplication of huge prime numbers, which will prove easily reversible by quantum computers requiring additional research into new per-message encryption techniques. The result of research in quantum computing for traffic inspection requires a concurrent policy debate around monitoring and privacy.

Conclusion

This research was conducted to address the need for improvement in the time to detect data breaches. Currently, 200 days pass, on average, before victimized organizations recognize the presence of malicious software installed as part of a cyberattack (McConnell, 2017). Lengthy delays between initial breach and mitigation provide attackers ample time to move laterally within target networks, exfiltrate and possibly destroy data, or disrupt normal network operation. In extreme cases, maintained access to victim networks had spanned one to five years in duration while exfiltration of hundreds of terabytes of data occurred (Mandiant, 2013).

Organizations of all sizes are affected by undetected cyberattacks resulting in a loss of 5.58 billion PII records in the previous five years (Ponemon Institute, 2018). Costs have soared to an average of $7.91 million per incident in the US. In addition, a mega-breach is a new category of a data breach, which describes losses of more than one million records. Organizations have a responsibility to notify consumers when a data breach occurs and to keep consumer PII safe through well-defined incident response planning and protection of networks. Consumers are vulnerable to identity theft because of data breaches and are encouraged to use identity theft monitoring services to protect themselves.

Research has shown that neither manual protocol analysis nor IDS, can detect 100 percent of cyberattacks resulting in data breaches. Given a known network traffic baseline, which is inclusive of all network services, a comparative analysis may be undertaken to detect anomalies. Encrypted payloads impose challenges on manual protocol analysis due to the presence of hidden attacks. Unencrypted data flows expose privacy concerns when applying manual protocol analysis techniques to anomaly detection. Manual protocol analysis remains an effective method of identifying infected hosts, which are exfiltrating data. Encrypted data payloads, the time required to perform manual analysis, and privacy concerns warrant a shift to the situational application of protocol analysis in combination with an IDS or IPS solution. Training and expertise in manual protocol analysis techniques is a critical component to incident response.

The self-similarity phenomenon, which exists for network traffic datasets, is key to the detection of intrusions and anomalies in IDS and IPS solutions. A layered implementation of host-based and well-positioned network-based IDS solutions based on known-signature and self-similarity analysis allow for more quickly detected anomalies. Automated IDS based on network traffic flows removes privacy concerns inherent in manual protocol analysis solutions while identifying attacks closer to real-time. Limitations to host-based IDS, including a lack of scale, signature compatibility, and adverse performance effects on the host preclude host-based IDS alone as a viable solution to network protection. Hidden attacks among regular network traffic flow challenge network-based IDS solutions, which require additional analysis through manual protocol analysis or host inspection.

In closing, a combination of well-trained analysts armed with protocol knowledge and the ability to perform manual protocol analysis, along with complementary host-based and network-based IDS solutions form the basis for protection of modern organization networks. In addition, future research may provide advancements in machine learning algorithms, which may predict the presence of anomalies in large datasets. A well-crafted incident response plan combined with detection methodologies outlined in this research should result in an overall reduction in the time that passes between an intrusion and when that intrusion is detected.

References

115th Congress. Data security and breach notification act (2017). S. Retrieved from https://www.congress.gov/115/bills/s2179/BILLS-115s2179is.pdf

18 U.S.C. 1030. Fraud and related activity in connection with computers (2009). Retrieved from https://www.govinfo.gov/content/pkg/USCODE-2016-title18/pdf/USCODE-2016-title18-partI-chap47-sec1030.pdf

Ablon, L., Paul Heaton, Diana Catherine Lavery, & Romanosky, S. (2016). *Consumer attitudes toward data breach notifications and loss of personal information*. *RAND Corporation*. Retrieved from http://www.jstor.org.proxy-um.researchport.umd.edu/stable/10.7249/j.ctt1bz3vwh.7?Search=yes&resultItemClick=true&searchText=(data&searchText=breach)&searchUri=%2Faction%2FdoBasicSearch%3FQuery%3D%2528data%2Bbreach%2529&refreqid=search%3Ae0c157c005b44b7577

Ahmed, M., & Mahmood, A. N. (2014). Network traffic analysis based on collective anomaly detection. *2014 9th IEEE Conference on Industrial Electronics and Applications*, 1141–1146. https://doi.org/10.1109/ICIEA.2014.6931337

Ahmed, M., Naser Mahmood, A., & Hu, J. (2016). A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, *60*, 19–31. https://doi.org/10.1016/j.jnca.2015.11.016

Baran, P. (1964). On distributed communications: Introduction to distributed communications networks. Retrieved from https://www.rand.org/pubs/research_memoranda/RM3420.html#download

Barford, P., & Plonka, D. (2001). Characteristics of network traffic flow anomalies. *Proceedings of the First ACM SIGCOMM Workshop on Internet Measurement - IMW '01*, 69. https://doi.org/10.1145/505202.505211

California Code. Section 1798.29 Information Practices Act of 1977 (1977). Retrieved from http://leginfo.legislature.ca.gov/faces/codes_displaySection.xhtml?lawCode=CIV&sectionNum=1798.29.

Cerf, V. G., & Kahn, R. E. (1974). A protocol for packet network intercommunication. *IEEE Trans on Comms*, *22*(5). Retrieved from https://www.cs.princeton.edu/courses/archive/fall06/cos561/papers/cerf74.pdf

Cichonski, P., Millar, T., Grance, T., & Scarfone, K. (2012). Computer security incident handling guide: Recommendations of the National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-61r2

Claise, B., Trammell, B., & Aitken, P. (2013). Specification of the IP Flow Information Export (IPFIX) Protocol for the exchange of flow information. Internet Engineering Task Force. Retrieved from https://tools.ietf.org/search/rfc7011

Corey, A. T., & Wilsker, N. R. (2015). Data breach preparedness: A look at the legal responsibilities specific to Granite State firms Northern New England Resource for Legal Business Matters. *New Hampshire Business Review*, *37*(8), 27. Retrieved from http://search.ebscohost.com.ezproxy.utica.edu/login.aspx?direct=true&AuthType=ip,cookie,url,uid&db=bwh&AN=114693093&site=ehost-live

Davies, D. W. (1966). Proposal for a digital communication network. Retrieved from http://www.dcs.gla.ac.uk/~wpc/grcs/Davies05.pdf

Densham, B. (2015). *Three cyber-security strategies to mitigate the impact of a data breach*. *Network Security* (Vol. 2015). https://doi.org/10.1016/S1353-4858(15)70007-3

Edwards, B., Hofmeyr, S., & Forrest, S. (2016). Hype and heavy tails: A closer look at data breaches. *Journal of Cybersecurity*, *2*(1), 3–14. https://doi.org/10.1093/cybsec/tyw003

Ehrlich, M. (2017). Strengthening data breach protection. *Risk Management*, *64*(7), 10–11. Retrieved from https://search.proquest.com/docview/1928357494?accountid=28902

El-Hajj, W., Al-Tamimi, M., & Aloul, F. (2015). Real traffic logs creation for testing intrusion detection systems. *Wireless Communications and Mobile Computing*, *15*(February 2015), 1851–1864. https://doi.org/10.1002/wcm.2471

European Commission. (2016). Directive 95/46/EC (General Data Protection Regulation), *2*, 1–78. Retrieved from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:02016R0679-20160504&from=EN

Experian. (2018). Data breach industry forecast 2018, 1–29. Retrieved from http://www.experian.com/assets/data-breach/white-papers/2018-experian-data-breach-industry-forecast.pdf

Federal Trade Commission. (2016). *Data breach response: A guide for business*. Retrieved from https://www.ftc.gov/system/files/documents/plain-language/pdf-0154_data-breach-response-guide-for-business.pdf

FindLaw. (2005). General Business Law - GBS § 899-aa. Retrieved from https://codes.findlaw.com/ny/general-business-law/gbs-sect-899-aa.html

Golling, M., & Koch, R. (2014). Towards Multi-layered Intrusion Detection in High-Speed Networks, 191–206. Retrieved from https://ieeexplore.ieee.org/document/6916403/

Goyal, P., & Goyal, A. (2017). Comparative Study of two Most Popular Packet Sniffing Tools-Tcpdump and Wireshark. *2017 9th International Conference on Computational Intelligence and Communication Networks*, 77–81. https://doi.org/10.1109/CICN.2017.19

Gupta, A. (2018). The Evolution of Fraud Ethical Implications in the Age of Large-Scale Data Breaches and Widespread Artificial Intelligence Solutions Deployment. *International Telecommunication Union Journal*, (1), 0–7. Retrieved from https://www.researchgate.net/publication/323857997_The_Evolution_of_Fraud_Ethical_Implications_in_the_Age_of_Large-Scale_Data_Breaches_and_Widespread_Artificial_Intelligence_Solutions_Deployment.pdf

Hong, J., Liu, C. C., & Govindarasu, M. (2014). Integrated anomaly detection for cyber security of the substations. *IEEE Transactions on Smart Grid*, *5*(4), 1643–1653. https://doi.org/10.1109/TSG.2013.2294473

Internet Society. (2018). RFC Editor. Retrieved from https://www.rfc-editor.org/

Jacobs, J. (2014). Analyzing Ponemon cost of data breach. Retrieved from http://datadrivensecurity.info/blog/posts/2014/Dec/ponemon/

Jirsik, T., Cermak, M., Tovarnak, D., & Celeda, P. (2017). Toward stream-based IP flow analysis. *IEEE Communications Magazine*, *55*(7), 70–76. https://doi.org/10.1109/MCOM.2017.1600972

Katz, O., Perets, R., & Matzliach, G. (2017). Digging Deeper - An in-depth analysis of a fast flux network, 1–17. Retrieved from https://www.akamai.com/us/en/multimedia/documents/white-paper/digging-deeper-in-depth-analysis-of-fast-flux-network.pdf

Kaur, G., Saxena, V., & Gupta, J. P. (2017). Detection of TCP targeted high bandwidth attacks using self-similarity. *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2017.05.004

Kaur, J., Wendzel, S., Eissa, O., Tonejc, J., & Meier, M. (2016). Covert channel-internal control protocols: attacks and defense. *Security and Communication Networks*, *9*, 2986–2997. https://doi.org/10.1002/sec.1471

Kenkre, P. S., Pai, A., & Colaco, L. (2015). Real Time Intrusion Detection and Prevention System. *Advances in Intelligent Systems and Computing*, *327*, 405–411. https://doi.org/10.1007/978-3-319-11933-5_44

Khader, M., Hadi, A., & Hudaib, A. (2016). Covert communication using port knocking. *Proceedings - 2016 Cybersecurity and Cyberforensics Conference, CCC 2016*, 22–27. https://doi.org/10.1109/CCC.2016.12

Khamphakdee, N., Benjamas, N., & Saiyod, S. (2014). Improving intrusion detection system based on Snort rules for network probe attack detection. *2014 2nd International Conference on Information and Communication Technology, ICoICT 2014*, 69–74. https://doi.org/10.1109/ICoICT.2014.6914042

Kleinrock, L. (1961). *Information flow in large communication nets*. Massachusetts Institute of Technology. Retrieved from https://www.lk.cs.ucla.edu/data/files/Kleinrock/Information Flow in Large Communication Nets.pdf

Kleinrock, L., & Naylor, W. E. (1974). On measured behavior of the ARPA network. In *Proceedings of the May 6-10, 1974, national computer conference and exposition*. https://doi.org/https://doi.org/10.1145/1500175.1500320

Lampson, B. W. (1973). A note on the confinement problem. *Communications of the ACM*, *16*(10), 613–615. https://doi.org/10.1145/362375.362389

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., … Wolf, S. (2009). A brief history of the internet. *ACM SIGCOMM Computer Communication Review*. https://doi.org/10.1145/1629607.1629613

Licklider, J. C. R. (1963). Memorandum for members and affiliates of the Intergalactic Computer Network. Retrieved from http://www.kurzweilai.net/memorandum-for-members-and-affiliates-of-the-intergalactic-computer-network

Liu, L., Jin, X., Min, G., & Xu, L. (2014). Anomaly diagnosis based on regression and classification analysis of statistical traffic features. *Security and Communication Networks*, *7*, 1372–1383. https://doi.org/10.1002/sec.843

Mandiant. (2013). Exposing one of China’s cyber espionage units. Retrieved from https://www.fireeye.com/content/dam/fireeye-www/services/pdfs/mandiant-apt1-report.pdf

Marchetti, M., Pierazzi, F., Colajanni, M., & Guido, A. (2016). Analysis of high volumes of network traffic for Advanced Persistent Threat detection. *Computer Networks*, *109*, 127–141. https://doi.org/10.1016/j.comnet.2016.05.018

McConnell, M. (2017). Admiral Mike McConnell speech at Utica College. Utica: . Retrieved from https://drive.google.com/file/d/0B_KlORlP8jhrc0pGbEN4bFBJM2s/view

Medine, D., Brand, R., Wald, P., Dempsey, J., Cook, E. C., Felten, E., … Solove, D. (2014). *Defining privacy forum*. Washington, D.C. Retrieved from https://www.pclob.gov/library/20141112-Transcript.pdf

Mehic, M., Slachta, J., & Voznak, M. (2016). Whispering through DDoS attack. *Perspectives in Science*, *7*, 95–100. https://doi.org/10.1016/j.pisc.2015.11.016

MIT Lincoln Laboratory. (1998). 1998 DARPA intrusion detection evaluation data set. Retrieved from https://www.ll.mit.edu/r-d/datasets/1998-darpa-intrusion-detection-evaluation-data-set

MIT Lincoln Laboratory. (1999). 1999 DARPA intrusion detection scenario specific data sets. Retrieved from https://www.ll.mit.edu/r-d/datasets/1999-darpa-intrusion-detection-evaluation-data-set

MIT Lincoln Laboratory. (2000). 2000 DARPA intrusion detection scenario specific data sets. Retrieved from https://www.ll.mit.edu/r-d/datasets/2000-darpa-intrusion-detection-scenario-specific-data-sets

Naik, N., Diao, R., & Shen, Q. (2018). Dynamic fuzzy rule interpolation and its application to intrusion detection. *IEEE Transactions on Fuzzy Systems*, *26*(4), 1878–1892. https://doi.org/10.1109/TFUZZ.2017.2755000

National Conference of State Legislatures. (2018). Security breach notification laws. Retrieved from http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx

Ponemon Institute. (2014a). 2014 Cost of Data Breach Study: Global Analysis, (May), 28. Retrieved from https://www-935.ibm.com/services/multimedia/SEL03027USEN_Poneman_2014_Cost_of_Data_Breach_Study.pdf

Ponemon Institute. (2014b). The Aftermath of a Data Breach: Consumer Sentiment, (April), 22. Retrieved from http://www.ponemon.org/local/upload/file/Consumer Study on Aftermath of a Breach FINAL 2.pdf

Ponemon Institute. (2015). 2015 Cost of Data Breach Study: Global Analysis. *Ponemon Institute*, (May), 1–30. Retrieved from https://nhlearningsolutions.com/Portals/0/Documents/2015-Cost-of-Data-Breach-Study.PDF

Ponemon Institute. (2016). *2016 Cost of Data Breach Study : Global Analysis*. *2016 Cost of Data Breach Study: Global Analysis*. Retrieved from https://public.dhe.ibm.com/common/ssi/ecm/se/en/sel03094wwen/SEL03094WWEN.PDF

Ponemon Institute. (2017). *2017 Cost of Data Breach Study, Global Overview*. *IBM Security*. Retrieved from https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=SEL03130WWEN&

Ponemon Institute. (2018). *2018 Cost of a data breach study: Global overview*. Retrieved from https://public.dhe.ibm.com/common/ssi/ecm/55/en/55017055usen/2018-global-codb-report_06271811_55017055USEN.pdf

Postel, J. (1981a). RFC 791: Internet protocol. Retrieved from https://tools.ietf.org/pdf/rfc791.pdf

Postel, J. (1981b). RFC 793: Transmission control protocol. Retrieved from https://tools.ietf.org/pdf/rfc793.pdf

Privacy Rights Clearinghouse. (2018). Data breaches. Retrieved from https://www.privacyrights.org/data-breaches

Rid, T., & Buchanan, B. (2015). Attributing Cyber Attacks. *Journal of Strategic Studies*, *38*(1–2), 4–37. https://doi.org/10.1080/01402390.2014.977382

Romanosky, S., Hoffman, D. a., & Acquisti, A. (2014). Empirical analysis of data breach litigation. *Journal of Empirical Legal Studies*, *11*(1), 74–104. https://doi.org/10.2139/ssrn.1986461

Rowland, C. (1997). Covert Channels.pdf. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/528/449

Ruefle, R., Dorofee, A., Mundie, D., Householder, A. D., Murray, M., & Perl, S. J. (2014). Computer Security Incident Response Team Development and Evolution. *IEEE Security & Privacy*, *12*(5), 16–26. https://doi.org/10.1109/MSP.2014.89

Sestito, G. S., Turcato, A. C., Dias, A. L., Rocha, M. S., Da Silva, M. M., Ferrari, P., & Brandao, D. (2018). A method for anomalies detection in real-time ethernet data traffic applied to PROFINET. *IEEE Transactions on Industrial Informatics*, *14*(5), 2171–2180. https://doi.org/10.1109/TII.2017.2772082

Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, *5*(3), 3. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Singh, G., & Baliya, S. (2015). Detection of malicious traffic and checksum error in network using Wireshark. *International Journal of Scientific Research in Science, Engineering and Technology*, *1*(3), 356–359. Retrieved from www.dell.com

Singh, R., Kumar, H., Singla, R. K., & Ketti, R. R. (2017). Internet attacks and intrusion detection system: A review of the literature. *Online Information Review*, *41*(2), 171–184. https://doi.org/10.1108/OIR-12-2015-0394

The Snort Project. (2018). *Snort 3 user manual*. Retrieved from https://snort-org-site.s3.amazonaws.com/production/release_files/files/000/007/161/original/snort_manual.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Credential=AKIAIXACIED2SPMSC7GA%2F20180823%2Fus-east-1%2Fs3%2Faws4_request&X-Amz-Date=20180823T224533Z&X-Am

Tripathi, N., & Hubballi, N. (2018). Slow rate denial of service attacks against HTTP/2 and detection. *Computers and Security*, *72*, 255–272. https://doi.org/10.1016/j.cose.2017.09.009

Umer, M. F., Sher, M., & Bi, Y. (2017). Flow-based intrusion detection: Techniques and challenges. *Computers and Security*, *70*, 238–254.

https://doi.org/10.1016/j.cose.2017.05.009

University of California Irvine. (1999). KDD cup 1999 data. Retrieved from http://kdd.ics.uci.edu/databases/kddcup99/kddcup99

Verizon. (2018). *2018 Data breach investigations report*. *Verizon Business Journal*. Retrieved from https://www.verizonenterprise.com/resources/reports/rp_DBIR_2018_Report_en_xg.pdf

Wendzel, S., & Keller, J. (2014). Hidden and under control: A survey and outlook on covert channel-internal control protocols. *Annales Des Telecommunications/Annals of Telecommunications*, *69*(7–8), 417–430. https://doi.org/10.1007/s12243-014-0423-x

Wheatley, S., Maillart, T., & Sornette, D. (2016). The extreme risk of personal data breaches and the erosion of privacy. *The European Physical Journal B*, *89*(7).

Yang, S., Wang, J., Zhang, J., & Li, H. (2016). Cyber Threat Detection And Application Analysis. *2016 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC)*, 46–49. https://doi.org/10.1109/CyberC.2016.17

Appendix A

Table A1

Top 100 Data Breaches 2014 – 2018

Company	Records	State	Date Made Public
Yahoo!	3,000,000,000	California	December 14, 2016
Yahoo!	500,000,000	California	September 22, 2016
FriendFinder	412,000,000	California	November 16, 2016
MySpace	360,000,000	California	May 31, 2016
Under Armour	150,000,000	California	March 30, 2018
Equifax Corporation	145,500,000	Georgia	September 7, 2017
Ebay	145,000,000	California	May 21, 2014
LinkedIn	117,000,000	California	May 17, 2016
Anthem	80,000,000	Indiana	February 5, 2015
J.P Morgan Chase	76,000,000	New York	August 28, 2014
T-Mobile	69,600,000	Texas	October 12, 2017
Tumblr	65,469,300	New York	May 13, 2016
Uber	57,000,000	California	November 21, 2017
The Home Depot	56,000,000	Georgia	September 2, 2014
Facebook, Inc.	50,000,000	California	September 28, 2018
Weebly	43,430,300	California	October 20, 2016
Twitter	32,000,000	California	June 13, 2016
Ticketfly	27,000,000	California	June 12, 2018
FourSquare	22,535,000	California	October 21, 2016
Office of Personnel Management (OPM)	21,500,000	District of Columbia	June 4, 2015
Experian	15,000,000	California	October 1, 2015
Premera Blue Cross	11,000,000	Washington	March 17, 2015
Excellus Blue Cross Blue Shield	10,000,000	New York	September 10, 2015
We Heart It	8,000,000	California	October 16, 2017
ClixSense	6,600,000	North Carolina	September 14, 2016

Company	Records	State	Date Made Public
Kansas Department of Commerce	5,500,000	Kansas	July 21, 2017
VTech	5,100,000	Illinois	November 30, 2015
Lord & Taylor's, Saks	5,000,000	New Jersey	April 1, 2018
Scottrade	4,600,000	Missouri	October 1, 2015
Community Health Systems	4,500,000	Tennessee	August 18, 2014
UCLA Health System	4,500,000	California	July 17, 2015
University of California, Los Angeles Health	4,500,000	California	July 17, 2015
Medical Informatics Engineering	3,900,000	Indiana	July 23, 2015
Banner Health	3,620,000	Arizona	August 3, 2016
Adult Friend Finder	3,500,000	Florida	May 22, 2015
Newkirk Products, Inc.	3,466,120	New York	August 9, 2016
21st Century Oncology	2,200,000	Florida	March 4, 2016
America's Job Link Alliance	2,100,000	Kansas	March 27, 2017
Adidas	2,000,000	California	June 28, 2018
Spiral Toys	2,000,000	California	February 27, 2017
PageUp	2,000,000	-	June 12, 2018
Washington Department of Fish and Wildlife	1,700,000	Washington	October 14, 2016
Imgur	1,700,000	California	November 27, 2017
SunTrust Banks, Inc.	1,500,000	Georgia	April 20, 2018
Systema Software	1,500,000	California	September 21, 2015
Schoolzilla	1,300,000	California	April 12, 2017
Staples Inc.	1,200,000	Massachusetts	October 20, 2014
Neiman Marcus	1,100,000	Texas	January 10, 2014
BeautifulPeople.com	1,100,000	New York	April 26, 2016
Montana Department of Public Health & Human Services	1,062,510	Montana	July 7, 2014

Company	Records	State	Date Made Public
Google Android	1,000,000	California	November 30, 2016
Goldenvoice/Coachella Music Festival	950,000	California	March 2, 2017
Valley Anesthesiology Consultants, Inc.	882,590	Arizona	August 12, 2016
Orbitz	880,000	Illinois	March 20, 2018
Goodwill Industries International Inc.	868,000	Maryland	July 14, 2014
Oregon Employment Department/WorkSource Oregon	850,000	Oregon	October 10, 2014
Epic Games Forums	808,000	North Carolina	August 23, 2016
US Postal Service	800,000	District of Columbia	November 10, 2014
County of Los Angeles Departments of Health and Mental Health	749,017	California	December 16, 2016
The Urban Institute	700,000	District of Columbia	February 24, 2015
Virginia Department of Medical Assistance Services	697,586	Virginia	March 12, 2015
Kardashian Website	663,200	California	September 17, 2015
National Stores, Inc.	609,064	California	January 22, 2018
Comcast	590,000	California	November 9, 2015
MSK Group	566,236	Tennessee	May 22, 2018
Georgia Department of Community Health	557,779	Georgia	March 2, 2015
LifeBridge Health, Inc	538,127	Maryland	May 15, 2018
Peachtree Orthopaedic Clinic	531,000	Georgia	November 18, 2016
Airway Oxygen, Inc.	500,000	Michigan	June 16, 2017
Equifax Inc.	431,000	Georgia	May 6, 2016
AU Medical Center, INC	417,000	Georgia	August 16, 2018
St Joseph Health System	405,000	Texas	February 5, 2014
Michigan State University	400,000	Michigan	November 18, 2016
Community Health Plan of Washington	381,504	Washington	December 21, 2016
Disney Consumer Products and Interactive Media	365,000	California	July 30, 2016
Arby's	335,000	Georgia	February 9, 2017
Time Warner Cable	320,000	California	January 8, 2016
University of Maryland	309,079	Maryland	February 19, 2014

Company	Records	State	Date Made Public
Beacon Health System	306,789	Indiana	May 22, 2015
Women's Health Care Group of Pennsylvania	300,000	Pennsylvania	July 26, 2017
Central Ohio Urology Group, Inc.	300,000	Ohio	September 23, 2016
North Dakota University	290,780	North Dakota	March 6, 2014
Oklahoma State University Center for Health Sciences	279,865	Oklahoma	January 5, 2018
Urology Austin, PLLC	279,663	Texas	March 22, 2017
Med Associates, Inc.	276,057	New York	June 14, 2018
Pacific Alliance Medical Center	266,123	California	August 10, 2017
Department of Homeland Security	246,167	District of Columbia	February 1, 2018
Paytime	233,000	Pennsylvania	May 14, 2014
CoPilot Provider Services Inc.	220,000	New York	January 19, 2017
Illinois Board of Elections	200,000	Illinois	August 30, 2016
Delta Air Lines, Inc.	200,000	California	April 6, 2018
Snapsaved.com	200,000	California	October 13, 2014
Guaranteed Rate, Inc.	187,788	Illinois	January 12, 2018
Bizmatics, Inc.	177,000	California	June 17, 2016
Peachtree Neurological Clinic, P.C.	176,295	Georgia	July 7, 2017
Butler University	163,000	Indiana	June 30, 2014
Boxee	158,128	New Jersey	April 2, 2014
Advantage Dental	151,626	Washington	March 16, 2015
[24]7.ai.	150,000	California	April 6, 2018
Total	5,573,149,693		

Note. Top 100 data breaches. Results of a search for exposed records related to hacking-only data breaches occurring between 2014 and October 2018. Adapted from “Data breaches,” by the Privacy Rights Clearinghouse, 2018.