2016-06

A model for real-time data reputation via cyber telemetry

Houser, Beau M.

Monterey, California: Naval Postgraduate School

NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A MODEL FOR REAL-TIME DATA REPUTATION VIA CYBER TELEMETRY

by

Beau M. Houser

June 2016

Thesis Advisor: Dorothy E. Denning
Co-Advisor: Phyllis Schneck

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2016	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE
A MODEL FOR REAL-TIME DATA REPUTATION VIA CYBER TELEMETRY

5. FUNDING NUMBERS

6. AUTHOR(S) Beau M. Houser

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/ MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The federal government faces a monumental task of protecting national security information, advanced warfighting capabilities and the personal information entrusted by hundreds of millions of American citizens. Each federal agency has now identified "High Value Assets" (HVA) as defined by information sets that our adversaries most typically target. The Continuous Diagnostic and Mitigation (CDM) initiative aims to establish a unified security posture across the federal space with a specific focus on HVAs.

This work examines federal cybersecurity initiatives and proposes how data reputation and telemetry can enhance the federal security posture, increase the costs of computer network attack (CNA) of our adversaries, and improve the ability of defenders to drive down the time between when malicious code is observed and when protections are put in place.

14. SUBJECT TERMS
data reputation, federal government, end point protection, telemetry

15. NUMBER OF PAGES
59

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	**18. SECURITY CLASSIFICATION OF THIS PAGE**	**19. SECURITY CLASSIFICATION OF ABSTRACT**	**20. LIMITATION OF ABSTRACT**
Unclassified	Unclassified	Unclassified	UU

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

A MODEL FOR REAL-TIME DATA REPUTATION VIA CYBER TELEMETRY

Beau M. Houser
Deputy Chief Information Security Officer, Centers for Medicare & Medicaid Services
B.S., Barry University, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN CYBER SYSTEMS AND OPERATIONS

from the

NAVAL POSTGRADUATE SCHOOL
June 2016

Approved by: Dr. Dorothy E. Denning
Thesis Advisor

Dr. Phyllis Schneck
Co-Advisor

Dr. Cynthia Irvine
Chair, Cyber Academic Group

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The federal government faces a monumental task of protecting national security information, advanced warfighting capabilities and the personal information entrusted by hundreds of millions of American citizens. Each federal agency has now identified “High Value Assets” (HVA) as defined by information sets that our adversaries most typically target. The Continuous Diagnostic and Mitigation (CDM) initiative aims to establish a unified security posture across the federal space with a specific focus on HVAs.

This work examines federal cybersecurity initiatives and proposes how data reputation and telemetry can enhance the federal security posture, increase the costs of computer network attack (CNA) of our adversaries, and improve the ability of defenders to drive down the time between when malicious code is observed and when protections are put in place.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF ACRONYMS AND ABBREVIATIONS

AM	Anti-malware
APT	Advanced Persistent Threat
ATO	Authority to Operate
AV	Antivirus
BLS	Bureau of Labor and Statistics
CDM	Continuous Diagnostic and Mitigation
CISA	Cybersecurity Information Sharing Act
CMS	Centers for Medicare and Medicaid Services
COTS	Commercial Off the Shelf
CSIP	Cybersecurity Strategy Implementation Plan
DISA	Defense Information Systems Agency
DLL	Dynamic Link Library
DNS	Domain Name Service
DOD	Department of Defense
EPO	ePolicy Orchestrator
EPP	Endpoint Protection
FDA	Food and Drug Administration
FEDRAMP	Federal Risk and Authorization Management Program
GOTS	Government Off the Shelf
GSR	Global Security Ratings
GTI	Global Threat Intelligence
HBSS	Host Based Security Service
HVA	High Value Assets
IOC	Indicator of Compromise
IOC	Initial Operating Capacity
IP	Internet Protocol
ISCM	Information Security Continuous Monitoring
IT	Information Technology
KSN	Kaspersky Security Network
NCCIC	National Cybersecurity and Communications Integration Center

NS	Name Server
OMB	Office of Management and Budget
OPM	Office of Personnel Management
PHI	Personal Health Information
PII	Personally Identifiable Information
PKI	Public Key Infrastructure
ROI	Return on Investment
SIEM	Security Information Event Management
SPE	Sony Pictures Entertainment
SSA	Social Security Administration
STIX	Structured Threat Information eXpression
TAXII	Trusted Automated eXchange of Indicator Information
TCP/IP	Transport Control Protocol/Internet Protocol
URL	Uniform Resource Locator
WoC	Wisdom of the Crowd

ACKNOWLEDGMENTS

I would first like to thank my thesis advisors, Dr. Dorothy Denning of the Naval Postgraduate School and Dr. Phyllis Schneck of the Department of Homeland Security. Dr. Denning's guidance and advice instilled academic rigor that steered me in the right direction. Dr. Schneck's interesting and complex topic truly challenged me. My knowledge and understanding of cyberspace was greatly expanded by Dr. Duane Davis, Dr. Cynthia Irvine, and the entire cyber academic group. I also want to thank my Department of Homeland Security cohort, whose friendship and commitment to the program made the difficult journey a little easier. The support and pressure of my supervisor, Mr. Emery Csulak, were also instrumental to the success of this thesis. Lastly, I want to recognize the sacrifice of my wife. Her unfailing support and continuous encouragement throughout this project were critical to this accomplishment.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

Network defenders have a monumental task when it comes to protecting the information maintained and used to power the world around us. Cybercriminals seemingly pillage unabated through network after network, leaving mayhem and destruction in their wake. Companies are learning that the true cost of cyber security goes far beyond the price of tools and experts. Neglecting information security can result in the loss of market share, customers and reputation, making it impossible to compete in the global economy. Companies and governments must continually work to turn the table on cybercriminals or risk severe consequences. With security of cyberspace leaning significantly toward the offenders, what can be done to improve defenses, detection and response? How do we lower costs of cyber security while raising the cost of computer network attacks? How can detected attacks on one party lead to an improved security posture for another?

As cybersecurity costs continue to rise, companies are forced to find creative ways to increase coverage, visibility, detection and protection while minimizing the associated expense. One field showing promise in achieving these seemingly contrasting motives is security intelligence technologies. According to the Ponemon Institute's 2015 global report on the cost of cyber crime, companies deploying security intelligence technologies realize a 23% return on investment when compared to companies not using these tools. This represented the highest ROI out of seven categories of security-enabling technologies. Examples of security intelligence technologies include security information and event management (SIEM), big data analytics, next-generation firewalls and reputation-based services.

Security intelligence technologies are playing a key role in the evolution and maturation of the security programs in the private sector. Next generation firewalls are able to dynamically block malicious IP addresses using a continually updated data source in the cloud. Malicious documents received by someone in the U.K. can be blocked based on a data reputation score established minutes before by hosts in the U.S. Real-time "whois" information can be used to protect email services against suspicious domain

names that recently stood up. Attackers executing malicious actions against a target may inadvertently raise the defenses of other potential targets benefitting from security intelligence technologies. A robust security intelligence program can turn the tables on attackers and begin to shift the cyber landscape from an offensive advantage to a level battlefield. Security intelligence capabilities, when combined with information sharing, can reduce exposure to emerging threats by focusing on the time period between when a threat is observed until a detection signature is developed and deployed. Creative and resourced malware developers continue to outpace signature-based defensive measures. Software companies mature in the endpoint protection (EPP) market have struggled to keep up with the continually evolving malware landscape. Closing the gap between evolving malware and signature-based technologies is a major challenge for EPP companies. Security intelligence offers numerous capabilities to reduce the exposure. The data reputation element of the security intelligence area strives to fill this gap by scrutinizing the originator's reputation through a crowdsourced methodology. This thesis will explore current commercial data reputation models and propose a federated data reputation solution leveraging telemetric data aggregation.

A. DATA REPUTATION SYSTEMS

Establishing trust is a critical and difficult endeavor in the dynamic and unruly cyber domain. Trust management can be implemented by strong policy-based methods, such as that achieved through a public key infrastructure (PKI) approach. Unlike policy-based trust management, reputation-based systems rely on distributed, self-certified information (Bonatti, Duma, Olmedilla, & Shahmehri, 2005).

MacMillan Dictionary (2016) defines reputation as the opinion that people have about how good or how bad someone or something is. The concept of reputation is an integral social factor that influences countless individual decisions. Business also recognizes how critical reputation can be to the success of their bottom line. Online services have incorporated reputation tools into their transaction to instill trust and protect against cyber threats.

Data reputation systems assign reputation scores to digital objects such as IP addresses, URLs, email addresses and attachments, and software and data files. These scores reflect the extent to which an object is associated with cyber threats and are based on security events and observations. They can be discrete values, such as 0, 1, and 2 for low, medium, and high reputation or continuous values, say between 0 and 1. The scores may be assembled in real time using telemetry data from numerous data sources, including sensors placed on networks and hosts, as well as from services that collect data such as the IP addresses and domains of malicious sites or the signatures of malicious code. Reputation scores are then used by security monitors to determine whether to allow, block or send an alert about an action such as accessing a website, opening an email attachment or running code.

The state of reputation for an object can remain in constant fluctuation based on temporal aggregation of observed events. The speed and dynamic nature of cyberspace and cyber threats requires a data reputation system that can adjust reputation scores in real time with a high degree of confidence. There are several factors required to achieve the speed and confidence needed for an effective data reputation system. Real time reputation scores incorporate numerous data points on a continuous basis and adjust accordingly in a risk range. Legitimate hosts suffering a malware infection will have their reputation score adjusted from low risk to high risk immediately based on observed malicious characteristics. When the infection is cleaned and the host returns to an expected baseline of behavior, the reputation score will climb into a neutral or positive range based on risk calculation algorithms (Barnett, 2010). Confidence is improved as the volume of telemetry data increases. More expansive and diverse input data sources result in a higher level of confidence that the reputation score accurately reflects behavior observed. Currency of data points is an important consideration that directly affects confidence in a data reputation system. Weighing recently observed data over aged observations ensure a reputation score that is reflective of the current state of a host. Data reputation systems also include specific mechanisms to authenticate the information received to ensure it originated from a valid source. The ability to correlate information received from all relevant sources is a strength of data reputation systems and is critical

to maintaining high confidence in the reputation score. Data reputation systems provide insight into the gray space between absolute good and absolute bad (Barnett, 2010). Trust is established and maintained through direct observation of numerous data points from vast and diverse sources allowing organizations to base policy decision on reputation scores to dynamically protect against emerging and evolving threats.

B. PROBLEM STATEMENT

From a cybersecurity perspective, 2014 can be classified as the year of massive data theft and destruction. The year culminated in what experts called the most destructive cyber attack reported to date against a company on U.S. soil (Grover, Hosenball, & Finkle, 2014). This statement referring to the unprecedented, well-planned cyber attack carried out by an organized group believed to be backed by the government of North Korean against Sony Pictures Entertainment (SPE) (Krebs, 2014a) that involved the theft of countless documents and emails as well as the destruction of data on 75% of company computers (Cieply & Barnes, 2014). Not only did the attack disrupt SPE business by causing the company weeks of recovery activities, but the attackers also caused damage to the company's reputation by releasing embarrassing emails, payroll information and sensitive personally identifiable information about employees. The attackers also tried to hurt SPE's bottom line by stealing and releasing upcoming movies. The attackers then took things to a new level when they threatened terrorist actions against movie theaters and movie goers. Never before has our government and economy dealt with an attack of this magnitude and purpose.

Other attacks occurring in 2014 follow a more familiar pattern where data theft is motivated by financial gain with retailers being the most targeted sector. Leading in this category was Ebay, which suffered an attack resulting in the loss of more than 233 million user records containing usernames, passwords, phone numbers and physical addresses (McGregor, 2014). Ebay assured customers that no financial information was stolen, but with the increased risk of identity theft, the loss of confidence is undeniable. Target also suffered a devastating cyber attack compromising personal information of over 70 million shoppers including financial information of 40 million customers (Krebs,

2014b). The financial information quickly found its way onto the black market, resulting in over $50 million in profit for the attackers. This attack ended up costing Target $148 million and financial institutions $200 million (Hardekopf, 2014).

The Target attack shows just how devastating and costly an intrusion can be and should cause companies to reconsider their cybersecurity budget. While increased emphasis on cybersecurity could have prevented the Target attack, it was not enough to prevent one of the largest cyber attacks against the financial sector. With an annual cybersecurity budget of around $250 million, J.P. Morgan is known as a premier banking institution. This budget, however, did not change the fact that the company networks were targeted, compromised and infiltrated (Goldstein, Perlroth, & Corkery, 2014). For two months, cyber criminals were busily collecting account information for over 80 million households and businesses. Although the data did not include social security numbers or account numbers, it is easy to see how the specific information can be used in other directed attacks, like spear phishing.

The tempo of cyber-attacks was sustained in 2015 with a noticeable shift toward bulk personally identifiable information (PII) and personal health information (PHI). Of note is the high profile Office of Personnel Management (OPM) breach that affected over 20 million government employees, contractors and family members (Nakashima, 2015). Included in the heist were the background investigations conducted on employees holding positions of national security. The success of our adversaries in cyberspace has resulted in the accumulation of personal, private, financial, historical, sensitive, travel and other associative information for nearly every government employee (Riley & Robertson, 2015). Combined with personal health information, such as that lost in the healthcare breaches, this allows our adversary to target individuals in a very customized and surgical manner without the need to reconnoiter a potential victim (Riley & Walcott, 2015).

The federal government's responsibility in protecting the massive troves of sensitive information has never been more urgent. Federal- and state-based entitlement programs collect and maintain PII and PHI for a majority of American citizens. The Social Security Administration (SSA), for example, maintains sensitive wage and earnings information on every American from birth and continually updates the data set

throughout their lives (Social Security Administration, 2007). The Centers for Medicare and Medicaid (CMS) administers health care programs that benefit over 100 million Americans, including children, adults and seniors (The Centers for Medicare and Medicaid Services, 2015). CMS also facilitates payments to health care providers in the amount of nearly 1 trillion dollars annually (Health and Human Services, 2015). There are numerous examples of sensitive information collected and maintained by the federal government requiring sophisticated cybersecurity strategies to ensure the protection and assurance from abuse.

The federal government has initiated several projects and programs to achieve an appropriate level of information security. The Continuous Diagnostic and Mitigation (CDM) initiative is one such program designed to achieve fundamental capabilities across several core aspects of cybersecurity in civilian departments and agencies with an ability to extend these capabilities to smaller agencies as well as to state and local governments The CDM program also enables the federal government to leverage the innovation of the private sector and buy the newest and most effective technologies to protect the departments and agencies (Department of Homeland Security, 2015b). Another initiative ordered by the White House's Office of Management and Budget (OMB) created as a result of the major breach suffered by the OPM is the Cybersecurity Strategy and Implementation Plan (CSIP) that came out of the 30 day cybersecurity sprints and focused on key components of cybersecurity to address current threats posed by host nation actors (Donovan, 2015). Congress has also weighed in by passing the Cybersecurity Information Sharing Act (CISA) of 2015 (Congress of the United States, 2015). CISA turns many of the cybersecurity initiatives into law. One capability missing from the current federal strategy is a data reputation repository and service. The ability to associate malicious files, emails or traffic across the federal space with a metric to measure the likelihood of malice in near real-time will significantly enhance the cybersecurity posture of government and shorten the dwell time of malicious software. This thesis argues that the federal government should implement a cloud-based data reputation capability to protect all federal, state, local and private-sector partners. The service will complement the current cybersecurity initiatives being implemented across

government and raise the cost of cyberattacks by sharing information about malicious content in real-time.

C. RESEARCH METHODOLOGY

The commercial industry created an emerging field dedicated to reducing risks associated with depending on signature-based technologies for protection from viruses and malware. Referred to as "security intelligence," this field of information security is quickly becoming a staple of effective cybersecurity programs. One specific element of this field posits that files and content sources can be tracked and monitored for trustworthiness by assigning them reputation scores. There are several examples of commercial data reputation services, each implemented slightly different with various strengths and weaknesses when viewed from the perspective of the federal government. Competition to increase capabilities and protections offered by today's anti-virus (AV) and anti-malware (AM) products has resulted in sector leaders developing near-real time data reputation capabilities that claim to have the ability to protect hosts from newly discovered variants of malware prior to the development and deployment of signatures. These capabilities are based on the providers' ability to collect data from events at their customers, thus creating a crowdsourced set of indicators that can be analyzed to form an assessment of risk to accompany items such as IP addresses, domain names, or malware hashes. This thesis will examine publicly available information from industry leaders to determine common elements of data reputations services. This information will inform a report and recommendation on the best approach for a federal government–wide data reputation service. The security intelligence elements of the following EPP products were chosen based on their rating as leaders in the 2016 Gartner Magic Quadrant for endpoint protection (Firstbrook & Ouellet, 2016):

- Symantec DeepSight
- McAfee Global Threat Intelligence Technology
- Kaspersky Security Network
- Trend Micro Smart Protection Network

- Sophos Live Protection

The thesis will also review alternative data reputation models that do not include EPP software, specifically:

- CISCO Advanced Malware Protection
- Akamai Cloud Security Intelligence

The structure for the remainder of this thesis is as follows:

- Chapter II reviews existing data reputation solutions from industry leaders identified above
- Chapter III identifies the most common and most beneficial elements of data reputation solutions in identifying and mitigating risks
- Chapter IV proposes an effective data reputation solution to implement across the federal government and partners
- Chapter V provides context with existing federal initiatives
- Chapter VI concludes the thesis and suggests further research

II. COMMERCIAL DATA REPUTATION SOLUTIONS

There are several data reputation services available today from commercial vendors. Most have added advanced cloud-based capabilities to AV and AM products in order to strengthen endpoint security and defense. Particularly, companies have added the ability for endpoints to reach out to the cloud, in real time, to determine if the specific communications pattern has been seen before, and if so, whether it is malicious or benign. There are several aspects of this capability that enhance the endpoint security posture.

A. CHARACTERISTICS AND BENEFITS

First, endpoint protection (EPP) software companies have struggled to keep AV and AM products relevant and valuable. When 44% of EPP software customers have suffered from breaches (Firstbrook & Ouellet, 2016), companies are questioning whether their investment into EPP solutions are providing an adequate return on investment. Driving the protection baseline up is the most important goal of the industry and security intelligence capabilities are a major part of the strategy. Second, sector leaders with vast deployments of EPP products have the ability to turn every instance of the deployed software into a standalone sensor able to provide a valuable stream of information regarding real time threats faced by the host. This is a microcosm of the vision for a self-healing Internet, and a contributor to competitiveness for the EPP providers who fear being commoditized as they offer similar services all predicated upon a need to have already formed a signature to recognize incoming malicious traffic.

Companies have developed sophisticated algorithms to inspect telemetric inputs from millions of deployed software agents which act as an Internet early warning system. This aggregation of information provides valuable intelligence and situational awareness of active threat actors, the location of the source and destination of active malicious attempts and the extent of an ongoing outbreak. As threats are unleased on the Internet, EPP software vendors are able to rely on information received from deployed sensors to detect these emerging threats. Mapping the outbreaks as the event progresses and spreads

from the source may provide an indication of the intention of the actor. Similarities among campaigns can also be determined earlier in the attack life cycle and result in quicker response times and improved security postures.

Data reputation determinations are made based on proprietary algorithms developed by each software vendor that consider the culmination of many factors. Often, these algorithms are standard data mining algorithms that can be used in a variety of datasets and are now being applied to cyber security. They tend to be complex and applied in series, so it is nearly impossible to derive the origin of or reason for a specific reputation score given to a specific entity. Longstanding, reputable websites that consistently serve legitimate content are trusted sources of information and have a positive reputation score. Note that a "positive reputation score" for purposes of this thesis means low risk or a good reputation. Providers of reputation scores have a variety of scales, some of which rate the worst reputation, others the best. The field has grown from the ground up by the cyber service providers themselves inventing creative ways to use their client base as sources of cyber threat indicators to make their products more competitive. Malicious traffic detected by sandbox technology platforms and by teams performing forensic analysis identifies the source IP address and URL, along with other relevant details. This information is fed into the reputation algorithm and drives the scores for the identified entities into a negative range. Newly established domains with unknown purpose and little content delivery history are scored as such offering the ability to address this situation with caution.

The accuracy and efficacy of any data reputation service is based largely on the effectiveness of the algorithms and the telemetry ingested from endpoints around the globe. A dedicated team of forensic analysts are also key to maintaining the integrity of the service. Each of the following EPP software vendors are industry leaders and have established robust data reputation services based upon telemetric data consumed from millions of endpoints.

B. SYMANTEC DEEPSIGHT

Symantec's data reputation service evolved from a capability called "Insight," which consisted primarily of a database of all known executables encountered in the digital world (Symantec Corporation, 2012). Through their extensive worldwide network of deployed EPP products, Symantec established a database that stored the name and attributes of every executable file seen. Based on a proprietary method, Symantec evaluated numerous characteristics and attributes and rated each unique file with a trust level. Executables from known software manufacturers were trusted with a high level of confidence while newly discovered executables from unknown vendors were rated as suspicious. Malicious files identified through analysis are also maintained and shared for quicker protection against emerging cyber threats. Today, the product has evolved into a robust, comprehensive, full featured security intelligence service known as "DeepSight," which tracks and maintains numerous attributes of various key indicators of cyber hygiene to determine a risk score for information (Symantec Corporation, 2015a). By establishing a global intelligence network comprised of more than 40 million sensors around the globe, Symantec can track malware, email, web traffic and botnet activity on a continual basis to assess threats posed by the observed information flow. Ownership, established or unknown reputation, and event information and correlation are key indicators used to establish a trust level for Internet domains and uniform resource locators (URL). Tracking malicious software in real-time is another important capability needed to swing the balance from attacker to defender.

C. INTEL SECURITY (MCAFEE) GLOBAL THREAT INTELLIGENCE TECHNOLOGY

At the time McAfee was acquired by Intel Security in 2011, they were already a leader in the data reputation arena. Artemis, Greek Goddess of the Hunt, was the program name originally chosen by McAfee to represent the company's real-time, proactive defense model based on data reputation (Brenesal, 2008). In accordance with the most prevalent threats at the time, Artemis tracked suspicious executables and dynamic link libraries (DLL) active on endpoints running McAfee products. McAfee Avert labs continuously updated the cloud-based service with newly identified malware and

variations of existing malware, enabling endpoints to defend against emerging threats. In 2008, McAfee Inc. acquired Secure Computing Corporation (Intel Security, 2015). Secure Computing specialized in network protection products, and had a reputation system for network traffic, specifically from their firewall, web gateway and mail gateway appliances that looked at widely different telemetry than the McAfee Avert Labs endpoint dataset. The combination of Artemis and the Secure Computing reputation system enabled McAfee to build a suite of security intelligence tools that all used an integrated approach to ingesting, sharing and operationalizing threat intelligence now known as the McAfee Global Threat Intelligence Technology product line (Intel Security, 2016).

McAfee Global Threat Intelligence (GTI) is the component specifically designed as a real-time, cloud-based reputation service that enables endpoints to dynamically defend against emerging attack vectors that manifest in files, web traffic, message traffic, and network traffic. Like the Symantec reputation service, McAfee relies on attributes of files, URLs, domains and IP addresses to provide reputation scores. McAfee combines information from several external sources as well, such as the Cyber Threat Alliance, to enhance the capability. Endpoints protected by McAfee VirusScan Enterprise are able to communicate in real time with the GTI cloud to determine if suspicious files and traffic have been seen before and, if so, obtain a reputation score. Organizations are able to set their risk tolerance level, according to the business need, to maintain an appropriate risk posture and proactively block malicious and suspicious events. Companies and organizations with a lower threshold for risk are able to increase their security posture by blocking unknown, suspicious and malicious files proactively. By only blocking known malicious files, companies are able to maintain an open, collaborative network at a lower security posture. This capability does not exist with most signature base AV and AM software.

McAfee Inc. was acquired by Intel Corporation in 2011 and these capabilities now exist today under the Intel Security brand (Vance, 2010). This sequence of business-driven events demonstrates the key role of corporate mergers and acquisitions and the general global business landscape on cyber security capabilities.

D. KASPERSKY SECURITY NETWORK

The Kaspersky approach to data reputation provides core EPP capabilities. This cloud-based service collects information about suspicious files, executables, websites visited, email attachments and peer-to-peer downloads to determine which may pose a threat to their customers (Kaspersky Lab, 2015). Kaspersky labs analyzes the information collected to determine whether each file or site is safe or malicious. Malicious content is added to the "Urgent Detection System" database while signatures are developed and deployed to endpoints. Confirmed safe content is added to a whitelist that is also continually updated on the endpoints. One method used by Kaspersky to authenticate content is by validating the digital signature vendors are now using to sign code. The goal of KSN is to improve protections for endpoints during the time between malicious software birth and when Kaspersky can deploy a signature for the new threat. One benefit of the reputation database is achieved through Wisdom of the Crowd (WoC) technology where malware can be rated using the popularity reported across the Kaspersky community. Kaspersky also added global security ratings (GSR) which is a customizable algorithm of data reputation elements allowing each organization to create a custom risk tolerance. By combining whitelists, blacklists, signatures, WoC, and GSR, Kaspersky is able to offer a robust EPP product that minimizes the risks associated with new and evolving malware.

E. TREND MICRO SMART PROTECTION NETWORK

As a leader in the EPP market, Trend Micro is well positioned to lead cloud-based data reputation service offerings by commercial vendors. The smart protection network includes several elements that together provide a comprehensive protection strategy for Trend Micro customers (Trend Micro, 2012). According to Trend Micro, there are hundreds of millions of sensors around the globe providing telemetric data and analysis of over one hundred terabytes of files, web addresses, mobile apps, network addresses. This extensive network, along with a cadre of malware and cyber specialists, identifies and protects against over five hundred thousand new threats daily. Web traffic analysis involves collecting all URLs encountered across all Trend Micro endpoints. After

filtering known good and known malicious URLs, analysis begins on remaining suspicious addresses leveraging sandbox technology to identify malicious behaviors. Finally, highly qualified cybersecurity analysts determine whether any outstanding URLs should be deemed malicious and added to the known malicious list. Trend Micro endpoints generate over sixteen billion queries each day, resulting in over two hundred and fifty million blocks of malicious traffic and files. Trend Micro's extensive capabilities are implemented using a cloud-based big data architecture based on the hadoop model. This distributed, unstructured approach to data allows Trend to ingest massive volumes of information. Sophisticated algorithms and big data visualization tools provide predictive intelligence and insight into the ever-changing threat landscape, proving why Trend Micro continues to be an industry leader in EPP.

F. SOPHOS LIVE PROTECTION

The Sophos EPP strategy depends upon a worldwide network of facilities called SophosLabs (Sophos Ltd., 2015). Using a "follow the sun" approach, the labs continually track and respond to emerging threats. Sophos maintains information on every IP address encountered along with IP classifications to create a reputation score and a protection policy. The score is based on several factors beyond whether an address was observed generating malicious traffic or spam messages, factors like whether or not the address is assigned to "end-users" are considered higher risk due to frequent infections affecting individual web users. Hostnames associated with suspicious or malicious IP addresses are also tracked and provided as a protection criteria. Sophos offers live protection as a part of their EPP product capability and as a network-based proxy appliance (Sophos Ltd., 2014). This flexibility allows customers to balance endpoint performance demands with network performance demands while still maintaining a high level of security for the environment.

G. ALTERNATIVE MODELS

Other vendors have created cloud-based data reputation services outside of the agent-based EPP model. Companies with extensive global Internet presence, like Cisco, are leveraging those capabilities to create and enhance visibility of cyber threats,

malicious code and targeted attacks. Cisco's published statistics indicate that they monitor 35% of worldwide email, one hundred terabytes of data and over one million malware samples every day (Cisco, 2015). Cisco's visibility into global Internet traffic through their dominance in the network equipment sector allows them to understand what is happening in the wild, without the need to interact directly with servers, workstations, laptops or mobile devices.

Akamai's content delivery service places them in a unique position between consumers of content and the creators of content. Akamai delivers 15–30% of the world's web traffic each day through more than 1,300 networks, and 175,000 servers in over 100 countries (Akamai, 2015). Akamai's direct visibility into terabytes of web traffic allows them to maintain a very accurate map of the Internet. By associating Internet addresses with traffic patterns, Akamai establishes and maintains a data reputation score for every known address. This information allows customers to refine security defenses by setting a threshold of risk based on the reputation score calculated by Akamai. Customers can choose to automatically block addresses that have been observed sending malicious traffic. The ability for organizations to strengthen perimeter defenses is key to a comprehensive cybersecurity strategy but only when measures are taken to eliminate rogue access points that subvert a trusted Internet connection.

THIS PAGE INTENTIONALLY LEFT BLANK

III. ANALYSIS OF AVAILABLE CLOUD-BASED DATA REPUTATION SERVICES

Technology will continue to expand into every aspect of human existence at an exponential pace leaving security professionals chasing from behind trying to instill structure and safety. This expansion is often referred to loosely as the Internet of things. The field of security intelligence offers a unique opportunity to close the gap between observing malicious intent and building protective measures to address the event. Numerous commonalities exist between the security intelligence capabilities of the products and services discussed in Chapter II. By identifying the core capabilities among competitors, we can establish a baseline of required functions to inform a government-wide security intelligence architecture. Several aspects of data reputation stand out when analyzing dominant industry solutions such as those identified in this work. Uniform Resource Locators (URLs), Internet Protocol (IP) addresses, email attachments and files form the basis of most cloud-based reputation services.

A. UNIFORM RESOURCE LOCATORS

URLs bridge the gap between machine speak and human speak, enabling non-technical individuals to understand and incorporate technology into their everyday lives. URLs also abstract the physical location of a destination host from a logical representation, allowing customers to use a familiar address without worrying about where the service will originate from. This flexibility allows content providers to adjust services and capabilities on the back end while allowing customers to rely on the same, easy to remember, URL (Kunze, 1995). Hackers leverage this flexibility to their advantage by creating URLs in malicious code that can always find a server to receive nefarious instructions from. Hackers commonly use URLs instead of IP addresses because it allows adjustments to be made to the hacker's infrastructure without affecting the malicious code already attached to host systems.

For the Internet to effectively translate URLs to IP addresses, there must be a domain name service (DNS) record associated with every URL that is registered on the

Internet (Mockapetris, 1983). The distributed nature of DNS means that when the name resolution process occurs on a host in a different DNS name space, the lookup will iterate through each level of the DNS structure starting at a top level domain. From there the lookup will locate the name space of the requested URL and iterate down each level until the request reaches the authoritative DNS server that is hosting the record associated with the URL (Microsoft Corporation, 2005). In the case of a malicious actor, the DNS server may be under their control or a compromised server configured to do their bidding.

Just as cyber criminals have discovered how to use the features of the Internet to their advantage, so too have security professionals. By tracking domain registrations and correlating URLs with telemetric data from sensors around the globe, security intelligence is able to quickly determine if a malicious actor is actively engaging in nefarious behaviors and adjust the reputation scores of URLs accordingly, thereby providing near real-time protection without delays associated with traditional signature creation and deployment.

B. INTERNET PROTOCOL ADDRESSES

Since the beginning of civilization, addresses have served as a cornerstone for moving goods and services. Cyberspace is no different. Each and every digital interaction occurring within a computer and without requires addresses. The Internet is built upon the continual disseminated implementation of the transport control protocol/Internet protocol (TCP/IP) which allows expansion of the Internet in a distributed fashion without the need of maintaining a master database of hosts and servers. This model allows the Internet to be extremely resilient because routing protocols are also dynamically built and maintained. Internet outages typically cause routing tables at the edge of the outage to simply update their routing tables and send packets along a different path to their destination (Gerich, 1993).

Requests for services over the Internet typically result in a session being established between the requesting host and the servicing host. Persistent sessions are created using the source and destination IP addresses of the hosts involved in the interaction and the roles of source and destination flip flop depending on the direction of

the packets (Kozierok, 2005). This means that hosts know who they are talking to, especially when secure protocols requiring an identity are used. They also know what information is being transmitted over the session. Visibility and protection from malicious transmissions over a TCP/IP session are key elements of any EPP product. Knowing the reputation of an IP address prior to establishing a session adds a valuable host protective capability that decreases risk. Maintaining a database of all available IP addresses and continually updating their reputation scores each time traffic to or from an address is observed accelerates protection against malicious activity by warning of danger so others can adjust.

It is important to note that as the world and Internet migrate to Internet protocol version 6 (IPv6), databases holding IP addresses will need to be changed to reflect new addresses and, with some degree of difficulty, the infrastructure and capabilities may rely for the foreseeable future upon a mapping of IPv4 to their new IPv6 counterparts. This will affect telemetry solutions and studies that are based on IP addresses.

C. EMAIL ATTACHMENTS

Email has become an integral part of modern life and continues to replace traditional methods of sending and receiving correspondence. In 2015, the average number of emails sent surpassed two hundred billion per day (The Radicati Group, 2015). Combine that statistic with the fact that approximately one in 1,000 emails are phishing attempts (Symantec Corporation, 2015b) targeting individual's personal, financial or corporate information and the scope of the problem is clear. Phishing is a long standing attack vector that continues to pay dividends as cybercriminals continually improve and refine their messages to increase the likelihood of success. Phishing messages are designed according to the objective of the sender. Organizations must protect against emails with generic themes likely to resonate with numerous employees, such as IT support messages or corporate broadcasts. Targeted "spear phishing" emails are also a major concern and much more difficult to detect due to the limited target audience. Spear phishing is a tactic typically used by advanced persistent threat (APT) actors with

specific objectives in mind, such as corporate espionage and intellectual property theft (FireEye Inc, 2014).

By monitoring all email and attachments, cybersecurity defenders can build a database of attachment names, hashes and other relevant details of files which have been determined malicious through sandbox technology, signature-based detection and security analysis. Email servers receiving malicious email query the reputation database prior to delivery of the message to the recipient's inbox. Since reputation scores continually evolve, it is also important that EPP products confirm the safety of attachments at the time the recipient opens the message. This model adds two layers to a defense in depth strategy and allows time for traditional signature based defenses to catch up.

D. MALICIOUS FILES

The field of AV and AM software emerged to confront a steady growing and evolving nuisance that, today, costs organizations and individuals billions of dollars annually (Ponemon Institute, 2015). Lost or corrupt information, system failures and increased maintenance are a few of the cost drivers resulting from malicious code. Creative individuals continue to find interesting ways to manipulate computer systems and their operators by inserting malicious logic and deceiving targeted individuals into executing the code. Once processed, the malicious code usually carries out the objectives of the hacker resulting in loss or damage to the victim.

In today's interconnected world, malicious files are constantly being delivered to potential victims through email, web browsing and file sharing services. Minimizing exposure to unknown files is a key element to protecting information systems, but eliminating the risk entirely is usually not an option. The inherit insecurity of software requires a constant flow of patches and hotfixes from every applicable software vendor. Therefore, tracking every known file encountered on the Internet along with establishing a reputation score increases the security posture for organizations benefitting from this capability. Known safe files from reputable software vendors are tracked and validated continually without impeding system operations. Known malicious files are also

identified and maintained allowing endpoints to immediately block files based on a negative reputation score. All newly discovered files from unknown or unverifiable sources fall into the suspicious category and are continually tracked in order to establish a risk rating. A key goal of the cybersecurity community, and one where reputation technology and the use of telemetry will help, is to shorten the time between first recognition of a malicious indicator to all Internet points having awareness and being able to detect and deflect the new threat.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. DATA REPUTATION ACROSS THE FEDERAL GOVERNMENT

The federal government is responsible for the protection of countless sensitive projects of all sorts and scopes. Next generation weapons systems, intelligence information, law enforcement information, personally identifiable information and protected health information are a few examples of the breadth of information maintained by the federal government. As a result of the Office of Management and Budget's cyber sprints, many projects across several departments and agencies have now been categorized as high value assets, meaning there is a high probability that the systems are actively being targeted by cyber criminals and adversaries of the United States (Scott, 2015). Since each information system could potentially have different threat actors using unique tactics to subvert the security of the system, it is necessary for any federal data reputation solution to be customizable and adaptable from a system perspective and a threat actor perspective. This capability will allow every department and agency to implement a customized risk posture that can be as strict or lenient as needed. Another key benefit of any data reputation solution is the number of sensors participating and providing telemetric data. The federal government has over four million employees and service members. It also spends over five hundred billion dollars on contract support although an exact number of contractors is not known. Establishing a total count of devices across the federal government capable of serving as a sensor providing telemetric data to a centralized threat monitoring service is a difficult task. The continuous diagnostic and mitigation (CDM) program is a federal wide initiative with a phase 1 goal of identifying every device active on every federal network. The CDM program, however, is not mature enough to provide this level of information yet. Having an extensive network of sensors providing real time information about what threats are actively targeting hosts across the federal government is a valuable source of intelligence. This information can be used at both a macro level to provide a mapping of active threats across the federal space and at a micro level specific to a particular department or agency to convey and inform the threat component of the risk equation. Knowing which specific

exploits are actively being attempted by threat actors allows organizations to prioritize existing vulnerabilities and exposures in a cost effective manner. Organizations must understand the threats they face in order to evolve into a risk management model.

A. ALTERNATIVES TO ACHIEVING A CLOUD-BASED DATA REPUTATION SERVICE ACROSS THE FEDERAL GOVERNMENT

There are several models the government could adopt to achieve a cloud-based data reputation service that encompassed the entirety of the federal government. Several levels of complexity and costs are primary factors to consider in weighing the options.

1. Sole Source

One alternative that would require the least amount of time but would most likely offer the least amount of flexibility would be for the federal government to adopt a singular commercially available option, like those described in this thesis. By leveraging the federal acquisitions process, the government could publish a succinct list of EPP requirements. Commercial vendors would then compete by offering valuable tools and services that satisfy the stated requirements. This strategy would allow the government to scrutinize each offer and select the proposal that provides the best value to the government. This approach would result in a uniform solution across the federal space that would lower complexity, simplify administration and establish predictable costs. These expected benefits would come at the cost of flexibility and customization due to the fact that every software vendor develops their products to accommodate a broad customer base with minimal concern for unique requirements that might exist. A sole source arrangement places the government requests for customization and enhancements at a level equal to other major customers of the product vendor. The government will need to ensure the requirements originally stated in the request for proposals fully address each circumstance reasonably expected over the life of the contract and that measures exist to address situations that might arise. The government will also need to develop a strategy that addresses cyber threats that are unique to the federal government, such as host nation actors engaging in espionage and theft of state secrets. This challenge becomes more difficult when viewed through a lens of diversity across the government.

For example, cyber actors targeting intellectual property housed by the Food and Drug Administration (FDA) will most likely use completely different tools techniques and tactics than a hostile nation targeting a contractor facility in order to obtain an advanced weapons system design. The data reputation service acquired by the government will need to accommodate numerous unique scenarios that cross the spectrum of malicious cyber actors.

The Department of Defense attempted to implement a sole-source model when it chose McAfee's EPP suite as its product of choice. Led by the Defense Information Systems Agency (DISA), under the direction of the U.S. Strategic Command, the program is named the Host Based Security System (HBSS) and includes several custom-configured EPP modules all centrally managed using McAfee's ePolicy Orchestrator (EPO) product (Intel Security, 2012). DISA was able to clearly state important and unique requirements that DOD faces, which resulted in a unique build from McAfee that is capable of detecting and protecting against known tactics and techniques that were specific to DOD. Each command and unit was ordered to replace all EPP products running in their environments with the McAfee ePO solution. Administration and maintenance of the product line is also left to each organization.

DOD faced many challenges throughout the implementation of HBSS, although the overall project is regarded as a success. First, many organizations were immature in their EPP strategy. Many simply implemented the required AV and assumed that was sufficient. Migrating from AV to a full suite of products, including intrusion prevention systems, host-based firewalls, application whitelisting and blacklisting and others represented a major move forward in technology that many organizations were not equipped to address. This situation caused training to be a major initiative in the success of the project. To address this challenge, DOD required training for all personnel implementing and administering the product. It even went so far as to require organizations to confirm that each HBSS administrator was trained by asking for their certificates of completion.

Another challenge that threatened the HBSS project was the idea that since the requirement originated from the information security community, it should be left to the

security teams to address. Some organizations ended up with security teams architecting, implementing and maintaining HBSS for the enterprise, leaving the operations staff wondering why their privileged administrators could no longer solve issues. Having security teams run small projects that do not encompass all hosts across the enterprise may be an acceptable arrangement, but with HBSS being a comprehensive suite of EPP tools, this was problematic. System administrators who were used to having free reign to maintain servers, workstations and laptops, were suddenly handcuffed by the new security tools that they had no insight or knowledge of. This led to many wasted hours by administrative staff struggling to achieve basic tasks (like deploying a hotfix) due to the restrictive posture of the endpoint products. Frustrations between the operations staff and the security staff led some to believe that the best course of action was to abandon the project. Other organizations realized very quickly that having the security teams responsible for such a powerful toolset may satisfy security concerns, but might not maintain a healthy balance between security, functionality and cost. Organizations are required to achieve stability among all the pressures influencing goals and priorities in order to maximize effectiveness in accomplishing the mission.

2. Develop a GOTS Solution

Another alternative to achieve a cloud-based data reputation service across the federal government is to plan, architect, design, develop and implement a complete solution from ground up. This approach would allow maximum flexibility to develop a solution that would address the unique requirements of the federal government. Government could follow the lead of private industry and create a federated software package that would implement all the capabilities referenced in Chapter III. The government can develop a logical framework and a modular design allowing organizations to select elements to deploy. The design will also allow organizations to develop custom modules specific to their mission and threats that can plug in to the overall architecture, as well as allow other organizations to leverage existing code.

This approach may seem like a logical choice but there are several challenges. First, the federal government has moved away from the days when computer

programmers were an integral part of the workforce. Today, the majority of programmers work in the private sector for software vendors and IT companies. Most programming tasks occurring for or on behalf of the federal government rely on private contractors. According to the Bureau of Labor Statistics (BLS) there were 130 software developers and programmers employed by the federal government in May of 2014, creating a challenge with taking on a major software development project involving the entire government enterprise (Bureau of Labor and Statistics, 2015). Hiring a private contracting company to accomplish a project of this magnitude requires thoughtful insight up front. Writing a statement of work for a complex endeavor without clear, complete requirements and goals introduces risk to completing the project on schedule and on budget. Poorly written requirements can lead to a project that spirals in circles without meaningful progress toward an ideal end state. On the other hand, properly captured goals and requirements allows the government to hire a well-qualified company with a background in developing similar solutions. Holding the vendor accountable and ensuring the government is receiving valuable services is also possible with a well written contract that clearly states the expectations of the vendor and the government.

For a project of this scope and complexity, the requirements gathering phase alone will be a laborious task, if taken as a federal wide software development initiative. Asking each and every federal department, agency and bureau to clearly state their requirements for a cloud-based data reputation service would be hit or miss at best. In order to achieve an initial operating capability, the government should focus on establishing the requirements for the centralized cloud architecture, code specifications, integration requirements, along with a modular framework capability that is agnostic to specific technologies. Clear specifications will help to identify information, protocols and format for reporting information to the centralized database. This approach will allow organizations to develop modules based on their specific needs and infrastructure. Following today's App model, the government can develop apps at various levels to allow organizations to integrate in a manner that works best for them and supports the overall goal of the program. For agencies with existing capabilities, apps can be leveraged to bridge host-based defense products to the government's cloud-based data

reputation service. By intercepting and transmitting key elements of the data stream generated by host-based products, the government can take advantage of existing capabilities to achieve visibility across organizations and agencies.

The design phase should focus on maximum flexibility and minimal information risk to increase the likelihood of adoption. Organizations need low cost, simple solutions that add value and are likely to work within existing environments. Many organizations are usually not interested in sharing information with external entities when there is the slightest possibility of damage to the agency. The damage can be in the form of a data breach, public perception or reputation. The design must consider how to minimize the amount of information collected and also how to anonymize what is collected. Any information considered sensitive must only be collected when absolutely necessary. Establishing the data reputation capabilities identified earlier requires very specific elements of Internet communications. The source of the communication is often an important characteristic in establishing the trustworthiness of data. The design phase must identify all the needed elements that must be collected. The design will also need to include risk algorithms that take inputs from all required sources needed to calculate a risk score. For example, newly registered domain names are higher risk than long standing, known domains. Ingesting domain registration information from domain registrars is a key element in determining the legitimacy of an Internet domain. For files and attachments, the system should be designed to ingest and track MD5 message digests instead of attempting to obtain a copy of every encountered file. This will minimize the collection of sensitive information such as PII. Files and attachments confirmed to be malicious should be collected in order to fully understand the objective of the malicious actors and establish the specific relationship to the targeted organization. The design should also include the ability to ingest and track IP addresses. Malicious activities should be attributed to the source IP address in order to establish an indicator of trustworthiness. To address the specific belief that information regarding current active threats targeting an organization is sensitive, the program should include measures to protect organizational attribution while also preserving the ability to understand where our enemies are focusing their efforts. These contradictory challenges require not only

technical controls around the data collected, but also administrative controls in the form of service level agreements, data sharing agreements or other instruments of negotiation. The design challenges are significant and must be fully addressed in order to increase the likelihood of success for the project.

During the development phase of the project the team will need to operationalize the design in the most appropriate manner that accounts for all requirements while maximizing flexibility and value to an organization. Considering the federal cloud first initiative, the government should look to FEDRAMP approved cloud service providers first (Kundra, 2010). Cloud-based service providers who have obtained a FEDRAMP provisional authority to operate (ATO) are capable of providing dynamic, elastic services on-demand in a secure manner (FedRAMP Program Management Office, 2014). Next, the project should implement a standard three tiered architecture to optimize and protect the information stored within. The architecture should also include a code promotion and change management process to ensure organization and verification of the software development activities. An integrated collaborative environment with the capability to share code and practices should also be a cornerstone of the development processes. This will allow organizations to leverage existing progress in developing software and shorten the time to adopt the service. The development model should follow an agile methodology which focuses on iterative cycles, known as sprints, to produce the capabilities of the system. Traditional waterfall models do not allow for the flexibility required for this project. Agile also supports the modularity to enable organizations to work together in a synthesized manner where development activities on the server side can be synchronized with development activities at organizations integrating the endpoints. The empirical feedback loop required of the scrum processes should benefit lagging organizations as they leverage lessons learned from early adopters to streamline their implementation. A key capability required to achieve the intelligence of crowds benefit of this project is to ingest telemetric data from millions of endpoints. This means that the system must have the ability to execute billions of transactions daily. The events generated, transmitted and processed by the service must be highly optimized and encrypted to maximize the effectiveness of the program.

During the implementation phase, care must be taken to minimize any negative effects on participating organizations. By starting out with a pilot set of hosts, organizations can control and limit harm caused by any unforeseen technical glitches. Sending and receiving live information to and from endpoints adds a level of complexity that should be closely monitored to ensure an appropriate service level. As hosts are added to the service, the increased throughput might place a strain on the central service. This is where cloud elasticity comes in to offer dynamically expanded resources to meet the increases in performance demands. The core reputation database should be seeded with all known trusted sources of data reputation information gained through open and closed sources along with that gained through partnerships with private and public sector organizations. This will provide an immediate benefit to organizations rather than waiting for the service to develop an aggregated repository of data points over time. Eventually, the federal government data reputation service will evolve into a repository of threats active across the federal space.

The operational phase will involve the continual upkeep of the service, expanding and calibrating threat information sources, tuning existing feeds from endpoints and all of the patching and administration required to maintain the service. Confidence in the reputation rating determined by the service is critical to the on-going success of the project. Therefore, continuous analysis and validation of reputation data is crucial to minimizing false positives and maximizing true positives. This requires a team of dedicated forensics analysts performing on-going analysis of artifacts deemed dangerous by the service. By continually performing quality control, the service will operate in an optimal manner maximizing value to participating organizations. Sophistication of reputation calculation algorithms can also be achieved through the analysis and findings of the forensics team. The service will also require a help desk to intake and manage issues identified by participants. Customer outreach is also an important aspect to success.

3. Hybrid Model

Another approach the federal government can take to implement a cloud-based data reputation model is to combine the elements of a sole source acquisition with those of a GOTS software development initiative in order to maximize the benefits of each. By leveraging existing software solutions, the government can accelerate the process of reaching an initial operating capability (IOC) while working to customize the solution by adding threat information specific to the federal sector. A dedicated data reputation database can be obtained through the acquisition process if clearly stated as a requirement in the request for proposals. In this model, the role of the developers is more aligned with integrating the commercial software with a set of federal sector specific threat information feeds. This approach offers the best of both approaches with costs likely falling between a pure GOTS and pure COTS. Combining capabilities also introduces complexities regarding acquisitions, implementation, customization and support. The acquisition strategy will involve both acquiring a commercial software solution that meets the needs of the program and a software development and integration support contract to customize the software in accordance with federal sector threat protections. Another consideration to take into account is the fact that every EPP software vendor provides many more capabilities beyond data reputation services. Most commercial data reputation services are an extension to endpoint AV, AM, firewall and intrusion detection/prevention. Software call back functionality should be investigated and integrated into the dedicated cloud service whenever practical to avoid unnecessary disclosures to the vendor. Support for technical issues raised by participating organizations, developers and maintainers is also complicated under this approach due to the need to investigate matters prior to assigning the problem to the software vendor, the software developers, the integrators or maintainers for resolution. Establishing a unified helpdesk to support all issues and problems encountered is a key step to ensure the success of the project. Based on the analysis and tradeoffs between the options presented, the optimal strategy to achieve a cloud-based data reputation service that offers sector specific threat protection via a hybrid approach to acquire COTS software from a known, experienced software vendor along with the development and integration services

required. The government should conceive an ideal end state capability and capture all relevant requirements at the beginning of the acquisitions process. If adequately envisioned, the government can accelerate the timeline for implementation while working to integrate all source threat feeds. By combining the approaches and maximizing the benefits of each strategy, agencies and departments can reduce the time to detect known and emerging cyber threats. Combining the approaches also ensure organizations will have a suite of EPP capabilities covering signature-based requirements and supplementing with data reputation services, reducing the exposure between when malicious software is created and when organizations are protected.

The U.S. Department of Homeland Security, with responsibility for providing greater capabilities in cybersecurity to protect federal civilian government and private sector systems, coordinates incident response, CDM and perimeter protection. In the case of federal departments and agencies, this is to augment the security policies, processes and capabilities that the agency CIOs already instantiate. The role of government in this modern view is to leverage the most innovative approaches and datasets from the private sector and combine them with information and capabilities that are unique to government to create a holistic and comprehensive knowledge base to be shared with all. This would indicate that the U.S. Government has been taking a hybrid approach and aggressively using its reputation system with combined private sector and government data, which has already been put into use in the EINSTIEN perimeter protection system for the federal civilian government (The Department of Homeland Security, 2015a).

V. RELATIONSHIP TO FEDERAL INITIATIVES

As long as cyber criminals, hacktivist and host nation actors continue to harm individuals, corporations and government, cybersecurity will remain in the spotlight. Protecting a fundamentally flawed media, like cyberspace, requires a comprehensive approach that leverages all available resources in a coordinated and synthesized manner. Fostering the partnerships and integration sometimes requires synchronized legislation, regulations, policies and procedures. There are a few current examples of federal initiatives that support the implementation of a federal data reputation service.

A. CYBERSECURITY INFORMATION SHARING ACT OF 2015

The Cybersecurity Information Sharing Act (CISA) of 2015 mandates much more than information sharing. There are four sections to the law, each focused on major initiatives across the federal government and the private sector (Congress of the United States, 2015). The first section covers information sharing by and with the federal government and is the section where the law gets its name. The fact that information sharing is covered first indicates just how critical this element is to protecting federal information systems and the dependence of national security. The effectiveness of a federal data reputation service will rely heavily on the ability to quickly share security incident information regarding infections and compromises of IT systems. The broader the scope of threat information sources, the higher the value of the service. CISA compels the federal government to establish an information sharing capability to serve federal, state, local government and also the private sector, to include classified indicators of compromise (IOC). A cloud-based data reputation service could be a central part of an information sharing strategy. Protecting private entities from liabilities incurred by participating in information sharing is also a key element of this section.

The second section of the law focuses specifically on initiatives to strengthen the security posture of federal networks. Mandates for establishing enhanced intrusion detection and prevention, along with mandatory agency participation are intended to strengthen the federal cyber security posture in a short time. Combined with the

aforementioned federal data reputation service, the intrusion detection and prevention system could be a valuable source of information related to malicious and benign hosts and attachments encountered on the Internet. By funneling all agency traffic through a centralize Internet connection and using the intelligence gained from aggregating the intrusion activity of each organization, a broad spectrum of risk will come into context. The value of this perimeter visibility is significant since the threat information is generated using network traffic that has not reached the endpoint. Using intelligence from intrusion detection and prevention tools, especially those with sandbox capabilities, is an important source in a federal data reputation service as it helps move threat detection closer to the source of the attack.

CISA also requires that agencies determine which IT systems contain information that may be of interest to our adversaries. Agencies providing public services have IT systems that contain personal information, which is valuable to cybercriminals seeking financial gain. Host nation intelligence services have also successfully targeted information systems containing rich personal information covering government employees, contractors, families and friends in order to add content and context to the foreign service's dossier on key U.S. Government personnel. Other information sources, such as the Food and Drug Administration's extensive amount of intellectual property, are also highly valuable to cyber criminals and nation states. Maintaining a complete and accurate inventory of valuable data sources not only allows agencies to prioritize cyber activities, it will also serve as a priority grouping of IT systems that might warrant a higher security posture. The data reputation service can be configured to only allow known good attachments, URLs and information sources. Protecting these high value IT systems will involve blocking malicious communications with a negative reputation score and also suspicious communications that may not have a fully established reputation score.

The information security continuous monitoring (ISCM) program has existed for well over a decade and focuses on the importance of monitoring the security posture of information systems on an ongoing basis as a way to make informed risk-based decisions. The current ISCM initiative intends to implement fundamental capabilities that provide

visibility into many aspects of information systems security management. Understanding what assets are connected to federal networks, knowing the software configuration and patching levels and proactively managing privileged users are all key goals of the current continuous diagnostic and mitigation (CDM) program. CISA adds advanced network security tool requirements to the current CDM program in order to improve visibility, detection and mitigation of intrusions and anomalous activities. A federal data reputation service fits this requirement by providing visibility across all participating organizations through telemetric aggregation and by mitigating threats seen on one part of the network across all other networks.

Of critical note is that the information sharing that has been enabled due to this and other legislation is based on the premise that it is coordinated by the Department of Homeland Security in the National Cybersecurity and Communications Integration Center (NCCIC), the cybersecurity watch operations center that manages all operations and incident response for cybersecurity. The NCCIC has implemented real-time machine-to-machine sharing of cyber threat indicators, which are not personal information, but rather hints or signs of threats or those associated with a threat, such as an IP address or hash of a malicious file. Agencies worked together to ensure that privacy and civil liberties were the priority, so that in the necessity to collect a large amount of indicators to inform as many others as possible of a threat, this data collection and real-time distribution does not pose a threat to personal privacy. This is a core tenet at DHS and part of the reason that the NCCIC was chosen and certified as the center for this machine-to-machine information sharing. This is the infrastructure and the way forward to create an ecosystem from the Internet where one intrusion or malicious attempt at one point is a learning moment for all other points on the network or greater Internet.

B. CYBERSECURITY STRATEGY AND IMPLEMENTATION PLAN (CSIP)

Another major initiative set forth by the White House and the Office of Management and Budget is the Cybersecurity Strategy and Implementation Plan (CSIP). There are several similarities between CSIP and CISA since each was developed in parallel. Of the five CSIP objectives, three overlap with CISA. (Scott, 2015) Those focus

on identification of high value IT systems, workforce recruitment and retention and detection and response to malicious activity. Since the three overlapping objectives of CSIP and CISA relate to a federal cloud-based data reputation service in the same way, repeating the benefits is not necessary. The remaining objectives cover incident recovery and streamlined acquisitions of emerging technologies. The ability of agencies to rapidly recover from a cyber event is important to maintaining mission effectiveness. Implementing lessons learned from cyber incidents creates continuous improvement and maturity over time. Information sharing during the recovery phase, including indicators of compromise and lessons learned, will enhance a federal data reputation service. Incident details learned in responding to an incident affecting one organization can be used to help protect numerous other agencies. Information sharing will continue to be a critical component of modern cyber security programs helping to move from reactive to proactive defense and protection. The ability to identify and implement advanced tools and services to secure federal networks is also a challenge. With strict laws and regulations governing federal acquisitions, agencies find themselves frustrated and vulnerable. CSIP establishes specific tasks and deadlines to streamline acquisition challenges that delay acquiring emerging technologies. This objective will have positive affects across government and allow organizations to better protect themselves against emerging threats. Since cybersecurity is a cat and mouse game, being able to adjust quickly and efficiently can make the difference between an effective security posture and a major breach. Since any data reputation service will likely include acquiring products and services, streamlining and adding efficiencies to the acquisitions process will benefit the program.

The focus on cybersecurity is now pervasive across public and private sector organizations. The steps taken by the legislative and executive branches of the federal government demonstrate how critical cybersecurity is to our national security. The steps also reinforce the importance of working together and sharing information. Collaboration is one action that we can take to address this sophisticated continual threat in a meaningful and comprehensive way.

VI. CONCLUSIONS AND FUTURE WORK

Cyberspace has become the new frontier for humanity. Technology continues to expand into every aspect of daily life. The more integrated technology becomes, the more efficient and informed are the subjects. Technology not only informs but also tracks and learns. Aggregating habits, interests, physical movements and personal relationships of technology consumers has proven to be a lucrative approach to businesses in the cyber and physical worlds. Building a personal profile of likes and habits of every connected individual requires astounding amounts of data storage and processing. In 2013, SINTEF reported that 90% of all the data in the world had been created in the previous two years (Dragland, 3013). This trend continues today as companies, governments, criminals and activists find innovative ways to collect and mine vast quantities of personal information. Peter Sondergaard from Gartner Research claimed that information is the oil of the 21st century (Gartner, 2011). With such value placed on data and information, protecting against misuse is now a top priority for governments and companies alike. The future of the Internet must include self-healing hardware and software that recognizes malicious objects and immediately informs and deflects threats across every node. Data reputation services are a key capability on the roadmap that will shift the balance of cyberspace from an environment where attackers hold a dominant advantage, to one where organizations and agencies benefit from each and every malicious action occurring on the Internet.

A. IMPACT

This thesis proposed a model for implementing a federal-wide data reputation service to protect federal agencies by allowing them to make policy decisions based upon the observed and calculated risk. This capability reduces exposures agencies experience between the time a new or evolved threat is observed and when a signature is developed and deployed. The federal government should extend current initiatives using a hybrid model to maximize the benefits of private sector innovations with government strengths and resources. Several commercial software vendors have developed data reputation

capabilities, each implemented differently without a common interoperable specification. Data reputation service is a critical capability in reducing the threats faced by the federal government. The government's ability to implement a federal wide data reputation solution will raise the security posture across all participating departments and agencies.

B. FUTURE WORK

Additional research and analysis is needed in several areas of data reputation. Establishing a standard interoperable specification for interchanging threat information will enhance reputation databases across all participating commercial and government entities. Having commercial EPP vendors add the ability to accept automated threat feeds, such as those implemented using the structured threat information expression (STIX) and trusted automated exchange of indicator information (TAXII) standards, will allow customers and vendors to benefit from information sharing regardless of the specific product. Enabling threat interchange at the end point will create a worldwide intelligence fabric able to report suspicious and malicious objects in near real time. This visibility will result in a fundamental shift in the protection of cyberspace.

Data reputation algorithms are a critical factor in maximizing the effectiveness and confidence of a service that ingests information from numerous diverse threat sources. Expanding current algorithms to include government specific threat sources will ensure comprehensive coverage against actors targeting federal agencies. Progress in the data mining field should be evaluated to determine benefits to the data reputation field.

LIST OF REFERENCES

Akamai. (2015, April 14). Akamai cloud security intelligence provides foundation for new advanced, data-driven cloud security services. Retrieved from Akamai: https://www.akamai.com/us/en/about/news/press/2015-press/akamai-cloud-security-intelligence-provides-data-driven-cloud-security-services.jsp

Barnett, J. (2010). Reputation: The foundation of effective threat protection. Retrieved from SCADAHacker.com: https://scadahacker.com/library/Documents/Threat_Intelligence/McAfee Reputation The Foundation of Effective Threat Protection.pdf

Bonatti, P., Duma, C., Olmedilla, D., & Shahmehri, N. (2005). An integration of reputation-based and policy-based trust management. Retrieved from Rewerse: http://rewerse.net/publications/download/REWERSE-RP-2005-116.pdf

Brenesal, B. (2008, November 19). McAfee total protection 2009 review. Retrieved from Computer Shopper: http://www.computershopper.com/software/reviews/mcafee-total-protection-2009

Bureau of Labor and Statistics. (2015, May). May 2015 national industry-specific occupational employment and wage estimates. Retrieved from U.S. Department of Labor: http://www.bls.gov/oes/current/naics4_999100.htm

The Centers for Medicare and Medicaid Services. (2015). 50 facts in 50 days. Retrieved from The Centers for Medicare and Medicaid Services: https://www.cms.gov/Outreach-and-Education/Look-Up-Topics/50th-Anniversary/50-Facts-in-50-Days-Pt1.pdf

Cieply, M., & Barnes, B. (2014, December 30). Sony cyberattack, first a nuisance, swiftly grew into a firestorm. *New York Times*. Retrieved from http://www.nytimes.com/2014/12/31/business/media/sony-attack-first-a-nuisance-swiftly-grew-into-a-firestorm-.html

Cisco. (2015, November 3). Content security update. Retrieved from Cisco: http://www.cisco.com/assets/global/MK/events/2015/cisco_day/presentations/Gyorgy_Acs-Content_Security_Update.pdf

Congress of the United States. (2015, December). S.754 – Cybersecurity information sharing act. Retrieved from C: https://www.congress.gov/bill/114th-congress/senate-bill/754

Department of Homeland Security. (2015a). Einstein. Retrieved from Official website of the Department of Homeland Security: https://www.dhs.gov/einstein

Department of Homeland Security. (2015b). Continuous diagnostic and mitigation. Retrieved from Official website of the Department of Homeland Security: https://www.dhs.gov/cdm

Donovan, S. (2015, October 30). Cybersecurity strategy and implementation plan (CSIP). Retrieved from The White House: https://www.whitehouse.gov/sites/default/files/omb/memoranda/2016/m-16-04.pdf

Dragland, A. (3013, May 22). Big data – For better or worse. Retrieved from SINTEF: http://www.sintef.no/en/latest-news/big-data--for-better-or-worse/

FedRAMP Program Management Office. (2014, June 6). Guide to understanding FedRAMP. Retrieved from FedRAMP: https://www.fedramp.gov/resources/documents/

FireEye Inc. (2014). Spear phishing attacks – Why they are successful and how to stop them. Retrieved from FireEye: http://www2.fireeye.com/rs/fireye/images/fireeye-how-stop-spearphishing.pdf

Firstbrook, P., & Ouellet, E. (2016, February 1). Magic quadrant for endpoint protection. Retrieved from Gartner: http://www.gartner.com/document/3196523

Gartner. (2011, October 17). Gartner says worldwide enterprise IT spending to reach $2.7 trillion in 2012. Retrieved from Gartner: http://www.gartner.com/newsroom/id/1824919

Gerich, E. (1993, May). Guidelines for management of IP address space. Retrieved from Internet Engineering Task Force: https://tools.ietf.org/html/rfc1466

Grover, R., Hosenball, M., & Finkle, J. (2014, December 3). Sony suffered the most devastating hack of a major U.S. company ever. *Business Insider*. Retrieved from http://www.businessinsider.com/the-size-and-scope-of-the-sony-hack-is-incredible-2014-12

Goldstein, M., Perlroth, N., & Corkery, M. (2014, December 22). Neglected server provided entry for JPMorgan hackers. *New York Times*. Retrieved from http://dealbook.nytimes.com/2014/12/22/entry-point-of-jpmorgan-data-breach-is-identified/?_r=0

Hardekopf, B. (2014, August 11). This week in credit card news: Hack costs Target $148 million, when will interest rates rise? *Forbes*. Retrieved from http://www.forbes.com/sites/moneybuilder/2014/08/11/this-week-in-credit-card-news-hack-costs-target-148-million-when-will-interest-rates-rise/#17024f2973be

Health and Human Services. (2015, February 2). HHS budget in brief – Centers for Medicare and Medicaid Services (CMS) overview. Retrieved from U.S. Department of Health and Human Services: http://www.hhs.gov/about/budget/budget-in-brief/cms/

Intel Security. (2012, March 6). U.S. Department of Defense extends McAfee key role in largest IT security system deployment. Retrieved from Intel Security: http://www.mcafee.com/us/about/news/2012/q1/20120306-02.aspx

Intel Security. (2015). Secure Computing is now part of McAfee. Retrieved from McAfee: http://securecomputing.intelsecurity.com/

Intel Security. (2016, May 1). McAfee GTI reputation & categorization services. Retrieved from Intel Security: http://www.mcafee.com/us/threat-center/technology/gti-reputation-technologies.aspx

Kaspersky Lab. (2015, August). Kaspersky security network. Retrieved from Kaspersky Lab: http://www.kaspersky.com/images/KESB_Whitepaper_KSN_ENG_final.pdf

Kozierok, C. (2005). *TCP/IP guide a comprehensive illustrated Internet protocols reference*. San Francisco: No Starch Press.

Krebs, B. (2014a). FBI: North Korea to blame for Sony hack. Retrieved from KrebsOnSecurity: http://krebsonsecurity.com/2014/12/fbi-north-korea-to-blame-for-sony-hack/

Krebs, B. (2014b). Target: Names, emails, phone numbers on up to 70 million customers stolen. Retrieved from KrebsOnSecurity: http://krebsonsecurity.com/2014/01/target-names-emails-phone-numbers-on-up-to-70-million-customers-stolen/#more-24297

Kundra, V. (2010, December 9). 25 point implementation plan to reform federal information technology management. Retrieved from Federal CIO Council: https://cio.gov/wp-content/uploads/downloads/2012/09/25-Point-Implementation-Plan-to-Reform-Federal-IT.pdf

Kunze, J. (1995, February). Functional recommendations for Internet resource locators. Retrieved from Internet Engineering Task Force: http://tools.ietf.org/html/rfc1736

MacMillan. (2016). Reputation. In *MacMillan Dictionary*. Retrieved from http://www.macmillandictionary.com/us/dictionary/american/reputation

McGregor, J. (2014, July 28). The top 5 most brutal cyber attacks of 2014 so far. *Forbes*. Retrieved from http://www.forbes.com/sites/jaymcgregor/2014/07/28/the-top-5-most-brutal-cyber-attacks-of-2014-so-far/#ad4adf121a65

Microsoft Corporation. (2005, January 21). How DNS query works. Retrieved from Microsoft Technet: https://technet.microsoft.com/en-us/library/cc775637(v=ws.10).aspx

Mockapetris, P. (1983, November). Domain names – Implementation and specification. Retrieved from Internet Engineering Task Force: https://tools.ietf.org/html/rfc883

Nakashima, E. (2015, July 9). Hacks of OPM databases compromised 22.1 million people, federal authorities say. *Washington Post*. Retrieved from https://www.washingtonpost.com/news/federal-eye/wp/2015/07/09/hack-of-security-clearance-system-affected-21-5-million-people-federal-authorities-say/

Ponemon Institute. (2015, October). 2015 cost of cyber crime study: Global. Retrieved from Ponemon Institute: http://www.ponemon.org

The Radicati Group Inc. (2015, March). Email statistics report, 2015–2019. Retrieved from Radicati Group: http://www.radicati.com/wp/wp-content/uploads/2015/02/Email-Statistics-Report-2015-2019-Executive-Summary.pdf

Riley, M., & Robertson, J. (2015, July 29). China-tied hackers that hit U.S. said to breach United Airlines. Bloomberg. Retrieved from http://www.bloomberg.com/news/articles/2015-07-29/china-tied-hackers-that-hit-u-s-said-to-breach-united-airlines

Riley, M., & Walcott, J. (2015, June 5). China's hack of U.S. data tied to health-care record thefts. Bloomberg. Retrieved from http://www.bloomberg.com/news/articles/2015-06-05/u-s-government-data-breach-tied-to-theft-of-health-care-records

Scott, T. (2015, June 17). Fact sheet: Enhancing and strengthening the federal government's cybersecurity. Retrieved from The White House: https://www.whitehouse.gov/blog/2015/06/17/fact-sheet-enhancing-and-strengthening-federal-government-s-cybersecurity

Social Security Administration. (2007, September 27). Privacy impact assessment for the earnings record maintenance system. Retrieved from Social Security Administration: https://www.ssa.gov/foia/piadocuments/FY07/Earnings Record Maintenance System.updtd Sept 28.htm

Sophos Ltd. (2014). Sophos cloud. Retrieved from Sophos: https://secure2.sophos.com/en-us/medialibrary/PDFs/factsheets/sophos-cloud-dsna.pdf?la=en

Sophos Ltd. (2015). Sophos data feeds. Retrieved from Sophos: https://www.sophos.com/en-us/medialibrary/PDFs/partners/sophos-data-feeds-dsna.pdf?la=en

Symantec Corporation. (2012). Turning the tables on malware. Retrieved from Symantec Corporation: http://www.symantec.com/content/en/us/enterprise/white_papers/b-turning_the_tables_on_malware_WP_21155056.en-us.pdf

Symantec Corporation. (2015a). Symantec cyber security services: Deepsight intelligence. Retrieved from Symantec Corporation: https://www.symantec.com/content/dam/symantec/docs/white-papers/deepsight-intelligence-overview-en.pdf

Symantec Corporation. (2015b). Symantec intelligence report. Retrieved from Symantec Corporation: https://www.symantec.com/content/en/us/enterprise/other_resources/b-intelligence-report-01-2015-en-us.pdf

Trend Micro. (2012). Smart protection network. Retrieved from Trend Micro: http://www.trendmicro.com/cloud-content/us/pdfs/about/ds_smart-protection-network.pdf

Vance, A. (2010, August 19). With McAfee deal, Intel looks for edge. *New York Times*. Retrieved from http://www.nytimes.com/2010/08/20/technology/20chip.html

THIS PAGE INTENTIONALLY LEFT BLANK

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California