JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Volume 15, Issue. 2
July 2022
ISSN: 1946-1836

In this issue:

iscap
Information Systems & Computing Academic Professionals

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is three to four issues a year. The first date of publication was December 1, 2008.

JISAR is published online (https://jisar.org) in connection with CONISAR, the Conference on Information Systems Applied Research, which is also double-blind peer reviewed. Our sister publication, the Proceedings of CONISAR, features all papers, panels, workshops, and presentations from the conference. (https://conisar.org)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%), other journal papers (top 30%), unsettled papers, and non-journal papers. The unsettled papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in the JISAR journal. Currently the target acceptance rate for the journal is under 38%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of ISCAP who perform the editorial and review processes for JISAR.

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Editors

2022 JISAR Editorial Board

Examining Cloud Data Security Vulnerabilities During Usage

Daniel Amoah
Azure Solutions Architect
daamoah@microsoft.com
Solutions Architect (Infrastructure and Cyber)
Microsoft Corporation
Denver, CO 80249

Samuel Sambasivam
Samuel.Sambasivam@Woodbury.edu
Computer Science
Data Analytics
Woodbury University
Burbank, CA 91504

Abstract

Cloud computing is a popular computing paradigm with overwhelming benefits, yet there are complex and unresolved cloud data security vulnerabilities in the usage stage of a cloud data life cycle. The purpose of this design science study was to examine cloud data security vulnerabilities during usage by developing a forensic artifact capable of determining cloud data security vulnerabilities. In line with the research question, the study was based on three propositions: 1) that unencrypted data vulnerability is detectable during usage in the cloud, 2) that detectable vulnerable data in the cloud is recoverable using forensics means, and 3) recoverable data is discernable to the extent that it provides value to the data collector. A total of 9 forensics experiments were conducted in three phases using different configurations to collect and analyze the forensic artifacts required to validate or disprove the research propositions. The findings of this design science study showed that both encrypted and unencrypted cloud datasets in memory during cloud data usage are detectable. Detectable unencrypted cloud data during usage is vulnerable, recoverable, and discernable. Encrypted cloud data during usage is also recoverable but not discernable. However, the practicality of homomorphic encryption, which allows the computation of encrypted data, remains a challenge. Therefore, security practitioners must adopt a defense-in-depth strategy that encompasses administrative, physical, and technical controls to minimize the risk of adversary access to volatile memory.

Keywords: Cloud Data Security, Data Lifecycle Security, Data Usage Vulnerability, Cloud Forensics, Memory Forensics.

1. INTRODUCTION

Cloud computing is a new computing paradigm that is more appealing due to benefits such as ubiquitous network access, easy on-demand self-service, rapid resource elasticity, location independence, resource pooling, and usage-based pricing (Sun et al., 2014). The cloud ecosystem can offer better computing services and other benefits such as business agility, cost savings from management, maintenance, and operations than privately owned on-premises data centers (Alam et al., 2018). However, cloud computing has introduced new and complex data

security concerns (Khan et al., 2017; Kumar & Goyal, 2019).

Studies have proposed various procedures to achieve the highest data security level for cloud data protection (Kumar & Goyal, 2019; Matloob, 2017; Mazonka et al., 2020; Singh & Chatterjee, 2017). Subramanian and Jeyaraj (2018) emphasized a need for data protection in all data lifecycle stages in cloud computing. Kacha and Zitouni (2017) described a data lifecycle's usage stage as performing computational processing on cloud data, where risks of misuse or abuse are very high due to many customers in the cloud. According to Mazonka et al. (2020), unlike data in transit and data at rest, which could be protected using encryption, data in use, or performing computation on sensitive data in the cloud, is a single point of failure in computing platforms because current processors operate entirely on plaintexts. To compute on encrypted sensitive data, existing computer architectures must first decrypt, operate on the data, and then re-encrypt. Unencrypted computational data in memory is vulnerable to attack (Singh & Chatterjee, 2017).

Verifying or validating the vulnerability of unencrypted cloud data requires the use of cloud forensic tools and methods (Arshad et al., 2018). However, there are unique challenges in conducting forensics in a public cloud computing environment (Nasreldin et al., 2015). There are architectural, access, jurisdictional, and multi-tenancy challenges associated with a complete forensic analysis of cloud data (Chaudhary & Siddique, 2017). Amato et al. (2020) described a novel semantic approach for conducting digital forensic that enhances evidence discovery and correlation in cloud computing.

This design science research examined the development of a forensic artifact capable of determining cloud data security vulnerabilities during cloud usage. The artifact development consisted of a cloud forensic investigation in different configurations to identify the configurations that offered the most likely source of unencrypted data vulnerability during cloud usage.

Problem Statement

The problem to be addressed in the research study was that the strategies cybersecurity specialists use to mitigate cloud data security vulnerabilities during usage are lacking (Singh & Chatterjee, 2017). Data security and privacy protection concerns remain the most critical issues in cloud computing (Barnwal et al., 2017; ISC2, 2020). According to International Information System Security Certification Consortium (ISC2) 2020 Cloud Data Security report, 69% of organizations are concerned about cloud data loss or leakage (ISC2, 2020). Another report by CloudPassage for Amazon Webservices showed that 63% of organizations are worried about cloud data loss or leakage (CloudPassage, 2020).

Barona and Anita (2017), Kacha and Zitouni (2017), Subramanian and Jeyaraj (2018), and Sun (2020) discussed different types of cloud data security vulnerabilities inherent in the cloud data lifecycle. During the usage stage, when the data is unencrypted, insiders, or outsiders' adversaries with malicious intentions, can gain access to private data used on cloud platforms illegally (Khan, 2016).

Research Question

The research question that guided the study was: What cloud data security vulnerabilities exist during usage? In line with the research question of the study, the following propositions were made:

Prop 1. Unencrypted data vulnerability is detectable during usage in the cloud.

Prop 2. Detectable vulnerable data in the cloud is recoverable using forensics means.

Prop 3. Recoverable data is discernable to the extent that it provides value to the data collector.

2. REVIEW OF THE LITERATURE

This section examined the existing academic and professional literature on cloud data lifecycle security. Cloud computing is a popular computing paradigm with substantial research on multiple interrelated topics, including data security (Barona & Anita, 2017; Kacha & Zitouni, 2017; Subramanian & Jeyaraj, 2018; Sun, 2020). However, as the section illustrates, there are no definitive studies in the literature on cloud data security vulnerabilities in the usage stage (Singh & Chatterjee, 2017).

Security Concerns in Cloud Computing

Over the last ten years, the cloud risk spectrum has expanded due to an increasing growth for cloud-based prospects for business (Kumar & Goyal, 2019). Critical or sensitive cloud storage data can be remotely accessed by attackers who now have the aptitude to utilize users' login information for remote access (Mattoo, 2017; Vumo et al., 2019). Security concerns in the cloud are a significant issue for 94% of

organizations (ISC2, 2020). Another cloud security report by CloudPassage showed that 95% of organizations are concerned about the security of their cloud workloads (CloudPassage, 2020).

Cloud Data Lifecycle Vulnerabilities

There is a need for data protection in all data lifecycle stages (Subramanian & Jeyaraj, 2018). The cloud data lifecycle describes the phases in data from creation to destruction (Kumar et al., 2017). The data lifecycle stages are creation, transmission, storage, usage, sharing, archiving, and disposal (Lin et al., 2014). Creation is the generation of new digital content or updating existing content (Kumar et al., 2017). Storing is the act of committing the digital data to some sort of storage repository and typically occurs nearly simultaneously with creation (Subramanian & Jeyaraj, 2018).

The viewing, processing, or using data in some activity describes the data usage stage (Subramanian & Jeyaraj, 2018). Kacha and Zitouni (2017) described data-in-use as performing computational processing on the cloud data, with a very high risk of misuse or abuse due to many customers in the cloud. The share stage describes activities such as exchanging data between users, customers, and partners (Kumar et al., 2017). In the archive phase, data leaves active use and enters long-term storage (Kumar et al., 2017). The disposal phase describes data destruction using physical or digital means (Kumar et al., 2017). Data deleted from storage media is not entirely erased because file systems cannot remove data; therefore, attackers may use data scavenging techniques to recover deleted data (Khan, 2016).

Data in use and remanence are green pastures for research (Subramanian & Jeyaraj, 2018). There are security vulnerabilities within the SaaS, PaaS, and IaaS models and all the cloud data lifecycle stages (Kumar et al., 2017). It is impossible to process encrypted data either in the cloud environment or in on-premises environments (Kumar et al., 2017). Static data used in cloud applications are usually unencrypted because encrypted data prompts for keys during processing (Kumar et al., 2017).

Encryption

Matloob (2017), Mazonka et al. (2020), and Lo'ai and Saldamli (2019) described encryption as one of the well-known and best solutions for securing data in the cloud. Encryption encodes information into a coded structure and transforms it back to the original state (Matloob, 2017). However, it is impossible to protect data-in-use with encryption either in the cloud environment or in on-premises environments because existing computer architectures must first decrypt, operate on the data, and then re-encrypt (Gaidhani et al., 2017). Other solutions in the academic literature from Alaya et al. (2020), Farokhi et al. (2017), Li et al. (2020), Tran et al. (2020), and Xiong and Dong (2019) focused on using some form of homomorphic encryption schemes to solve the cloud computing data security problems in the usage stage. However, homomorphic encryption has practical implementation challenges for widespread deployment (Alabdulatif et al., 2020; Alloghani et al., 2019; Geng, 2019; Ullah et al., 2019).

Digital Forensics

Digital forensics is a practice that uses scientifically driven and verified methods toward the identification, preservation, acquisition, analysis, interpretation, and documentation of digital data and source analysis and presentation of evidence for reconstructing suspicious events (Palmer, 2001). Digital forensics focuses on forensic procedures, legal approaches, and evidence (Serketzis et al., 2019).

Conducting forensics in a cloud environment is problematic due to the highly distributed and complex cloud architecture (Arshad et al., 2018). Also, established digital forensics practices such as searching and collecting data are not feasible in the public cloud environment due to the lack of individual ownership of devices and the volatile nature of data stored in the cloud (Arshad et al., 2018).

Challenges in Cloud Forensics

There are many unique challenges for conducting digital forensics in a public cloud computing environment (Nasreldin et al., 2015). Some of the cloud forensic challenges include architecture, data collection, evidence analysis, incident first responder, legal, standards, and training (Chaudhary & Siddique, 2017). Other forensic challenges unique to cloud computing are jurisdiction, multi-tenancy, and CSP dependency (Chaudhary & Siddique, 2017).

Traditionally, the forensic investigator controls the evidence collection, but in cloud computing forensics, access to the evidence may not be physically available (Chaudhary & Siddique, 2017). The investigator also faces challenges in analyzing available logs and artifacts (Tak et al., 2018). The forensic investigation challenges in the cloud computing environment are also

related to evidence control, collection, preservation, and validation (Tak et al., 2018). There are also unique digital forensics challenges within the IaaS, PaaS, and SaaS models (Chaudhary & Siddique, 2017).

Gaps in the Literature

Studies have proposed various procedures to achieve the highest data security level for cloud data protection (Kumar & Goyal, 2019; Matloob, 2017; Mazonka et al., 2020; Singh & Chatterjee, 2017). Mazonka et al. (2020) posited that unlike data in transit and data at rest, which could be protected using encryption, data in use, or performing computation on sensitive data in the cloud is a single point of failure in computing platforms because current processors operate entirely on plaintexts. To compute on encrypted sensitive data, existing computer architectures must first decrypt, operate on the data, and then re-encrypt. Public cloud data usage security remains an unresolved concern affecting critical user information privacy and requires more research (Singh & Chatterjee, 2017).

3. METHOD

Design Science was the most appropriate research methodology for this forensic study. According to Edmondson and McManus (2007), implemented research is a mature theory because components used to create an artifact are meticulously studied and documented in the body of knowledge but lacks a developed artifact for the research purpose. Peffers et al. (2007) stated that design science methodology is used to create a knowledge discovery artifact for a research problem. The result of a design science research study is the purposeful creation of an artifact, which can be a product, process, technology, tool, methodology, technique, procedure, or any combination for achieving some purpose (Lapão et al., 2017; Peffers et al., 2007).

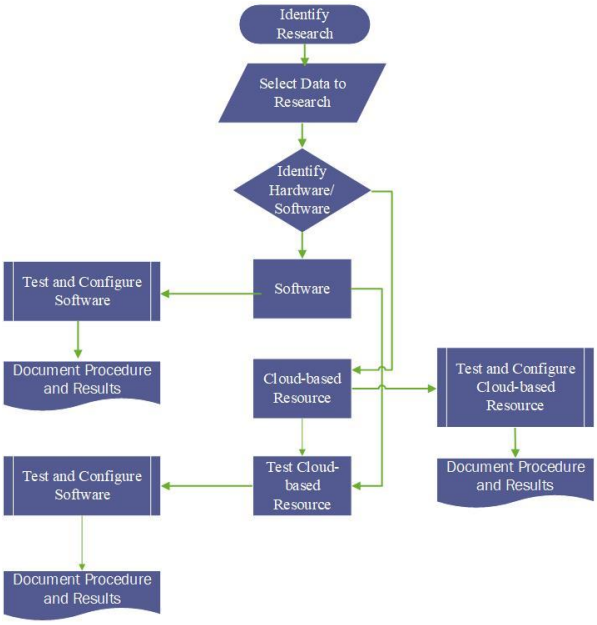
Research Design

The research design was implemented in a standard public cloud operational environment using standard vendor installation instructions. The overall design consisted of two virtual machines (VM) servers hosted in a public cloud, two VM workstations hosted in the public cloud, and a physical workstation. Memory and other research data were collected from the cloud servers using forensics tools and procedures during data computation analysis. The setup of the design allowed for a repeatable process that was easily documented.

Artifact Design

Digital forensics is a practice that uses scientifically driven and verified methods toward the identification, preservation, acquisition, analysis, interpretation, and documentation of digital data and source analysis and presentation of evidence for reconstructing suspicious events (Palmer, 2001). Cloud forensic investigation involves five primary dimensions: data collection, evidence segregation, virtualized environment, preservation of evidence, and reporting and documentation (Chaudhary & Siddique, 2017). Dynamic digital forensics is a forensic data collection and analysis of a running state system or distributed across multiple locations (Arshad et al., 2018). Forensics includes specialized forensic software or hardware that enables a complete digital investigation (Alenezi et al., 2019).

Figure 1
Methodology for Forensic Evaluation



Note. Methodology for forensic evaluation

Forensic methods were used to validate or disprove the research propositions through a rigorous process of data collection. Data collection approaches were tested to identify controlled data sets from the testing environment. The research was conducted in three phases. Phase I of the study involved installing hardware, software, and testing without external or internal manipulations. The VM servers and workstations were deployed in Microsoft Azure public cloud with default settings. Initial data were collected and analyzed

to determine if there were identifiable data to document.

In phase II, controlled use of client-server applications with encrypted cloud data was introduced to the same configuration in phase I. The encrypted data was downloaded to the VM server and opened through a client-server interaction via Simple Message Block (SMB), making the encrypted data available in memory (data-in-use). Data was collected using forensics tools from the Azure VM servers and analyzed. In phase III, the same default configuration settings from phase I was used but with controlled use of client-server applications using unencrypted cloud data to determine data vulnerability in memory. Figure 1 illustrates the methodology used for the forensics evaluation using free and publicly available specialized forensics software (FireEye's Redline) and hardware for the research.

Figure 1 illustrates the basic flow of the methodology used for the forensic evaluation, from identifying the problem, selecting data, identifying hardware and software for testing and configuration, and documenting the procedures and results at each stage.

Collection of Running Memory

Data was collected from the VM servers in the public cloud and examined according to standard forensic guidelines to provide unaltered data supported by documented collection procedures used in each phase of the collection and analysis process. Data were categorized in each phase of the collection process according to data type, date and time collected, test case number, and test case descriptions. Forensics data collection and storage procedures were applied in all data collection for this study.

4. FINDINGS

Description of the Study Sample

The research used random samples of Indicators of Compromise (IOC) obtained from the following publicly available, accessible, and open-source projects:
https://github.com/topics/ioc
https://cyberwarzone.com/download-indicators-of-compromise/
IOCs are forensic artifacts observed in an operating system or on a network and utilized to indicate a computer intrusion and detect cyber-attacks in an early stage (Catakoglu et al., 2016).
The sample IOC data and two non-IOC data were used in the study. Table 1 summarizes the sample data used to validate cloud data security vulnerabilities during usage.

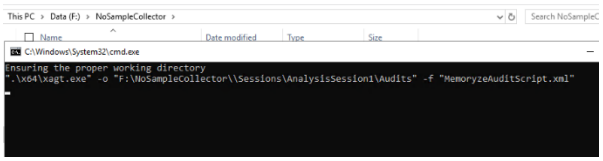
Results

In phase I, the test environment (two VM servers and two VM workstations) was built on Microsoft Azure public cloud with default settings on Windows operating systems as described in Section Three. Various techniques and tools can be employed in digital forensics to analyze live memory (Al-Sharif et al., 2018). The VM servers and workstations were initially analyzed using Redline forensic software and manual hex searches of the file system to ensure the datasets were not present. Figure 2 shows Redline Command run to capture active memory of VM Server1 during interaction with VM Workstation1 with no dataset on the Server. Volatile memory analysis can be performed using four unique methods: file carving, process-object searching, string search, and file signature search (Thantilage & Jeyamohan, 2017). This study used string searches and process-object searches for the analysis of the collected memory artifacts.

Table 1
Description of Sample Data Sets Used in Study

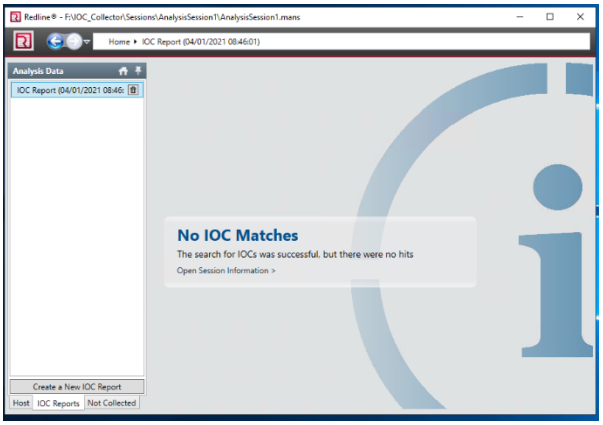
Dataset	Source	Deployment Method	Errors on Client	Operating System
www.apicola.cl	IOC	Notepad	None	Windows Server 2019
halkbankasi.cf	IOC	Word Document	None	Windows Server 2019
paypalll.ga	IOC	Word Document	None	Windows Server 2019
quiroga.cl	IOC	Notepad	None	Windows Server 2019
$Daniel&Amoah$	Non-IOC	Word Document	None	Windows Server 2019
COVID-19	Non-IOC	Word Document	None	Windows Server 2019

Figure 2
Commands run on VM Server1 to Capture Memory with No Dataset



The captured memory data from VM Server 1 was analyzed, as shown in Figure 3. The forensic analysis showed no indication of the presence of the research dataset in memory during the interaction between VM Workstation 1 and VM Server 1.

Figure 3
Forensics Analysis of VM Server1 Memory with No Dataset in Memory

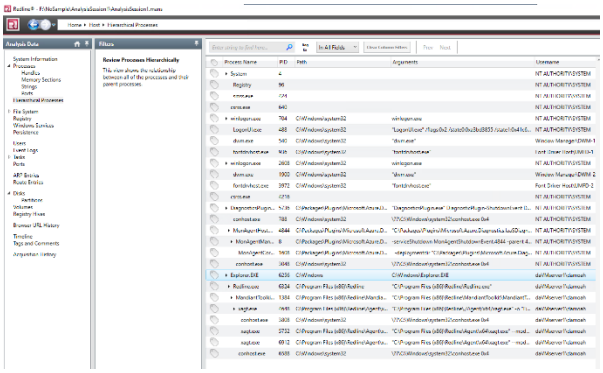


Note. Figure 3 shows an initial view of the IOC search report for possible matches in the sample_ioc dataset in the collected memory.

Figure 3 shows that the captured memory has no elements of the sample_ioc dataset in the memory of VM Server1.

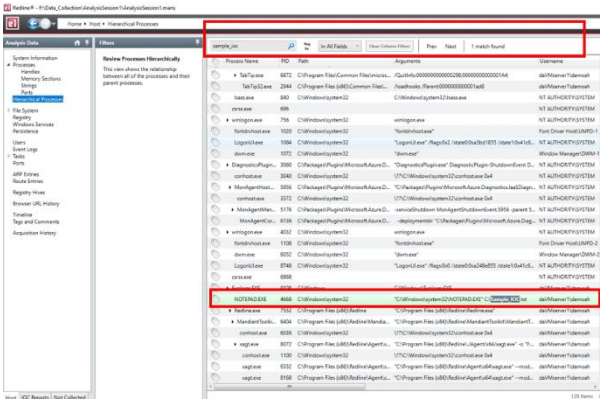
In phase II, controlled use of a client-server application with encrypted cloud dataset was introduced to VM Server1 using methods described in Section Three. The encrypted data was accessed via VM Workstation1 but not decrypted. VM Server1's live memory was captured and analyzed during the client-server application interaction, as shown in Figure 5.

Figure 4
Forensics Analysis of VM Server1 with No Dataset



Note. Figure 4 shows that while no sample_ioc data was found in memory, other data elements not considered were available in memory.

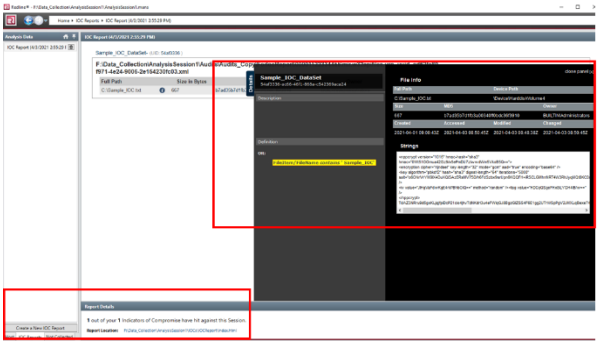
Figure 5
Forensics Analysis of VM Server1 Memory with Encrypted Dataset Match



Note. As shown in Figure 5, the forensics analysis showed the encrypted sample IOC dataset in memory.

A search for "sample_ioc" on hierarchical processes in memory returned one match, but the dataset file was encrypted and, therefore, not discernable. Encrypted dataset elements were detected in the memory analysis of VM Server1 during the client-server interaction.

Figure 6
Forensics Analysis of VM Server1 Memory with Encrypted Dataset Match Details



Note. In Figure 6, the memory analysis of VM Server1 with the encrypted dataset match was expanded to show the contents of the dataset file.

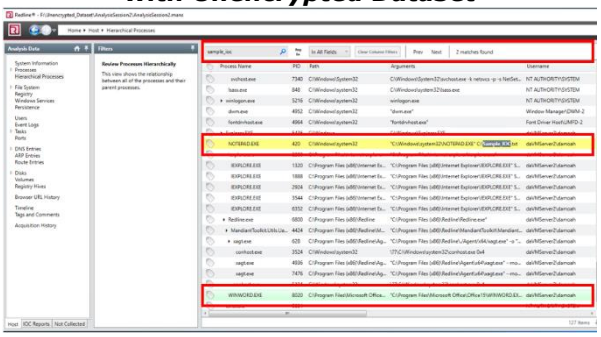
As shown in Figure 6, the contents of the sample_ioc encrypted dataset were not discernable.

Figure 7
Forensics Analysis of VM Server1 with Search Terms for IOC Dataset Elements

Note. In Figure 7, the forensics analysis of VM Server1 Memory was further expanded with specific search terms for known IOC dataset elements in the sample_ioc dataset.

The dataset elements "COVID-19", "paypall.ga", "halkbankasi.cf", and "$Daniel&Amoah$" were used individually at different times as search criteria on the captured memory of VM Server1. Each of the searches resulted in "no matches found." The results clearly showed that an encrypted dataset in memory is not discernable.
In phase III, the unencrypted sample dataset was introduced to VM Server2 with the same default configuration settings as in phases I and II. A client-server application interaction was initiated from VM Workstation2 to VM Server2 to access and use the unencrypted datasets. A live memory of VM Server2 was captured with the forensic tool and analyzed, as shown in Figure 9.

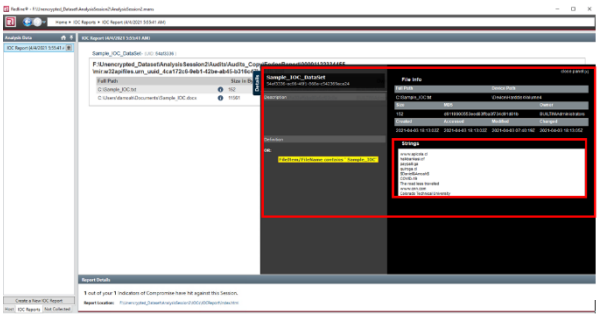
Figure 8
Forensics Analysis of VM Server2 Memory with Unencrypted Dataset



Note. As shown in Figure 8, the forensics analysis showed the unencrypted sample IOC dataset in memory with a search for "sample_ioc on hierarchical processes.

The search returned two matches for sample_ioc datasets in Notepad and Microsoft Word, representing a match for each deployment method for the sample_ioc dataset. However, further trace analysis of the sample_ioc on the captured memory showed all the unencrypted sample_ioc dataset in memory, as shown in Figure 9.

Figure 9
Forensics Analysis of VM Server2 Memory with Unencrypted Dataset Match Details

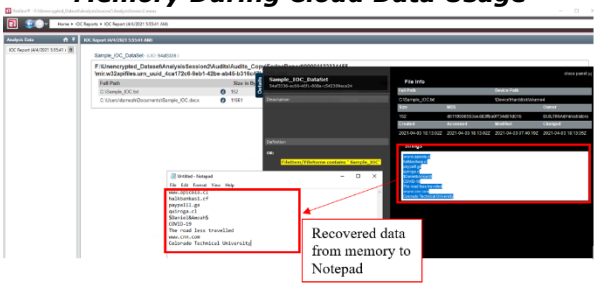


Note. In Figure 9, the complete unencrypted sample_ioc dataset was discernable and accessible in memory.

As shown in figure 9, the IOC search report on the captured memory image returned one match, but the dataset file was encrypted and

not discernable. The unencrypted dataset elements were detected in the memory analysis of VM Server2 during the client-server interaction and usage of data.

Figure 10
Forensic Data Recovery from VM Server2 Memory During Cloud Data Usage



Note. Figure 10 shows a detectable and discernable sample_ioc dataset that was easy to highlight and copy into the Notepad application on a standalone forensic workstation. The copied dataset provides great value to the data collector because it reveals secret information.

The collected artifacts' examination and analysis reviewed three significant themes: data detectability in memory, discernability of data in memory, and recoverability of data in memory.

Data is Detectable During Cloud Data Usage

The collected memory artifacts' analysis showed that both encrypted and unencrypted datasets were detectable in memory during cloud data usage. The artifacts in phases I, II, and III indicate that encrypted and unencrypted data is detectable in memory during usage in the cloud. In phase I, where no sample data was introduced in the examination, collection, and analysis, other non-sample data were observed in memory, as shown in the captured forensic memory analysis in Figure 4. In phase II, encrypted sample_ioc data was introduced to VM Server1, and the encrypted data was accessed via a client-server interaction. The collected live memory analysis showed the encrypted sample_ioc dataset, as shown in Figures 5 and 6. In phase III, the unencrypted sample_ioc dataset was also observed and captured in the analysis shown in Figures 8 and 9. The finding in the three phases addresses the first research proposition: that unencrypted data vulnerability is detectable during usage in the cloud.

Data is Recoverable During Cloud Data Usage

The collected artifacts' analysis showed that detected cloud data in memory could be recovered using forensic tools, as shown in Figure 10. The forensic examination and analysis also showed that both encrypted and unencrypted data could be recovered in memory. However, encrypted data in memory does not provide immediate value to the data collector because data confidentiality is not compromised. On the other hand, unencrypted data in memory is vulnerable and provides immediate value to the data collector because there is no data confidentiality, as shown in Figure 10. The forensic artifact in Figure 10 supports the second research proposition: detectable vulnerable data in the cloud is recoverable using forensic means.

Data is Discernable During Cloud Data Usage

Data discernability describes the ability to identify specific or unique datasets in memory valuable to the data collector. In phase II, the forensic analysis showed that encrypted data in memory is not discernable, as shown in Figure 6. Encrypted data does not reveal any specific data elements and, therefore, retains data confidentiality. Unencrypted cloud data during usage, on the other hand, is discernable in memory, as shown in the collected and analyzed artifacts in Figure 10. Unencrypted data in a file system can be viewed and recovered (Shashidhar & Novak, 2015). The collected forensic artifacts showed that unencrypted cloud data during usage is discernable and, therefore, vulnerable.

5. DISCUSSION

The purpose of the design science study was to examine cloud data security vulnerabilities during usage by developing a forensic artifact capable of determining cloud data security vulnerabilities. The study determined whether unencrypted data vulnerability was detectable, recoverable, and discernable during usage in the cloud.

Theme 1: Defense-in-Depth Strategy to Safeguard Data Detectability in Memory

As indicated by the collected memory artifacts, encrypted and unencrypted cloud datasets in memory during cloud data usage are detectable. The ability to detect datasets in memory during cloud data usage means data is vulnerable while in memory. Since data in memory is detectable, unencrypted data in memory is a serious threat to data security. There is, therefore, a need for cybersecurity specialists and practitioners to consider strategies and technologies to protect data in memory.

There are different strategies and approaches for safeguarding datasets in memory. According to Mazonka et al. (2020) and Lo'ai and Saldamli (2019), one of the well-known and best solutions for securing datasets in the cloud is encryption. Encryption is a process that converts plaintext data into cyphertext. However, it is currently impractical to protect data-in-use with encryption (Gaidhani et al., 2017; Kumar et al., 2017; Miyan, 2017). Homomorphic encryption is an encryption scheme that allows computation on encrypted data without first decrypting the data (Gaidhani et al., 2017). However, homomorphic encryption has practical implementation challenges for widespread deployment and adoption (Alabdulatif et al., 2020; Alloghani et al., 2019; Geng, 2019; Ullah et al., 2019).

A significant part of the data detectability in memory vulnerability is access to the volatile computer memory. It is, therefore, critical for cybersecurity specialists and practitioners to adopt comprehensive layers of different controls (defense-in-depth) to minimize the risk of access to the vulnerable memory (Mazonka et al., 2020; Rocha et al., 2013). Controls such as policies, identity and access management, personnel security, physical security, network security, host-based security, and application security, among other controls, effectively reduce the risk (Jeganathan, 2018). Cybersecurity specialists can implement layers of technical and administrative controls to reduce the risk of vulnerabilities (Kumar & Goyal, 2019).

Theme 2: Use Available CSP Tools and Controls to Reduce Recoverability of Data in Memory

Recoverability of data in memory was the next theme from the findings of the collected and analyzed artifacts in phase III. The forensic examination and analysis showed that both encrypted and unencrypted data could be recovered in memory. The study artifacts showed that encryption provides data confidentiality because recovered encrypted datasets from memory remained encrypted and did not reveal any data secrets to the data collector. The study has shown that encrypted cloud data remained encrypted when accessed through client-server interaction. However, performing a computation or using encrypted data in computing platforms remains a challenge because current processors operate entirely on plaintexts (Mazonka et al., 2020).

The study also showed that unencrypted cloud data in use are vulnerable and recoverable. It is, therefore, critical for cybersecurity specialists and practitioners to adopt available cloud service provider (CSP) tools and strategies to secure cloud data during usage. For instance, within the Azure cloud platform, enabling Just-in-Time VM access restricts the VM's management ports and grants access on-demand for a limited time to only pre-approved IP addresses. Using a bastion service to connect the VMs also protects the VMs against exposing the public IP on the VM. Using conditional access policies to restrict access and auto-shutdown VMs also reduces the risk of data recoverability in memory. There are multiple administrative and technical controls and strategies to safeguard unencrypted data in memory to prevent unauthorized recoverability (Subramanian & Jeyaraj, 2018). There is no silver bullet when it comes to protecting unencrypted data in use. No single technology ultimately provides the required protection (CSA, 2017). However, using available CSP tools and controls to enforce administrative and technical controls reduces the risk of recovering unencrypted data from memory.

Theme 3: Device Management and Isolation to Reduce Discernability of Data in Memory

The study artifacts showed that collected encrypted cloud data usage in memory is not discernable, as demonstrated in phase II. It is impossible to identify unique data elements from encrypted cloud data collected from memory without decrypting the data, as shown in Figure 6. On the other hand, unencrypted cloud data in use is vulnerable, recoverable, and discernable without decrypting the collected data, as shown in Figure 10 in the study artifacts. Unencrypted discernable data in memory is vulnerable to bus snooping attacks (Tavana et al., 2017). The risk of volatile memory vulnerability depends on access to the cloud-based resources memory; therefore, cybersecurity specialists and practitioners should implement strong authentication mechanisms through identity and access control, device management, zero-trust security model principles, and device isolation as part of broader layers of controls to minimize the risk to unencrypted data in use.

6. CONCLUSIONS

The results of the design science study showed that data could be detected during cloud usage in memory. The results also indicated that cloud data detected during usage could be recovered from memory. Finally, the results showed that encrypted cloud data usage in memory was not

discernable while unencrypted cloud data in use was vulnerable, recoverable, and discernable.

The findings of this study apply to all information technology settings that use sensitive data in public cloud computing. A quantitative or qualitative study on cloud data usage security would add to the body of knowledge a comprehensive list of practical approaches cybersecurity professionals can use to minimize the risk of cloud data usage vulnerability. The practicality of homomorphic encryption also requires more research.

7. ACKNOWLEDGEMENTS

I want to acknowledge the support, encouragement, and guidance I received from my research supervisor, Dr. Samuel Sambasivam. Thank you, Dr. Sambasivam. Your counsel made the path to success on the doctoral journey clearer.

8. REFERENCES

Alabdulatif, A., Khalil, I., & Yi, X. (2020). Towards secure big data analytic for cloud-enabled applications with fully homomorphic encryption. *Journal of Parallel and Distributed Computing*, *137*, 192-204. https://doi.org/10.1016/j.jpdc.2019.10.008

Alam, S., Muqeem, M., & Suhel, A. K. (2018). Review on security aspects for cloud architecture. *International Journal of Electrical and Computer Engineering, 8*(5), 3129-3139. http://doi.org/10.11591/ijece.v8i5.pp3129-3139

Alaya, B., Laouamer, L., & Msilini, N. (2020). Homomorphic encryption systems statement: Trends and challenges. *Computer Science Review*, *36*, 100235. https://doi.org/10.1016/j.cosrev.2020.100235

Alenezi, A., Atlam, H. F., & Wills, G. B. (2019). Experts reviews of a cloud forensic readiness Framework for organizations. *Journal of Cloud Computing*, 8(1), 1-14. http://doi.org/10.1186/s13677-019-0133-z

Alloghani, M., Alani, M. M., Al-Jumeily, D., Baker, T., Mustafina, J., Hussain, A., & Aljaaf, A. J. (2019). A systematic review on the status and progress of homomorphic encryption technologies. *Journal of Information Security and Applications*, *48*, 102362. https://doi.org/10.1016/j.jisa.2019.102362

Al-Sharif, Z. A., Bagci, H., Zaitoun, T. A., & Asad, A. (2018). Towards the memory forensics of MS word documents. In: Latifi S. (eds) Information Technology - New Generations. Advances in Intelligent Systems and Computing, vol 558. Springer, Cham. https://doi.org/10.1007/978-3-319-54978-1_25

Amato, F., Castiglione, A., Cozzolino, G., & Narducci, F. (2020). A semantic-based methodology for digital forensics analysis. *Journal of Parallel and Distributed Computing*, *138*, 172-177. https://doi.org/10.1016/j.jpdc.2019.12.017

Arshad, H., Jantan, A. B., & Abiodun, O. I. (2018). Digital Forensics: Review of Issues in Scientific Validation of Digital Evidence. *Journal of Information Processing Systems*, *14*(2). https://doi.org/10.3745/JIPS.03.0095

Barnwal, A., Pugla, S., & Jangade, R. (2017). Various security threats and their solutions in cloud computing. https://doi.org/10.1109/ccaa.2017.8229923

Barona, R., & Anita, E. A. M. (2017). A survey on data breach challenges in cloud computing security: Issues and threats. https://doi.org/10.1109/iccpct.2017.8074287

Catakoglu, O., Balduzzi, M., & Balzarotti, D. (2016, April). Automatic extraction of indicators of compromise for web applications. In *Proceedings of the 25th international conference on world wide web* (pp. 333-343). https://doi.org/10.1145/2872427.2883056

Chaudhary, O., & Siddique, A. S. (2017). Cloud computing application: Its security issues and challenges faced during cloud forensics and investigation. *International Journal of Advanced Research in Computer Science, 8*(2). http://www.ijarcs.info/index.php/Ijarcs/article/view/2916

CloudPassage. (2020). AWS Cloud Security Report.
https://pages.cloudpassage.com/rs/857-FXQ-213/images/2020-AWS-Cloud_Security-Survey-Report.pdf

CSA. (2017). *Cloud Controls Matrix*. https://cloudsecurityalliance.org/research/ccm/

Edmondson, A. C., & McManus, S. E. (2007). Methodological fit in management field research. *Academy of management review*, *32*(4), 1246-1264. https://doi.org/10.5465/amr.2007.26586086

Farokhi, F., Shames, I., & Batterham, N. (2017). Secure and private control using semi-homomorphic encryption. *Control Engineering Practice*, *67*, 13-20. https://doi.org/10.1016/j.conengprac.2017.07.004

Gaidhani, D., Koyeerath, J., Kudu, N., & Mehra, M. (2017). A survey report on techniques for data confidentiality in cloud computing using homomorphic encryption. *International Journal of Advanced Research in Computer Science, 8*(8). https://doi.org/10.26483/ijarcs.v8i8.4746

Geng, Y. (2019). Homomorphic encryption technology for cloud computing. *Procedia Computer Science*, *154*, 73-83. https://doi.org/10.1016/j.procs.2019.06.012

ISC2. (2020). 2020 Cloud Security Report. https://www.isc2.org/resource-center/reports/2020-cloud-security-report

Jeganathan, S. (2018). Practical approaches to overcome security challenges in cloud Computing. *ISSA Journal*, *16*(12), 30–41.

Kacha, L., & Zitouni, A. (2017, September). An overview on data security in cloud computing. *In Proceedings of the Computational Methods in Systems and Software* (pp. 250-261). Springer, Cham. https://doi.org/10.1007/978-3-319-67618-0_23

Khan, H., Ahamad, M. V., & Samad, A. (2017). Security challenges and threats in cloud computing systems. *International Journal of Advanced Research in Computer Science, 8*(2).

Khan, M. A. (2016). A survey of security issues for cloud computing. *Journal of Network and Computer Applications*, *71*, 11-29. https://doi.org/10.1016/j.jnca.2016.05.010

Kumar, R., & Goyal, R. (2019). On cloud security requirements, threats, vulnerabilities and countermeasures: A survey. *Computer Science Review*, *33*, 1-48. https://doi.org/10.1016/j.cosrev.2019.05.002

Kumar, S., Verma, R. S., & Mohan, K. (2017). Survey on data security issues in cloud computing. *International Journal of Advanced Research in Computer Science, 8*(3)

Lapão, L. V., da Silva, M. M., & Gregório, J. (2017). Implementing an online pharmaceutical service using design science research. *BMC Medical Informatics and Decision Making*, *17*(1). https://doi.org/10.1186/s12911-017-0428-2

Li, J., Kuang, X., Lin, S., Ma, X., & Tang, Y. (2020). Privacy Preservation for Machine Learning Training and Classification Based on Homomorphic Encryption Schemes. *Information Sciences*. https://doi.org/10.1016/j.ins.2020.03.041

Lin, L., Liu, T., Hu, J., & Zhang, J. (2014, December). A privacy-aware cloud service selection method toward data lifecycle. In *2014 20th IEEE International Conference on Parallel and Distributed Systems (ICPADS)* (pp. 752-759). IEEE. https://doi.org/10.1109/PADSW.2014.7097878

Lo'ai, A. T., & Saldamli, G. (2019). Reconsidering big data security and privacy in cloud and mobile cloud systems. *Journal of King Saud University-Computer and Information Sciences*. https://doi.org/10.1016/j.jksuci.2019.05.007

Matloob, G. (2017). A Survey on cloud computing security issues and its possible solutions. *International Journal of Advanced Research in Computer Science*, *8*(2).

Mattoo, I. A. (2017). Security issues and challenges in cloud computing: A conceptual analysis and review. *International Journal of Advanced Research in Computer Science*, *8*(2).

Mazonka, O., Sarkar, E., Chielle, E., Tsoutsos, N. G., & Maniatakos, M. (2020). Practical data-in-use protection using binary decision diagrams. *IEEE Access*, *8*, 23847-23862. https://doi.org/10.1109/ACCESS.2020.2970120

Miyan, M. (2017). FHE implementation of data in cloud computing. *International Journal of Advanced Research in Computer Science*, *8*(3).

Nasreldin, M. M., El-Hennawy, M., Aslan, H. K., & El-Hennawy, A. (2015). Digital forensics evidence acquisition and chain of custody in cloud computing. *International Journal of Computer Science Issues (IJCSI), 12*(1), 153-160.

Palmer, G. (2001). *A roadmap for digital forensics research*. Report for the First Digital Forensics Research Workshop (DFRWS), DTR-T0010-01, DFRWSDTR-T0010-01.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems, 24*(3), 45-77. https://doi.org/10.2753/MIS0742-1222240302

Rocha, F., Gross, T., & Van Moorsel, A. (2013). Defense-in-depth against malicious insiders in the cloud. In *2013 IEEE International Conference on Cloud Engineering (IC2E)* (pp. 88-97). IEEE. https://doi.org/10.1109/IC2E.2013.20

Serketzis, N., Katos, V., Ilioudis, C., Baltatzis, D., & Pangalos, G. J. (2019). Actionable threat intelligence for digital forensics readiness. *Information and Computer Security*, 27(2), 273-291. https://doi.org/10.1108/ICS-09-2018-0110

Shashidhar, N. K., & Novak, D. (2015). Digital forensic analysis on prefetch files. *International Journal of Information Security Science*, *4*(2), 39-49.

Singh, A., & Chatterjee, K. (2017). Cloud security issues and challenges: A survey. *Journal of Network and Computer Applications*, *79*, 88-115. https://doi.org/10.1016/j.jnca.2016.11.027

Subramanian, N., & Jeyaraj, A. (2018). Recent security challenges in cloud computing. *Computers & Electrical Engineering*, *71*, 28-42. https://doi.org/10.1016/j.compeleceng.2018.06.006

Sun, P. J. (2020). Security and privacy protection in cloud computing: Discussions and challenges. *Journal of Network and Computer Applications*, 102642. https://doi.org/10.1016/j.jnca.2020.102642

Sun, Y., Zhang, J., Xiong, Y., & Zhu, G. (2014). Data security and privacy in cloud computing. *International Journal of Distributed Sensor Networks,* https://doi.org/10.1155/2014/190903

Tak, V., Kachhwaha, R., & Mahia, R. N. (2018). Secure log forensics as a service in cloud computing. *International Journal of Advanced Research in Computer Science, 9*(1). http://doi.org/10.26483/ijarcs.v9i1.5373

Tavana, M. K., Fei, Y., & Kaeli, D. R. (2017). Nacre: Durable, secure and energy-efficient non-volatile memory utilizing data versioning. *IEEE Transactions on Emerging Topics in Computing*. http://doi.org/10.1109/TETC.2017.2787622

Thantilage, R., & Jeyamohan, N. (2017, September). A volatile memory analysis tool for retrieval of social media evidence in windows 10 OS based workstations. In *2017 National Information Technology Conference (NITC)* (pp. 86-88). IEEE. https://doi.org/10.1109/NITC.2017.8285664

Tran, J., Farokhi, F., Cantoni, M., & Shames, I. (2020). Implementing homomorphic encryption based secure feedback control. *Control Engineering Practice*, *97*, 104350. https://doi.org/10.1016/j.conengprac.2020.104350

Ullah, S., Li, X. Y., Hussain, M. T., & Lan, Z. (2019). Kernel homomorphic encryption protocol. *Journal of Information Security and Applications*, *48*, 102366. https://doi.org/10.1016/j.jisa.2019.102366

Vumo, A. P., Spillner, J., & Köpsell, S. (2019, July). A Data security framework for cloud computing adoption: Mozambican

government cloud computing. In *European Conference on Cyber Warfare and Security* (pp. 720-XX). Academic Conferences International Limited.

Xiong, L., & Dong, D. (2019). Reversible data hiding in encrypted images with somewhat homomorphic encryption based on sorting block-level prediction-error expansion. *Journal of Information Security and Applications*, *47*, 78-85. https://doi.org/10.1016/j.jisa.2019.04.005

Cybersecurity Maturity Model Certification Initial Impact on the Defense Industrial Base

Hala Strohmier
strohmierh@uncw.edu

Geoff Stoker
stokerg@uncw.edu

Manoj Vanajakumari
vanajakumarim@uncw.edu

Ulku Clark
clarku@uncw.edu

Jeff Cummings
cummingsj@uncw.edu

Minoo Modaresnezhad
modaresm@uncw.edu

University of North Carolina Wilmington
Wilmington, NC 28412 USA

Abstract

The Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) published the Cybersecurity Maturity Model Certification (CMMC) framework in January 2020. The CMMC is a major effort intended to strengthen the ability of Defense Industrial Base (DIB) members to protect Federal Contract Information (FCI) and Controlled Unclassified Information (CUI). In this article, we briefly recount the history of unclassified information handling in the U.S. Federal Government that led to the current situation and explain why the CMMC was created, what it is, and what it entails. Through a series of interviews with a convenience sample of current large and small DIB members, we explore some of the perceptions, perceived challenges, and expected impacts of the CMMC on the DIB. We also consider the chances that the CMMC will accomplish its intended goals and describe a planned future larger study of the CMMC effort and its effects on the DIB.

Keywords: Cybersecurity Maturity Model Certification (CMMC)

1. INTRODUCTION

In February 2018, the Council of Economic Advisors (CEA, 2018) released a report that estimated the cost of malicious cyber activity to the U.S. economy in 2016 was between $57 and $109 billion. These costs stemmed from 42,000+ cybersecurity incidents that compromised the confidentiality, integrity, and/or availability (CIA) of information systems

and nearly 2,000 breaches resulting in confirmed unauthorized disclosure of data.

In addition to outright theft of intellectual property, there is concern, heightened since the 9/11 attacks, that the loss of many small pieces of seemingly insignificant information can aggregate to create a grave intelligence concern (Pozen, 2005). Referred to by some as the mosaic theory, this is where:

> Disparate items of information, though individually of limited or no utility to their possessor, can take on added significance when combined with other items of information. Combining the items illuminates their interrelationships and breeds analytic synergies, so that the resulting mosaic of information is worth more than the sum of its parts. (Pozen, 2005, p. 630)

Most of the data theft appears to be attributable not to a lack of effective security control guidance, but rather to poor cybersecurity habits and posture. Because of this, the Department of Defense (DoD) has embarked on an earnest effort to enhance the protection of sensitive data – especially among defense contractors. The Office of the Under Secretary of Defense for Acquisition and Sustainment (OUSD(A&S)) worked with Johns Hopkins University Applied Physics Laboratory and Carnegie Mellon University Software Engineering Institute to create a new cybersecurity certification standard for DoD contractors. The goal of the new standard, the Cybersecurity Maturity Model Certification (CMMC), is to provide cybersecurity guidance to the Defense Industrial Base (DIB) and hold them accountable for protecting Federal Contract Information (FCI) and Controlled Unclassified Information (CUI) within the supply chain.

Any vulnerabilities introduced to the supply chain ecosystem by the least cybersecurity-capable company very likely weakens the cybersecurity posture of the entire supply chain. Given the interdependencies between the customer (DoD) systems, prime contractor, and sub-contractor, a breach of one can affect all. Use of a maturity model with built-in accountability is a way to reduce the inherent vulnerabilities stemming from the use of interdependent systems.

In this study, we investigate how ready the DIB is for the CMMC process by conducting a set of interviews with a group of small and large DoD contractors. We discuss the cybersecurity protocols and or standards currently in place in those companies, the current state of their cybersecurity posture, the CMMC level each company feels they need to achieve, concerns about achieving certification, and explore the differences reflected by the size of the company.

2. BACKGROUND

How to prudently handle non-classified information is something that the U.S. Government has wrestled with for quite some time. What follows is a brief history to set the stage and provide context from which the CMMC has emerged. President Carter's 1977 Presidential Directive to manage the security of unclassified telecommunications information transmitted among U.S. Government agencies and contractors, was arguably the first high-level U.S. policy dealing with unclassified information (Brzezinski, 1977). In 1984, this information was referred to as sensitive but unclassified (SBU) (National Security Decision Directive [NSDD], 1984) and later, was specifically defined as "information the disclosure, loss, misuse, alteration, or destruction of which could adversely affect national security or other Federal Government interests" (National Telecommunications and Information Systems Security Policy [NTISSP], 1986, p. 166). For the next 20+ years, the definition, handling, and sharing of SBU was problematic as was the proliferation of agency-specific labels for similar type information such as For Official Use Only (FOUO), Law Enforcement Sensitive (LES), etc.

Controlled Unclassified Information

In 2008, President G.W. Bush, in an effort to standardize government information handling practices and improve information sharing, issued a memorandum establishing a framework for managing CUI and defined it as:

> the single, categorical designation henceforth throughout the executive branch for all information within the scope of that definition, which includes most information heretofore referred to as Sensitive But Unclassified (SBU) in the Information Sharing Environment (ISE), and establishes a corresponding new CUI Framework for designating, marking, safeguarding, and disseminating information designated as CUI. (Bush, 2008)

Maintaining focus and momentum on this issue, President Obama issued a memorandum four months after inauguration that set up a task

force to review government procedures used to categorize and share SBU information as well as to consider measures for tracking government agencies' progress implementing the CUI framework (Obama, 2009). The task force report provided 40 recommendations, key elements of which were included 15 months later in Executive Order 13556 which also broadened the scope of CUI to include all SBU information within the Executive Branch (Holder & Napolitano, 2009; Exec. Order No. 13556, 2010).

After nearly four years of work to codify the CUI program, the Information Security Oversight Office, an organizational component of the National Archives Record Administration (NARA) which is the Federal Government's Executive Agent for CUI, issued a rule to establish policy for executive branch agencies on "designating, safeguarding, disseminating, marking, decontrolling, and disposing of CUI" as well as other aspects of the CUI program (Federal Register, 2016b). The guidance entered the Code of Federal Regulations (Electronic Code of Federal Regulations [e-CFR], 2021; National Archives, 2020) creating the CUI registry (125 categories of CUI currently) and formally defining CUI as:

> information the Government creates or possesses, or that an entity creates or possesses for or on behalf of the Government, that a law, regulation, or Government-wide policy requires or permits an agency to handle using safeguarding or dissemination controls. (e-CFR, Title 32, Vol. 6, Part 2002.4(h), 2021)

Contractor Protection of CUI

Around the same time, the DoD published a final rule on the Defense Federal Acquisition Regulation Supplement (DFARS) clause requiring that contractors implement the security requirements in NIST SP 800-171 no later than December 31, 2017 (Federal Register, 2016a). Two key problems with this guidance were that (1) DoD had no process for certifying compliance (contractors could simply self-attest to their compliance) and (2) contractors were allowed to continue providing goods and services even if they were not fully compliant with 800-171 so long as any gaps were documented in a Plan of Action and Milestones (POAM) (National Institute of Standards and Technology [NIST], 2018).

Because of problems with implementation of DFARS 252.204-7012 (DFAR, 2019), the OUSD(A&S) issued a memorandum in January 2019 that directed the Defense Contract Management Agency (DCMA) to "validate compliance with the requirements of DFARS clause 252.204-7012" for certain contractors (Lord, 2019). As a direct result, DCMA stood up the Defense Industrial Base Cybersecurity Assessment Center (DIBCAC) in June 2019 to begin conducting assessments of some of the DoD's largest contractors (Tremblay, 2019).

Birth of the CMMC

The OUSD(A&S) announced in May 2019 the initiative to create the CMMC framework (Doubleday, 2019). Figure 1 depicts the key events in the CMMC development and implementation timeline.

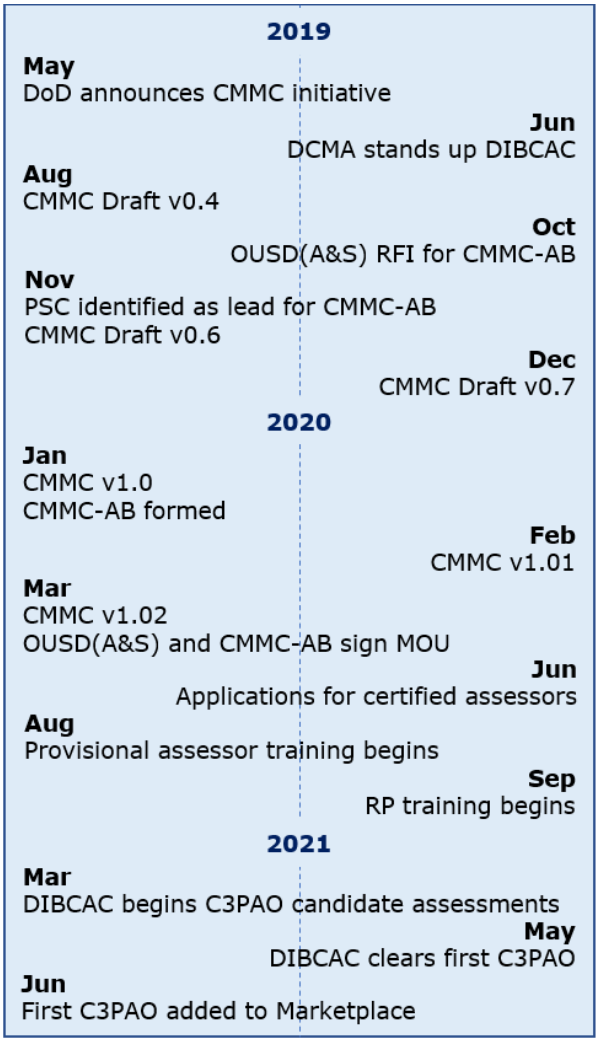


Figure 1 – CMMC development key event timeline

As some in the DoD iterated through draft versions of the CMMC, others worked to create the organizational structure required to

implement it. In early October 2019, the OUSD(A&S) published a request for information (RFI) on "how to define the long-term implementation, functioning, sustainment, and growth of the CMMC Accreditation Body" (RFI HQ0034SS10032019, 2019). In November 2019, an Accreditation Body kickoff meeting was held out of which the Professional Services Council (PSC, 2021) emerged as the lead to create a volunteer board to establish a nonprofit to act as the accreditation body for the CMMC process (Barnett, 2020). The PSC, founded in 1972, is the 400+ member-company national trade association of the government technology and professional services industry.

The CMMC Accreditation Body (CMMC-AB) formed as a non-profit organization in January 2020 with a 15-person volunteer board and signed a formal Memorandum of Understanding (MOU) with the OUSD(A&S) in March 2020 (Lord & Schieber 2020). The CMMC-AB manages and oversees all certification, training, and accreditation aspects of the CMMC including training of Registered Practitioners (RPs); marketplace listing of Registered Provider Organizations (RPOs); accreditation of CMMC Third Party Assessment Organizations (C3PAOs); and, most importantly, contractor CMMC certification.

Key to getting 300,000+ defense contracting companies through the certification process over the next several years are the C3PAOs. Each C3PAO must be Level 3 certified (CMMC Accreditation Body [CMMC-AB], 2021) by DIBCAC and meet various administrative and personnel requirements from the CMMC-AB before they can begin conducting contractor assessments. DIBCAC assessments of C3PAOs, which began in March 2021 (Goepel, 2021), take approximately 6 weeks, including scheduling and pre-assessment reviews, virtual and on-site assessments, and post-assessment analysis. The CMMC-AB Marketplace reflected in early June 2021 that there were 156 C3PAO candidates pending Level 3 assessment and a single company, Redspin, officially designated as a certified assessment organization.

3. CMMC Details

The CMMC is a framework designed to provide the DoD with verification that DIB members can adequately protect FCI and CUI flowing through the supply chain from customer to prime contractors to sub-contractors. It builds upon existing regulations, other models' best practices, and combines multiple existing cybersecurity standards both from within the U.S. government and internationally (DoD, 2019).

CMMC Components

Based on early work conducted by the Software Engineering Institute to improve software processes (Paulk, et. al, 1993), the framework uses five levels to designate an organization's cybersecurity maturity. Each of these levels is defined by the *processes* an organization has established and is following, as well as the *practices* that are implemented. This relationship between processes and practices across the five maturity levels of the CMMC is reflected in figure 2. Processes range from Performed, at level 1, to Optimizing, at level 5. With CMMC required practices in place, level 1 is considered Basic Cyber Hygiene, while level 5 is Advanced/Progressive. An organization certified at any level of the CMMC is meeting the processes/practices of that level as well as those below it.

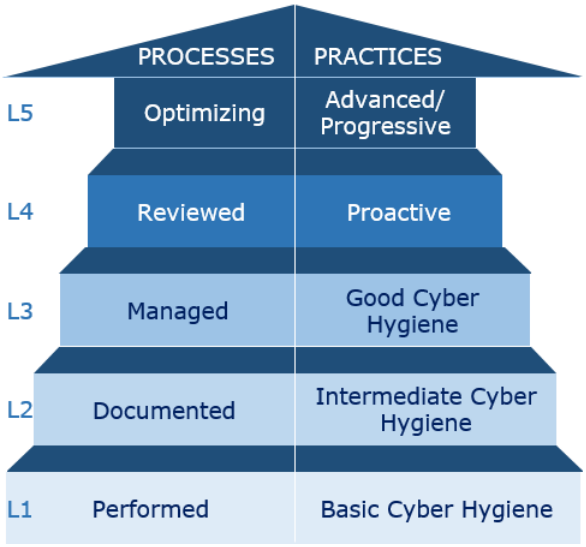


Figure 2 – CMMC processes and practices at each maturity level

General descriptions of the five levels are:

- Level 1: Protecting FCI is the focus and is achieved by meeting the basic requirements of 48 CFR 52.204-21.
- Level 2: This is a transitional stage for organizations working towards Level 3. The focus is on replacing ad-hoc processes/practices with well-documented processes and corresponding regular practices.
- Level 3: Protecting CUI is the focus and is achieved with well-established processes accompanied by implementation of all

regular practices outlined in NIST SP 800-171, plus 20 additional practices.
- Level 4: This level could be viewed as a transitional stage for organizations working towards Level 5. Reviewing and measuring existing practices to gauge effectiveness and enhancing security to protect CUI from Advanced Persistent Threats (APTs) is the focus.
- Level 5: At this highest level, organizations would be continually optimizing existing processes and practices. Being capable of defending CUI from APTs would include, noticing missing logs, verifying the integrity of security critical software, responding in real-time to anomalous network activities, recording network traffic crossing organizational boundaries, etc.

The number of practices that must be met and verified at each level are depicted in figure 3. Note that each level requires all practices from previous levels. For example, Level 1, Basic Cyber Hygiene, requires 17 practices be met, while Level 2, Intermediate Cyber Hygiene, requires 72 practices be met, 17 from Level 1 plus 55 from Level 2 (17 + 55 = 72).

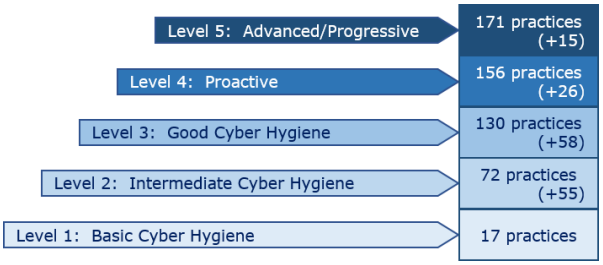


Figure 3 – number of practices required at each CMMC maturity level

As the goal of CMMC is to change the supply chain culture, every DIB member will need to be at least Level 1 certified. As emphasized by the OUSD(A&S) CISO:

> Level 1 reflects the basic cyber hygiene skills that we should be using every day, regardless. I've been asked, "Ma'am, I do landscaping for the government. Should I have CMMC certification?" And my answer has actually been, "Yes, I want you to at least get to Level 1". (Anderson, 2020).

The CMMC framework organizes practices within 17 domains, which includes the 14 domains enumerated in NIST 800-171 as well as 3 additional domains: Asset management (AM), Recovery (RE), and Security Assessment (CA). These domains are listed in table 1 where we present a crosswalk of the number of required practices across domains and levels.

The 17 practices (Appendix B) required for Level 1 certification come from just 6 of the 17 domains while at Level 3, organizations must meet practice requirements across all 17 domains. The 17 domains are: Access Control (AC), Asset Management (AM), Audit and Accountability (AU), Awareness and Training (AT), Configuration Management (CM), Identification and Authentication (IA), Incident Response (IR), Maintenance (MA), Media Protection (MP), Personnel Security (PS), Physical Protection (PE), Recovery (RE), Risk Management (RM), Security Assessment (CA), Situational Awareness (SA), System and Communications Protection (SC), System and Information Integrity (SI).

Cybersecurity Practice Crosswalk by Domain and Level						
DOMAIN	L1	L2	L3	L4	L5	Domain Totals
AC	4	10	8	3	1	**26**
AM			1	1		**2**
AU		4	7	2	1	**14**
AT		2	1	2		**5**
CM		6	3	1	1	**11**
IA	2	5	4			**11**
IR		5	2	2	4	**13**
MA		4	2			**6**
MP	1	3	4			**8**
PS		2				**2**
PE	4	1	1			**6**
RE		2	1		1	**4**
RM		3	3	4	2	**12**
CA		3	2	3		**8**
SA			1	2		**3**
SC	2	2	15	5	3	**27**
SI	4	3	3	1	2	**13**
Totals	17	55	58	26	15	**171**

Table 1 – domain crosswalk for the number of required practices at each level.

CMMC Phased Implementation

The DFARS Clause 252.204-7021 states that OUSD(A&S) must approve the use of the clause for new acquisition until October 2025 after which CMMC is expected to be fully implemented and required of all new contracts. Table 2 illustrates the roll-out plan over the next five fiscal years for the number of contracts that will contain a CMMC requirement.

Number of Contracts with CMMC Requirement				
FY21	FY22	FY23	FY24	FY25
15	75	250	479	479

Table 2 – CMMC roll-out by # of contracts

Table 3 shows the initial CMMC roll-out numbers of prime contractors and sub-contractors across that same time horizon.

	Number of Prime/Sub-Contractors with CMMC Requirement				
	FY21	FY22	FY23	FY24	FY25
L1	895	4,490	14,981	28,714	28,709
L2	149	748	2,497	4,786	4,785
L3	448	2,245	7,490	14,357	14,355
L4	4	8	16	24	28
L5	4	6	16	24	28
Tot	1,500	7,500	25,000	47,905	47,905

Table 3 – CMMC roll-out by # of contractors

Costs associated with acquiring and maintaining certification will vary by the level of the certification and the size of the organization. The availability of resources among DIB is a concern that we noted during our interviews with the pilot group of DoD contractors. We also noted that there was a consensus among most of the pilot group regarding the importance of CMMC and the security it will add to the CUI and FCI data.

4. METHODOLOGY

To investigate DIB understanding of, readiness for, and opinion of the CMMC process, we conducted interviews with 10 defense contractors: six small and four large businesses. Company size was established using Small Business Administration (SBA) standards related to number of employees and/or average annual receipts according to their North American Industry Classification System (NAICS) code (NAICS Association, 2019). It should be noted that there is no medium-size category in SBA classification of companies. All interviews were transcribed in their entirety and kept anonymous.

The 10 companies we interviewed were a convenience or opportunity sample. While the sample is nonrandom, we tried to include a mix of industries and blend of large and small companies to provide a reasonable approximation of the larger contractor population. The interviewees were mid-level managers of information technology departments or decision makers of small companies that outsource information technology needs. The open-ended survey questions were designed to collect information on the nature of the firms, their readiness for CMMC assessment, and their concerns. Appendix A lists the survey questions used in the interviews.

5. SURVEY RESULTS AND DISCUSSION

We note generally that all four large businesses in our pilot study have conducted several discussions regarding CMMC and have formed teams that include information security specialists assigned specifically to CMMC adoption. Small businesses, on the other hand, had widely ranging responses from "we are starting to analyze the current state" to "almost compliant with our desired CMMC level." It was apparent from the responses that the small businesses that do not primarily provide IT consulting services were struggling most with CMMC.

In the following subsections, we discuss the responses to key questions (3, 4, 5, 6, 7, 9, & 10) from among the 10 asked.

Other Cybersecurity Framework Adoption

Responses to question 3, *Has your company adopted a cybersecurity framework or standard, if so, which one?*, indicate familiarity with cyber-related standards generally and some existing standards specifically. Given the DFARS clause deadline of December 31, 2017 that currently applies to all DoD contractors, this is not surprising and probably should have been a reason for greater awareness. Frameworks mentioned include: NIST Risk Management Framework (RMF), ISO 27001, NIST 800-171, Capability Maturity Model Integration (CMMI) Level 2, Payment Card Industry Data Security Standard (PCI DSS), Federal Risk and Authorization Management Program (FedRAMP), Health Insurance Portability and Accountability Act (HIPAA) provisions, Health Information Technology for Economic and Clinical Health (HITECH) requirements, and CMMC.

Providing these standards in response to a question about "cybersecurity frameworks" may cause some concern/questions by readers, but it was insightful for researchers both to see familiarity with implementing governance requirements and for how some companies seemingly lumped many requirements into a single, large mental bin. Three companies have not formally adopted any cybersecurity framework, though they are aware of the importance of cybersecurity practices generally and are following them in an ad-hoc manner. These findings seem to validate concerns of compliance with self-attestation.

CMMC Level Targeted

Question 4 asked: *Which CMMC Level (1-5) does your organization need per current/anticipated DoD contracts? What CMMC Level is your Prime requiring [if applicable] of your company?* Three of the four large companies indicated that they believe they currently meet level 3 certification requirements, while one was unsure of their status and likely not yet at level 1.

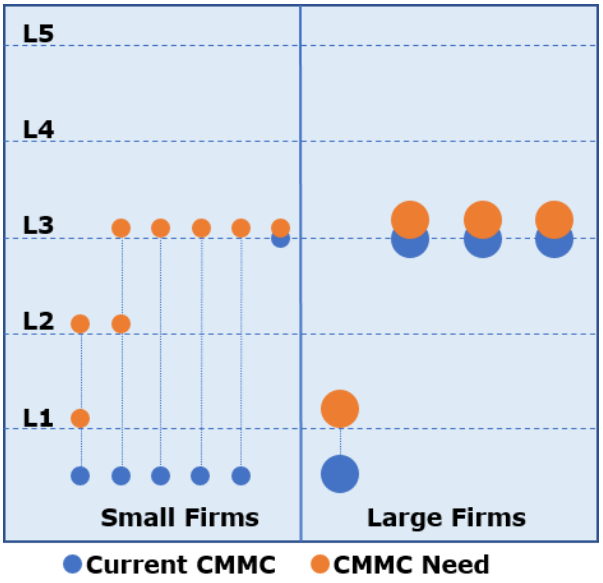


Figure 4 – pilot study group of 10 companies self-attested current CMMC level and future required level

One small firm indicated they were already meeting level 3 requirements, while none of the others professed to be currently meeting any level. All small companies seemed to understand the importance of achieving CMMC and were considering how to get to the level they felt they needed. Figure 4 shows the current CMMC readiness level attested by the pilot study group as well as the future anticipated level.

Note the clear differences in the current readiness posture between small and large businesses. It seems apparent that small businesses will struggle more with the CMMC mandate, specifically the ones performing in industries outside the cyber domain, while most large businesses appear positioned to rapidly meet the new requirements.

When asked about plans to continue to level-up on CMMC, two of the large firms stated they will likely push to levels 4 and 5 even if not required by contract. Two of the small firms expressed a view toward taking the CMMC levels in steps – get certified at level 1, then work on level 2, etc.

Short-term CMMC Achievement

In response to question 5, *what CMMC Level can your organization achieve in the short term (within 12-18 months)?*, all of the large contractors and cyber-focused small contractors interviewed had a good understanding of the requirements, budgeted CMMC readiness and certification costs, and had a plan to achieve the required certification level within the next 12-18 months. It was common among the small non-cyber companies that they had a more loosely sketched plan to get their systems ready. This could be linked to the lack of understanding by the leadership in those businesses of the cyber systems used in their companies and what it takes to meet the required CMMC levels. All companies interviewed understood the need to achieve CMMC in order to continue doing business with DoD.

Cost Concerns

Question 6 got into the question of CMMC cost: *Has your organization budgeted for CMMC costs? If yes, approximately how much, if no, why not? Please, choose a range: $0-$25k, $26k-$50k, $51k-$75k, $76k and above.* All companies had concerns regarding the CMMC costs. Note that the NAICS code provides a lower bound on the number of employees required to be classified as a large organization. Thus, the number of employees in a large firm can vary significantly, e.g., two large firms having employee numbers differing by tenfold would not be surprising. For that reason, it is hard to enumerate the anticipated cost per employee with the sample size. Without exception, all companies were concerned about the resources and the cost of this new mandate. They were concerned not only about the initial cost to bring their processes and practices up to certification

standards, but also the ongoing cost of maintaining the certifications. The CMMC financial outlay suggested by the large businesses varied from $600K to $3M for the initial certification, while the small businesses' estimates ranged from $1K to $50K. While most companies expressed their willingness to do whatever it takes to keep working with DoD, they also expressed that the cost of tools/licenses that provide functions to maintain the certification would be an internal challenge as it will exceed what they normally spend for cybersecurity.

Little Concern about Inability to Adapt

For question 7, *is your organization concerned that it will not be able to adapt to CMMC required changes? What are your concerns (e.g., leveling up, losing contract)?*, the overriding theme of the response was that companies expressed willingness to do whatever it takes to keep working with DoD. An interesting concern voiced by some larger companies was the possibility of being held responsible for getting/keeping sub-contractors certified. In previous work by some of the authors (Vanajakumari, Mittal, Stoker, Clark, & Miller, 2021), this idea was proposed and, according to some of the interviewees' comments, it may be gaining traction in the DoD. The concern is understandable, especially if a small company is the prime contractor and a large company is a sub. However, generally, we continue to believe that in the highly interconnected cyber environment of today, the lead contractor (typically the more powerful member) must take special initiative and leadership to ensure the highest level of cybersecurity attainment.

Will CMMC Help?

Question 9 asked, *do you think CMMC will help your organization, or the supply chain of which you are a part, mitigate cybersecurity risks?* While all the companies in the pilot-group expressed a degree of cybersecurity concern and agreed that there is a need to secure supply chain data, three were not sure about CMMC helping. The responses from the three that had low enthusiasm for CMMC ranged from probably not to possibly. The lack of excitement among this subset mostly stemmed from confidence in their own current cybersecurity posture, which caused them to see CMMC as yet another top-down driven requirement that added little value. 80% of the companies interviewed believed that CMMC would certainly help in ensuring accountability when it comes to supply chain cybersecurity.

Final Interviewees' Thoughts

Outside the context of the 10 questions, the interviewed DIB members generally agreed that the third-party assessment will help with keeping businesses honest and thus complying with the cybersecurity requirements at the certified level. However, there were doubts expressed regarding the extent to which complying with CMMC practices would help avoid and/or contain cybersecurity events. Some of the concern stems from the fact that CMMC compliance is only checked once every three years and thus the reliability of compliance in between certification periods might be questionable.

Emergence of CMMC 2.0

After the completion of this study, a new version, CMMC 2.0, emerged from an internal review of CMMC's implementation which included more than 850 public comments (Office of the Under Secretary of Defense, Acquisition & Sustainment, 2021). The implementation of CMMC 2.0 was undertaken to build on and refine the original requirements from CMMC. Key changes include a more streamlined model, reliable assessments, and flexible implementation (DoD, 2021).

The CMMC Model 2.0 was streamlined from 5 levels to 3 (Appendix C). These three levels are labeled Foundational (Level 1), Advanced (Level 2), and Expert (Level 3). Level 1 remained the same as CMMC 1.0 with 17 controls/practices that enable organizations to handle only Federal Contract Information (FCI). Level 2 is now aligned with the 110 controls in NIST 800-171 and is a combination of Levels 2 and 3 from the CMMC 1.0 model. The highest level of CMMC 2.0, Level 3, is aligned with NIST 800-172 and combines the previous Levels 4 and 5.

Additionally, there are reduced assessment requirements and flexibility around implementation. Under CMMC 2.0, organizations can now perform annual self-assessments for Level 1 as well as a subset of Level 2 (non-prioritized). The companies that fall under prioritized Level 2 group need to get C3PAO assessments every three years. Level 3 organizations will have government-led assessments every three years. There is also more flexibility under certain circumstances that would allow for waivers to the CMMC requirements (DoD, 2021).

As for companies who currently have CMMC 1.0 compliance, companies will maintain that compliance until the DoD finalizes the rule-

making process which will take 9–24 months to complete. Once this process is complete, CMMC 2.0 will become a contract requirement for most DoD contracts.

6. CONCLUSIONS AND FUTURE WORK

To investigate the CMMC readiness of DoD contractors and sub-contractors, we conducted a pilot survey of 10 large and small government contractors. Our findings show that all DIB members are aware of the compliance requirements; however, their state of readiness and understanding of the certification requirements vary markedly depending on their size and the nature of their business. In light of some items revealed by our pilot study, like the respondents' concerns with contractors maintaining a proper cybersecurity posture during the three years between required certifications, we plan to conduct a follow-on CMMC study. That investigation will include more companies and delves more deeply into some of the questions raised from this pilot study regarding the differences in preparedness between small and large contractors. Additionally, the respondents' concerns may have changed with the introduction of CMMC 2.0 (e.g., the move to a yearly self-assessment may alleviate some concerns).

On May 12, 2021, Presidential Executive Order 14028 was signed outlining the desired path to improve the nation's cybersecurity posture and protect federal government networks (Exec. Order No. 14028, 2021). Recent high-profile attacks (SolarWinds, Microsoft Exchange, the Colonial Pipeline, and JBS) reveal how vulnerable federal and private sectors are to cyberattacks from other nations and cyber criminals. Executive Order 14028 specifically requires implementation of some items, such as multi-factor authentication (MFA), that are currently part of CMMC Level 3, as a base requirement. At the time of this research, this seemed to signal that there might soon be some modifications to the list of CMMC controls at each level and that the CMMC framework might soon become more generally applied to other parts of the federal government. With the introduction of CMMC 2.0, we are seeing the DoD move toward standardization across federal organizations with the adoption of NIST standards directly into the CMMC.

There is also an increasing number of ransomware attacks cutting across all sectors. On June 3, 2021, White House released a memo asking business leaders to step up their cybersecurity measures. Though currently CMMC compliance is a requirement for DIB members only, considering recent events it is likely to become a standard for all U.S. businesses. Many attacks can be prevented by companies adopting CMMC 1.0 Level 1 (basic cyber hygiene) which is also the same in CMMC 2.0. Our findings provide insights to companies on the challenges associated with improving their cybersecurity stance.

Limitations of the Study

As mentioned previously, the intent of this study was to serve as a pilot to inform future studies on CMMC. However, this may be considered a limitation of the current results due to the size of the study (10 companies). Additionally, based on feedback, some questions may need to be modified in future studies to elicit clearer responses. For example, questions surrounding budgeting for CMCC costs were asked without placing time constraints (e.g., have you budgeted for CMMC adoption in the next 12 to 18 months). Furthermore, since the submission of this research, CMMC 2.0 has emerged which will change the approach of future research as these new changes are approached and implemented by various organizations. These limitations will be addressed in future studies.

7. REFERENCES

Anderson, M. (2020, October 6). SME. *Your Best Cyber Defense Isn't a '60's Super Spy. It's You*. https://www.sme.org/technologies/articles/2020/october/your-best-cyber-defense-isnt-a-60s-super-spy.-its-you/

Barnett, J. (2020, June 23). FedScoop. The DOD wants better cybersecurity for its contractors. The first steps haven't been easy. https://www.fedscoop.com/cmmc-dod-cyber security-requirments-contractors-timeline/

Brzezinski, Z. (1977, November 16). Presidential Directive/NSC-24. Telecommunications Protection Policy. https://fas.org/irp/offdocs/pd/pd24.pdf

Bush, G. W. (2008, May 7). Memorandum for the Heads of Executive Departments and Agencies. *Designation and Sharing of Controlled Unclassified Information (CUI)*. https://fas.org/sgp/bush/cui.html

Council of Economic Advisers. (2018, February). *The Cost of Malicious Cyber Activity to the U.S. Economy*. https://www.hsdl.org/?abstract&did=808776

CMMC Accreditation Body. (2021, June). Cybersecurity Maturity Model Certification. https://cmmcab.org/

Defense Federal Acquisition Regulation. (2019). *Safeguarding Covered Defense Information and Cyber Incident Reporting* (Supplement 252.204-712). https://www.acq.osd.mil/dpap/dars/dfars/html/current/252204.htm#252.204-7012

Department of Defense. (2019, November 7). Cybersecurity Maturity Model Certification (CMMC) Draft V0.6. https://www.acq.osd.mil/cmmc/docs/CMMC-V0.6b-20191107.pdf

Department of Defense. (2021, November 4). Strategic Direction for Cybersecurity Maturity Model Certification (CMMC) Program [Press Release]. https://www.defense.gov/News/Releases/Release/Article/2833006/strategic-direction-for-cybersecurity-maturity-model-certification-cmmc-program/

Doubleday, J. (2019, June 3). *Defense Dept. to require new cybersecurity certification from contractors*. Inside Cybersecurity. https://www.the-center.org/getattachment/Our-Services/Cybersecurity-Services/Cybersecurity/Defense-Dept-to-require-new-cybersecurity-certification-from-contractor-2.pdf.aspx?lang=en-US

Electronic Code of Federal Regulations. (2021, May 27). *Controlled Unclassified Information* (Title 32, Vol. 6, Part 2002.4(h)). https://www.ecfr.gov/cgi-bin/text-idx?SID=54ce48937eb0b451c823363c49411eb2&mc=true&node=pt32.6.2002&rgn=div5#se32.6.2002_14

Exec. Order No. 13556, 3 C.F.R. 267 (2010). https://www.govinfo.gov/content/pkg/CFR-2011-title3-vol1/pdf/CFR-2011-title3-vol1-eo13556.pdf

Exec. Order No. 14028, 86 Fed. Reg. 26633 (May 12, 2021). https://www.federalregister.gov/documents/2021/05/17/2021-10460/improving-the-nations-cybersecurity

Federal Register, 81 FRM 72986. (2016a, October 21). *Defense Federal Acquisition Regulation Supplement: Network Penetration Reporting and Contracting for Cloud Services*. https://www.govinfo.gov/content/pkg/FR-2016-10-21/pdf/2016-25315.pdf

Federal Register, 81 FR 63323. (2016b, November 16). *Controlled Unclassified Information*. https://www.federalregister.gov/documents/2016/09/14/2016-21665/controlled-unclassified-information

Goepel, J. (2021, April 28). *DIBCAC Releases C3PAO CMMC Maturity Level 3 Lessons Learned*. CMMC Information Institute. https://cmmcinfo.org/2021/04/28/dibcac-releases-c3pao-cmmc-maturity-level-3-lessons-learned/

Holder, E. & Napolitano, J. (2009, August 25). Report and Recommendations of the Presidential Task Force on Controlled Unclassified Information. https://www.archives.gov/files/cui/documents/2009-presidential-task-force-report-and-recommendations.pdf

Lord, E. (2019, January 21). Under Secretary of Defense for Acquisition and Sustainment Memorandum. *Addressing Oversight as Part of a Contractor's Purchasing System Review*. https://www.acq.osd.mil/dpap/pdi/cyber/docs/USA000140-19%20TAB%20A%20USD(AS)%20Signed%20Memo.pdf

Lord, E. & Schieber, T. A. (2020, March). Memorandum of Understanding (MOU) between the Department of Defense, Office of the Undersecretary for Acquisition and Sustainment (OUSD(A&S)) and Cybersecurity Maturity Model Certification Accreditation Body, Inc. (CMMC-AB). https://assets.documentcloud.org/documents/6935675/CMO001673-20-CMMC-AB-MOU-Fully-Executed-20200323.pdf

National Archives. (2020, April 13). *CUI Categories*. https://www.archives.gov/cui/registry/category-list

National Institute of Standards and Technology. (2018, December). *Risk Management Framework for Information Systems and Organizations*. https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-37r2.pdf

National Security Decision Directive Number 145 (1984, September 17). *National Policy on Telecommunications and Automated Information Systems Security*. https://fas.org/irp/offdocs/nsdd145.htm

National Telecommunications and Information Systems Security Policy. (1986, October 29). *National Policy on Protection of Sensitive, but Unclassified Information in Federal Government Telecommunications and Automated Information Systems* (No. 2).

https://www.princeton.edu/~ota/disk2/1987/8706/870611.PDF

North American Industry Classification System Association. (2019, August 19). SBA Table of Small Business Size Standards. https://www.naics.com/sba-size-standards/

Obama, B.H. (2009, May 27). Presidential Memorandum. *Classified Information and Controlled Unclassified Information*. https://obamawhitehouse.archives.gov/the-press-office/presidential-memorandum-classified-information-and-controlled-unclassified-informat

Office of the Under Secretary of Defense, Acquisition & Sustainment. (2021, December 3). *About CMMC*. https://www.acq.osd.mil/cmmc/about-us.html

Paulk, Mark., Curtis, William., Chrissis, Mary Beth., & Weber, Charles. (1993). *Capability Maturity Model for Software (Version 1.1)* (CMU/SEI-93-TR-024).

Pozen, D. E. (2005). The Mosaic Theory, National Security, and the Freedom of Information Act. Yale LJ, 115, 628. https://www.yalelawjournal.org/pdf/358_fto38tb4.pdf

Professional Services Council. (2021). https://www.pscouncil.org/

Request for Information (RFI) HQ0034SS10032019. (2019, October 3). RFI Cybersecurity Maturity Model Certification Accreditation Body. https://sam.gov/opp/4a4b539a0e347e540b30b3121916031c/view

Tremblay, P. (2019, June 24). Defense Contract Management Agency (DCMA) website article. *Building a cybersecurity assessment capability*. https://www.dcma.mil/News/Article-View/Article/1885182/building-a-cybersecurity-assessment-capability/

Vanajakumari, M., Mittal, S., Stoker, G., Clark, U., & Miller, K. (2021). Towards a Leader-Driven Supply Chain Cybersecurity Framework. JISAR, 14(2), 42. http://jisar.org/2021-14/n2/JISARv14n2p42.pdf

Editor's Note:

This paper was selected for inclusion in the journal as an CONISAR 2021 Distinguished Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2021.

Appendix A – Survey Questions

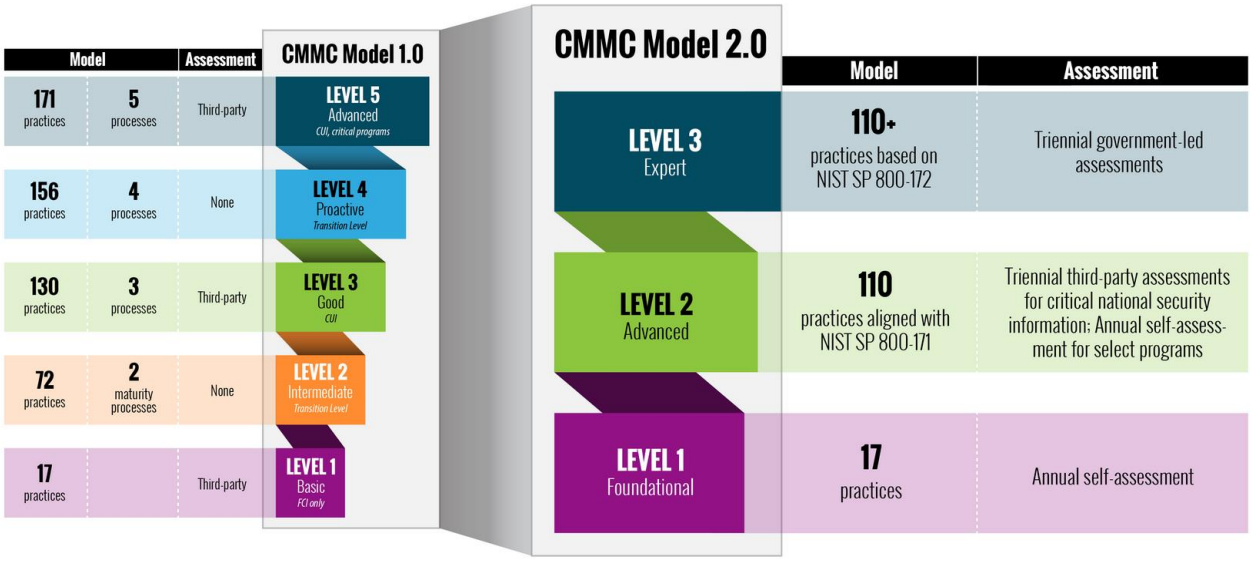
1. Which industry does your organization primarily support?
2. What is the size of your organization (small/large)? What is your main NAICS code?
3. Has your company adopted a cybersecurity framework or standard, if so, which one?
4. Which CMMC Level (1-5) does your organization need per current/anticipated DoD contracts? What CMMC Level is your Prime requiring [if applicable] of your company?
5. What CMMC Level can your organization achieve in the short term (within 12-18 months)?
6. Has your organization budgeted for CMMC costs? If yes, approximately how much, if no, why not? Please, choose a range: $0-$25k, $26k-$50k, $51k-$75k, $76k and above.
7. Is your organization concerned that it will not be able to adapt to CMMC required changes? What are your concerns (e.g., leveling up, losing contract)?
8. Do you have major homegrown software systems?
9. Do you think CMMC will help your organization, or the supply chain of which you are a part, mitigate cybersecurity risks?
10. Given your experience, what do you think are the major obstacles your organization will have in adopting CMMC?

Appendix B

CMMC Level 1 – Basic Cyber Hygiene

AC.1.001: Limit information system access to authorized users, processes acting on behalf of authorized users, or devices (including other information systems).
AC.1.002: Limit information system access to the types of transactions and functions that authorized users are permitted to execute.
AC.1.003: Verify and control/limit connections to and use of external information systems.
AC.1.004: Control information posted or processed on publicly accessible information systems.
IA.1.076: Identify information system users, processes acting on behalf of users, or devices.
IA.1.077: Authenticate (or verify) the identities of those users, processes, or devices, as a prerequisite to allowing access to organizational information systems.
MP.1.118: Sanitize or destroy information system media containing FCI before disposal or release for reuse.
PE.1.131: Limit physical access to organizational information systems, equipment, and the respective operating environments to authorized individuals.
PE.1.132: Escort visitors and monitor visitor activity.
PE.1.133: Maintain audit logs of physical access.
PE.1.134: Control and manage physical access devices.
SC.1.175: Monitor, control, and protect organizational communications (i.e., information transmitted or received by organizational information systems) at the external boundaries and key internal boundaries of the information systems
SC.1.176: Implement subnetworks for publicly accessible system components that are physically or logically separated from internal networks.
SI.1.210: Identify, report, and correct information and information system flaws in a timely manner.
SI.1.211: Provide protection from malicious code at appropriate locations within organizational information systems.
SI.1.212: Update malicious code protection mechanisms when new releases are available.
SI.1.213: Perform periodic scans of the information system and real-time scans of files from external sources as files are downloaded, opened, and executed.

Appendix C
CMMC 2.0 Model



*Retrieved from https://www.acq.osd.mil/cmmc/about-us.html on November 11, 2021

The COVID-19 Pandemic's Impact on Information Technology Employment, Salaries, and Career Opportunities

Patricia Sendall
sendallp@merrimack.edu
Girard School of Business
Merrimack College
N. Andover, MA 01845 USA

Alan Peslak
arp14@psu.edu
Department of Information Sciences and Technology
Penn State University
Dunmore, PA 18512 USA

Wendy Ceccucci
wendy.ceccucci@quinnipiac.edu
Department of Computer Information Systems
Quinnipiac University
Hamden, CT 06518 USA

D. Scott Hunsinger
hunsingerds@appstate.edu
Department of Computer Information Systems
Appalachian State University
Boone, NC 28608 USA

Abstract

This is a an empirical study of the effect of the COVID-19 pandemic on salary and employment trends in Information technology (IT) jobs over the period January 2019 to April 2021. The study is an effort to determine the impact of COVID on IT jobs and salary. Data was extracted from Burning Glass Labor Insight which includes over 40 million US job postings per year. We downloaded monthly data for the time period Jan 2020 to April 2021. These data included all job postings as well as job postings in science, engineering, information technology. They were analyzed using SPSS 26 and Microsoft Excel. We attempted to determine through correlation the degree of similarity between IT jobs and other technical and non-technical work. We gain key insights into IT jobs during the pandemic compared to other STEM jobs as well as variances among IT positions.

Keywords: Information Technology, COVID, COVID-19, IT jobs, pandemic

1. INTRODUCTION

The COVID-19 pandemic took a toll on workers across the globe. Social distancing and mask wearing became the norm. People were separated not only from their workplaces, but also from their loved ones. There was disruption across all industries with business closures and work-from-home (WFH) mandates. Women and under-represented populations took the hardest hit with increased domestic responsibilities for children and elders. The Information Technology sector was not immune to the disruption. While there was a need for new technologies, for example Zoom, to support learning and working from home, the tech industry also took a hit when it came to total job postings. This paper analyzes over 40 million job postings on Burning Glass Labor Insights during the period January 2019 to April 2021. The authors conclude that there will continue to be a need for tech workers especially in cybersecurity, software development and artificial intelligence. Tech workers will have to continue to learn new skills in order to keep up with the demands of the post-COVID economy.

2. LITERATURE REVIEW

The COVID-19 pandemic has brought about the greatest economic disruption since the Great Depression and job postings by American companies have been dramatically altered by the pandemic (Campello, Kankanhalli, Muthukrishnan, 2020). The authors analyzed data from LinkUp, a leading labor market research firm. Figure 1. shows an irregular drop in job postings at the beginning of the pandemic in March 2020. This drop also coincides with an extraordinary spike in initial jobless claims.

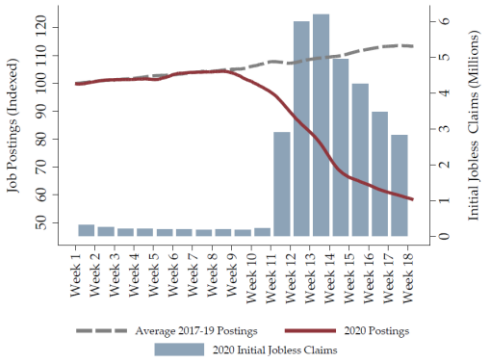


Figure 1. Job postings vs. unemployment claims

The McKinsey Global Institute discusses the future of work after COVID-19. They cite work trends that have increased by the pandemic (Lund, et al, 2021, pg. vii). They are:

1. Remote Work--20-25% of workers in advanced economies could work remotely 3+ days a week on a long-term basis
2. Digitization--2-5x growth in e-commerce, as a surge in digital platforms is underway
3. Automation—An uptick in use of robotics, robotic process automation and AI

Remote work has been supported by new digital solutions, such as "videoconferencing, document sharing tools, and expansion of cloud-based computing capacity" (Lund, et al, pg. 5).

Internet of Things (IoT) technologies provided "mechanical and digital technologies to transfer the data through the Internet without any human interaction." (Javaid & Khan 2021, pg. 209). IoT enabled healthcare workers to interact with and diagnose patients remotely; it opened new doors for medical professionals in the ways that they or their patients could never have imagined.

The COVID-19 pandemic also magnified structural inequalities such as class and ethnicity. In addition, there was a surge in cyberbullying and racial discrimination of Asian people (Zheng & Walsham, 2021). Lockdown has sadly increased the occurrences of domestic violence against women and children (Roy, 2021). In addition, queer people have found inequality during the pandemic given the "heteronormative" IT industry that often makes special allowances for men and women in heterosexual unions, but not for queer people (Roy, 2021).

How did women in tech, specifically, fare during the pandemic? According to Landry (2021), half of the women surveyed working in technology believe the "effects of COVID-19 have delayed their career progression, despite a similar percentage believing that much needed gender equality is more likely to be achieved through remote working structures" (para 1). Almost half of the women surveyed who were employed in technology struggled to balance work and family life since March 2020.

However, Greszler (2021) asserts this is no longer the case. COVID-19 initially disproportionately affected women because they lost more jobs and were more likely to stay at home.

The IT industry and employee compensation has changed due to the COVID-19 pandemic. As employees are working increasingly remotely and are leaving urban areas, employers are looking at different salary strategies. For example, Zuckerberg revealed that Facebook Inc. employees who work remotely and elect to move will be paid based on their new location. Many other firms, including Box, Inc., and Slack Technologies are investing similar strategies (Melin & Grant, 2020).

The number of jobs in the US IT market appears to have recovered those jobs that were lost due to COVID-19 (Gruman, 2021) According to the Bureau of Labor and Statistics at the end of 2020 there were 33,200 IT jobs in the US.

Dice.com, a leading database firm for IT employees, analyzed more than 6 million job tech job postings in the US during the first 4 months of 2020 (Bhalerao, 2020). In its 2020 Tech Job report, the company cited the top 15 tech jobs during that period for which companies were hiring. They were:

1. Software Developer
2. Network Engineer
3. Systems Engineer
4. Senior Software Developer
5. Java Developer
6. Software QA Engineer
7. IT Project Manager
8. Application Developer
9. Computer Support Specialist
10. Business Analyst
11. Computer Programmer
12. Systems Administrator
13. Graphic Designer
14. Cybersecurity Engineer
15. DevOps Engineer

Bhalerao (2020) argues that many employers were de-prioritizing new projects during the pandemic to "focus their efforts on their core product offerings and infrastructure maintenance." (para 5).

According to the Tech Salary report by Dice.com (2021), overall technologist salaries in the US increased by 3.6% between 2019 and 2020, averaging $97,859. The report indicated that the fastest growing salaries in tech were in the areas of cybersecurity, data scientist, DevOps Engineer, Tech Support Engineer, and Cloud Engineer. Table 1 shows the salaries and changes in the salaries for occupations in the IT field. The report also investigated the salary change of IT Professionals. In 2020, 52% received a salary increase, 35% experienced no change and 13% experienced a salary decrease. An item worth noting was the 40% of the people surveyed indicated that their potential salary increase was put on hold during the COVID-19 pandemic.

OCCUPATION	2020	YEAR/YEAR CHANGE
IT Management CEO, CIO, CTO, VP, Dir.	$143,416	▼ 1.7%
Systems Architect	$140,658	▲ 1.7%
Cloud Engineer	$136,479	▲ 6.3%
Cybersecurity Engineer	$134,340	▲ 4.3%
Data Architect*	$133,064	▲ 3.2%
Program Manager	$122,818	– N/A
Management Consultant	$121,619	– N/A
Product Manager	$120,584	▼ 0.6%
Data Scientist*	$119,898	▲ 12.8%
MIS Manager	$119,877	▲ 2.5%
Data Engineer	$118,621	▲ 4.7%
Project Manager	$116,911	▲ 0.8%
DevOps Engineer*	$115,125	▲ 12.2%
Systems Engineer	$113,272	– N/A
Software Developer	$111,297	▲ 1.9%
Cybersecurity Analyst	$103,106	▲ 16.3%
Database Administrator	$99,038	▼ 4.9%
Business Analyst	$97,633	▲ 5.3%
UX/UI Designer*	$91,941	▲ 1.8%
Network Engineer	$91,561	▲ 1.4%
Mainframe Programmer*	$91,386	▼ 11.2%
Application Support Engineer	$90,039	– N/A
QA Engineer	$89,543	▲ 1.7%
Systems Analyst	$88,401	– N/A
Systems Administrator	$83,490	▲ 0.6%
Web Developer	$81,550	▲ 4.9%
Data Analyst	$76,001	– N/A
Technical Support Engineer	$68,651	▲ 8.2%
Help Desk Technician	$51,553	▼ 4.0%

Table 1. Average Salaries by Occupation from Dice.com

In a different report by the CEO of Talent, Colin Etheridge (2021), found that the two major factors that have increased the demand and opportunities for IT workers in the US is the move to remote working and the rise in the digital economy. The brick-and-mortar sector is finding the increasing need to go digital.

While the demand for IT talent is increasing, the job postings for other jobs have increased dramatically in the last few months. According to ZipRecruiter.com, the number of job postings has steadily increased but the labor force participation rate has remained flat (Figure 2). Popken (2021) speculates that the reason for

the lack of change in the labor market is due to several reasons:

- Those not seeking employment due to lack of confidence, after trying to look for a job earlier in the year.
- Ongoing concerns about the virus, and childcare.
- Economic impact payments.

Figure 2. ZipRecruiter Labor Force Participation and job postings

Jobs with the fastest growing demand, according to Lewis (2021), include big data developer and quality assurance engineer. April 2021 LinkedIn job posts showed jobs with the most demand overall included software engineer, application developer and project manager (Lewis, 2021).

3. METHODOLOGY

In order to analyze the specific job trends in information technology and how they compared to other employment demands, we analyzed raw data from Burning Glass technologies. Burning Glass has the following claim, "Powered by the world's largest and most sophisticated database of labor market data and talent, we deliver real-time data and breakthrough planning tools that inform careers, define academic programs, and shape workforces." (Burning Glass Technologies, 2021). We downloaded monthly data for the time period Jan 2020 to April 2021. These data included all job postings as well as specific job postings in science, engineering, information technology. They were analyzed using SPSS 26 and Microsoft Excel.

4. RESULTS

Appendix 1 and Table 2 show the top ten significant IT job titles and the correlation of their job trends over the period January 1, 2019 to April 30, 2021 according to Burning Glass Labor Insight. Software/Developer Engineer job trends far exceeds all the other job titles and shows a trend that maps to the pandemic timeframe. Job postings generally grew until March 2020 when there was a steep decline (Figure 3). This trend continued until November 2020 when a slow recovery started, and which continues today.

	Software Developer / Engineer	
	Pearson Correlation	Sig. (2-tailed)
Software Developer / Engineer	1	
Computer Support Specialist	.894**	0.00
IT Project Manager	.929**	0.00
Systems Analyst	.946**	0.00
Computer Systems Engineer / Architect	.946**	0.00
Network Engineer / Architect	.916**	0.00
Network / Systems Administrator	.935**	0.00
Cyber / Information Security Engineer / Analyst	.873**	0.00
Web Developer	.922**	0.00
Software QA Engineer / Tester	.949**	0.00

Table 2. Top Ten Job Postings

To analyze each job title, the numbers were normalized to a 100% January 2019 base (Appendix 2). The normalized chart shows that generally each job category rose and fell and rose consistently across all job categories. A correlation analysis of trends across the ten jobs all have paired correlation coefficients above .873 and all are significant at $p < .001$. We can therefore suggest that all IT jobs were affected similarly by the pandemic.

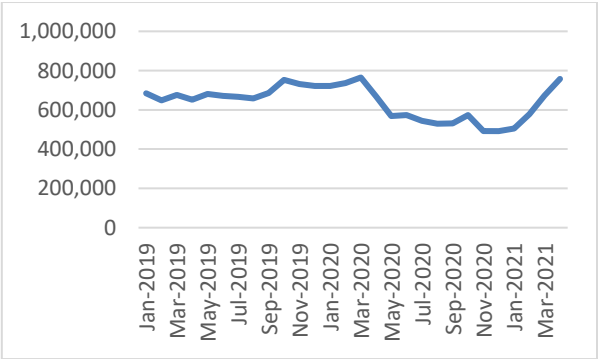


Figure 3. Total IT Job Postings

Figure 3 and Table 3 shows the full category of IT job postings during the period January 1, 2019 to April 30, 2021. The growth is fairly steady through March 2020. The COVID decline starts in April 2020 and continues through December 2020. The rebound growth began in January 2021 and continues through our available data period of April 2021. It should be noted that April 2021 job postings have recovered fully, and April 2021 was 11% above January 2019.

Period	IT	ALL
Jan-19	0%	0%
Feb-19	-5%	-5%
Mar-19	-1%	0%
Apr-19	-5%	2%
May-19	0%	6%
Jun-19	-2%	1%
Jul-19	-2%	-2%
Aug-19	-4%	-1%
Sep-19	0%	1%
Oct-19	10%	9%
Nov-19	7%	3%
Dec-19	6%	3%
Jan-20	6%	6%
Feb-20	8%	10%
Mar-20	12%	15%
Apr-20	-2%	-8%
May-20	-17%	-17%
Jun-20	-16%	-9%
Jul-20	-20%	-3%
Aug-20	-23%	2%
Sep-20	-22%	8%
Oct-20	-16%	18%
Nov-20	-28%	9%
Dec-20	-28%	8%
Jan-21	-26%	11%
Feb-21	-15%	17%
Mar-21	-2%	37%
Apr-21	11%	51%

Table 3. Change in IT and all job postings

Though this seems to be an excellent recovery, we next examined how IT jobs fared compared to the economy as a whole. Figure 4 shows normalized job postings for All jobs in the Burning Glass database versus solely IT jobs. As is apparent, total job postings have far exceeded IT jobs since the COVID rebound.

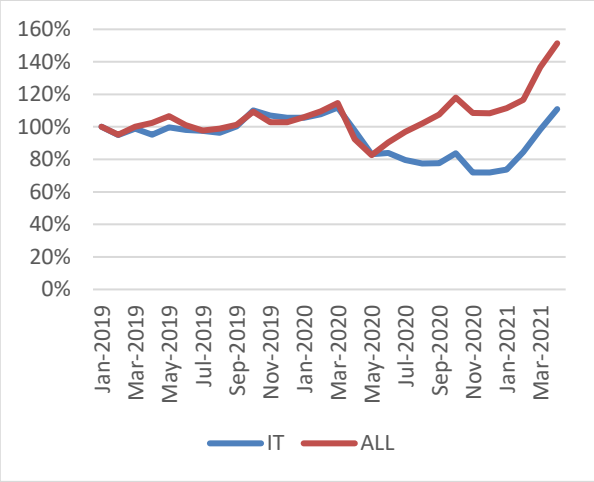


Figure 4. IT jobs Posting Vs All Job Postings

The match between total jobs was nearly perfect before and during the pandemic. But this has not been the case in the job market return. From January 2019 to March 2020 the correlation coefficient is .844 and $p < .000$. For the entire period though correlation is .219 and $p < .262$

Correlations

		Science	Engineering	IT
Science	Pearson Correlation	1	.612**	.304
	Sig. (2-tailed)		.001	.115
	N	28	28	28
Engineering	Pearson Correlation	.612**	1	.874**
	Sig. (2-tailed)	.001		.000
	N	28	28	28
IT	Pearson Correlation	.304	.874**	1
	Sig. (2-tailed)	.115	.000	
	N	28	28	28

**. Correlation is significant at the 0.01 level (2-tailed).

Table 4. Science, Engineering, IT Demand

If we examine how IT jobs have recovered since the trough compared to other STEM positions, we see a similar puzzling lag in job growth. There is a significant correlation between Science and Engineering over the COVID time period to present. There is not a significant correlation between Science and Engineering and IT (figure 5).

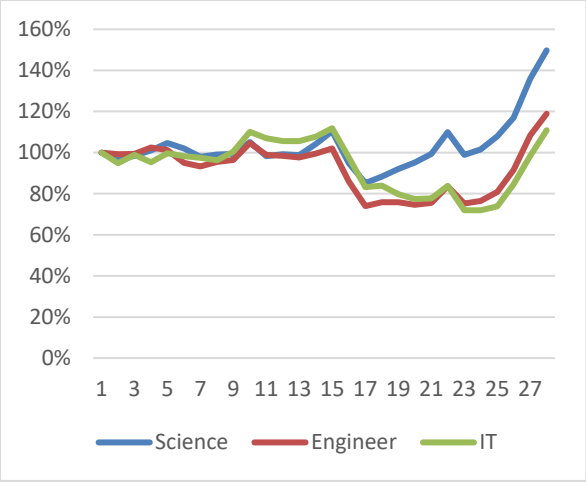


Figure 5. Relative Stem Job Growth

An area where there is demand that correlates with Science and Engineering is the IT sub-specialty, Cybersecurity. There is significant correlation for cybersecurity with both science and engineering (Table 5).

		Science	Engineering	CYBER
Science	Pearson Correlation	1	.612**	.427*
	Sig. (2-tailed)		.001	.023
	N	28	28	28
Engineering	Pearson Correlation	.612**	1	.662**
	Sig. (2-tailed)	.001		.000
	N	28	28	28
Cyber	Pearson Correlation	.427*	.662**	1
	Sig. (2-tailed)	.023	.000	
	N	28	28	28

Table 5. Science, Engineering, Cybersecurity Demand

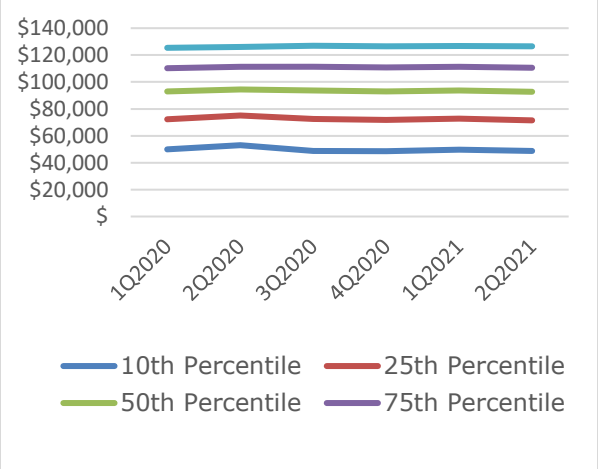


Figure 6. IT Salary Trends Over Covid Timeframe

Finally, we examined IT salary trends over the COVID timeframe (figure 6). Here we see that, in general, IT salary levels at all breakpoints are nearly unchanged from beginning to end.

5. DISCUSSION & CONCLUSIONS

COVID-19 disrupted business as usual for the worldwide workforce. Many workers lost their jobs, some permanently. For information technology workers, the good news is that IT employment is less than one percent below pre-COVID levels (Davis, 2020) and has created new opportunities. The economic fallout from the pandemic temporarily reduced demand in some skill sets, but, fortunately, strong job growth underscores longstanding talent shortages in IT. Expanding digital infrastructure was and still is important, "given the pandemic-fueled boost to the online economy" (Lund, 2021, pg. 20).

According to Ishani (2021), IT and software services companies shifted their focus to newer technologies such as data analytics, artificial intelligence, cloud computing and cybersecurity during the pandemic because of the impact these services have on the economy as a whole. These technologies, which were already growing, will require newer skill sets.

Furthermore, COVID-19 may "propel faster adoption of automation and artificial intelligence" (Lund, et al, 2021, pg. 11). Futhermore, workers will need to "learn more social and emotional skills, as well as technological skills, in order to move into occupations in higher wage brackets" (pg. 18). Women, young, less-educated workers, ethnic minorities, and immigrants "may need to make more occupation transitions after COVID-19" (pg. 19).

Our results show that job postings for information technology workers since the pandemic are lagging behind all jobs as well as science and engineering jobs. The reasons for this are unclear and require further study. Specifically, however, our results show that there is a strong demand for cybersecurity specialists. According to Vohra (2020), "the role of cybersecurity will gain greater traction in the post-COVID-19 era" (para. 2) and cybersecurity startups will "earn the favour of the investors" (para. 8). Davis (2020) asserts that professionals will want to focus on high-demand skills such as AI, cloud and cybersecurity.

6. REFERENCES

Bhalerao, S. (2020, May 23). *Top 15 Information Technology Jobs With Maximum Demand & Salary In Q1, 2020.* Retrieved

June 1, 2021, from Trak.in: https://trak.in/tags/ business /2020/05/23/top-15-information-technology-jobs-with-maximum-demand-salary-in-q1-2020/

Burning Glass Technologies. (2021) "Labor Insight™ Real-Time Labor Market Information Tool." [Burning Glass Technologies. (2019) "Labor Insight™ Real-Time Labor Market Information Tool." [

Campello, M., Kankanhalli, G., & Muthukrishnan, P. (2020, May). Corporate Hiring Under COVID-19: Labor Market Concentration, Downskilling, and Income Inequality. Cambridge, MA, USA: National Bureau of Economic Research. Retrieved June 9, 2021, from http://www.nber.org/papers/w27208

Davis, J. (2020, November 11). IT Employment Looks Up; Data, Cybersecurity Skills in Demand. Information Week India. Retrieved June 9, 2021, from https://www.informationweek.com/ strategic-cio/it-employment-trending-up-data-cybersecurity-skills-in-demand/d/d-id/1339418

Dice.com (2021) The Dice.com Salary Report the 2021 Edition for Technologists, retrieved June 22, 2021 from https://techhub.dice.com/Dice-2021-Tech-Salary-Report.html

Dice COVID-19 Jobs Resource Center Launches Alongside Q1 Tech Job Report, Shows Early Impact of Coronavirus on Technology Hiring. (2020, April 22). Retrieved June 2, 2021, from Cision PR Newswire: https://www.prnewswire.com/news-releases/dice-covid-19-jobs-resource-center-launches-alongside-q1-tech-job-report-shows-early-impact-of-coronavirus-on-technology-hiring-301044800.html

Etheridge, C. (2021) Market Overview, North America retrieved June 22, 2021 from https://www.talentsalaryguide.com/north-america

Feng, Z., & Savani, K. (2020). COVID-19 Created a Gender Gap in Perceived Work Productivity and Job Satisfaction: Implications for Dual-Career Parents Working From Home. *Gender in Management: An International Journal, 35*(7/8), 719-736.

Greszler, R. (2021, May 6). How Has COVID-19 Affected Women in the Workplace? (3617). Washington, DC: The Heritage Foundation. Retrieved June 6, 2021, from http://report.heritage.org/bg3617

Gruman, G. (2021) US IT jobs growth continues, with pandemic in rearview mirror. ComputerWorld. Retrieved June 22, 2021 from https://www.computerworld.com/article/3542681/us-it-jobs-growth-continues-with-pandemic-in-the-rearview-mirror.html

Javaid, M., & Khan, I. H. (2011, January). Internet of Things (IoT) Enabled Healthcare Helps to Take the Challenges of COVID-19 Pandemic. *Journal of Oral Biology and Craniofacial Research, 11*, 209-214. doi:https://doi.org/10.1016/j.jobcr.2021.01.015

Landry, G. (2021, May). Female Lockdown Barriers. *USA Today, 149*(2912), p. 27.

Lewis, G. (2021, May 27). *The Most In-Demand Jobs Right Now.* Retrieved June 2, 2021, from LinkedIn Talent Blog: https://www.linkedin.com/business/talent/blog/talent-strategy/most-in-demand-jobs

Lund, S., Madgavkar, A., Manyika, J., Smit, S., Ellingrud, K., Meaney, M., & Robinson, O. (2021). *The Future of Work After COVID-19.* New York: McKinsey Global Institute. Retrieved June 9, 2021, from https://www.mckinsey.com/featured-insights/future-of-work/the-future-of-work-after-covid-19

Melin, M., & Grant, N. (2020) American Tech Workers Face Pay Cuts for Relocating During Covid: Bloomberg.com retrieved June 22, 2021 from https://www.bloomberg.com/news/articles/2020-10-05/covid-effect-tech-workers-face-salary-cuts-if-they-relocate-to-cheaper-places

Popkin, B., The U.S. is having a "vaccination job boom" as confidence in the job market plummets, retrieved June 22, 2021 from https://www.nbcnews.com/business/economy/there-are-now-more-jobs-available-pandemic-so-why-aren-n1263669

Roy, R. (2021, Spring). Working from Home: Women in Indian Tech-Industy through the Pandemic. *Journal of Comparative Literature and Aesthetics, 44*(1), 56-67.

Vajpai, I. (2021, March). Qualitative Assessment of the Impact of Post-COVID-19 on Indian IT Sector Employees. *International Conference on Post Covid Challenges on Life and Livelihood (ICPCC).* Kota, Rajasthan, India.

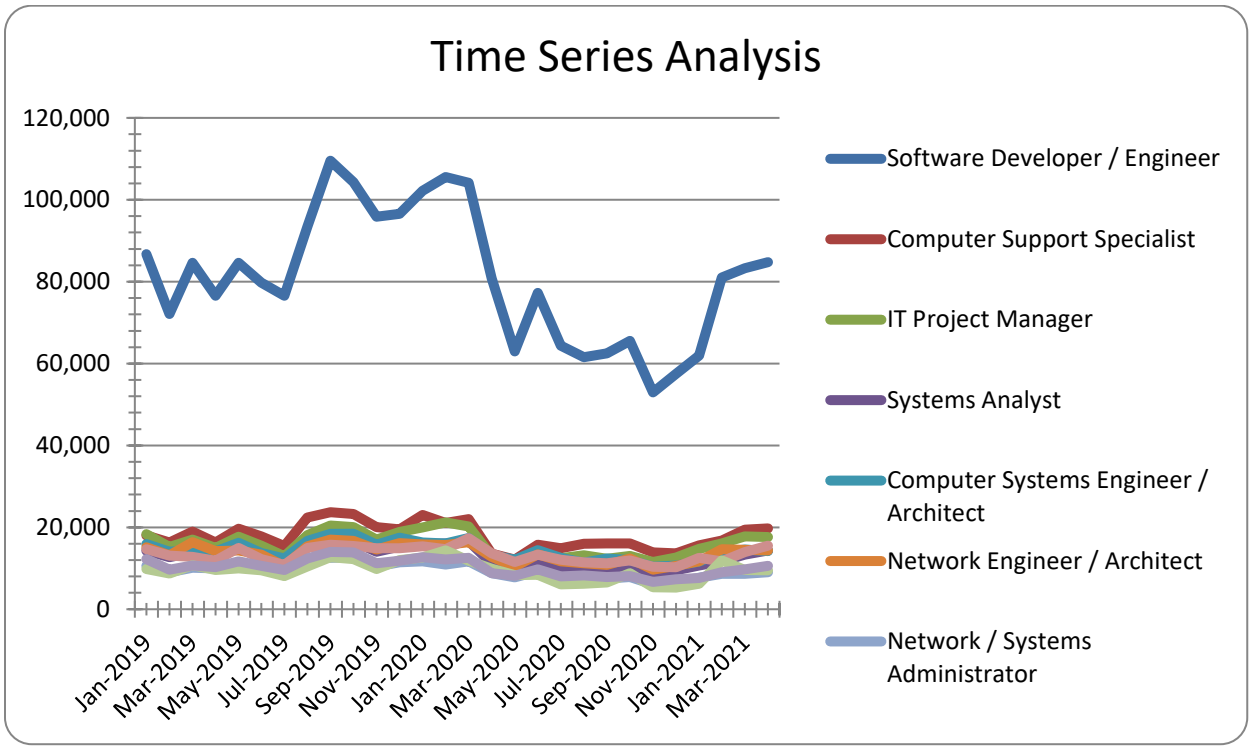
Vohra, G. (2020, May 10). *Why Career In Cybersecurity Will Be In Demand Post COVID-19.* Retrieved June 1, 2021, from BWEducation: http://bweducation.businessworld.in/article/Why-Career-In-Cybersecurity-Will-Be-In-Demand-Post-COVID-19-/10-05-2020-191587/

Zheng, Y., & Walsham, G. (2021, February). Inequality of What? An Intersectional Approach to Digital Inequality Under COVID-19. *Information and Organization, 31*, 1-6.

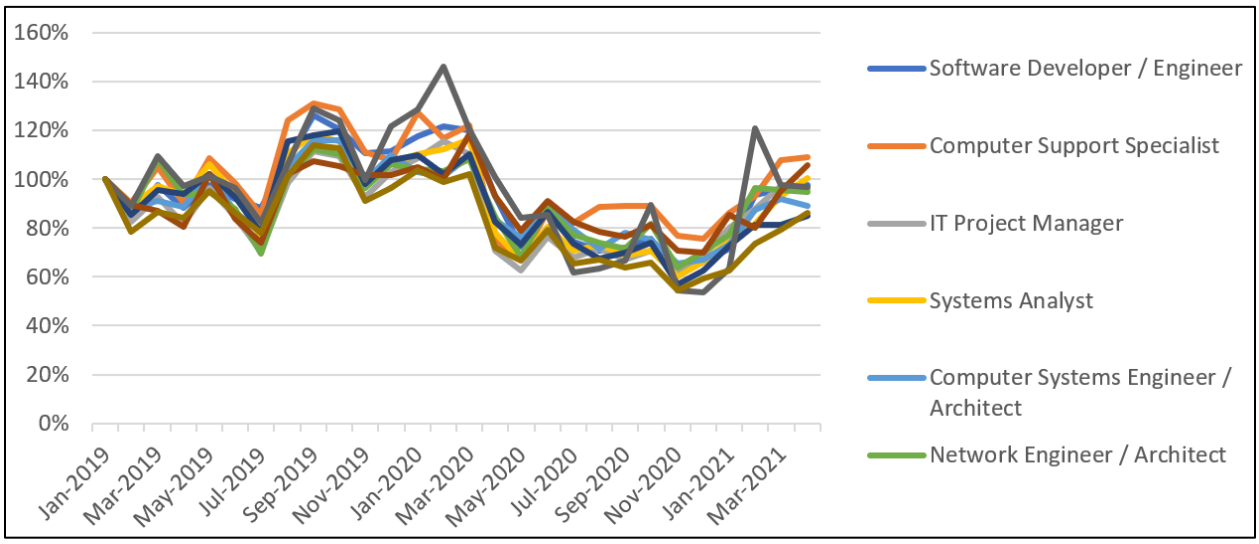
Editor's Note:

This paper was selected for inclusion in the journal as an CONISAR 2021 Meritorious Paper. The acceptance rate is typically 15% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2021.

Appendices



Appendix 1. Job Postings by IT Position



Appendix 2. Normalized Job Postings by IT Position

A Comparison of Internationalization and Localization Solutions for Web and Mobile Applications

Pen Wang
wangkevin@cityuniversity.edu
Pinterest, Inc.

Hee Jung Sion Yoon
yoonhee@cityu.edu

Sam Chung
chungsam@cityu.edu

School of Technology and Computing
City University of Seattle
Seattle, WA

Abstract

Building web and mobile applications that quickly adapt to the language, currency, number formatting, etc., of different regions – called internationalization and localization – has become more critical for most companies since the Internet allows these applications to reach foreign customers easily. However, the high development and maintenance cost and negative performance impact are two significant problems for implementing internationalization and localization functionalities. This paper analyzes current solutions that are handling the internationalization and localization problem for web and mobile applications. The advantages and disadvantages of each approach are listed and compared. Based on the information from the analysis, a new system is designed to offer a better internationalization and localization solution with a low cost and a low-performance impact.

Keywords: Internationalization, Localization, Web Application, Mobile Application, Cloud Computing

1. INTRODUCTION

Nowadays, companies grow faster when they can ship their products globally. The software industry is also taking advantage of the Internet to deliver applications or solutions to foreign markets more than ever. However, it is hard to make an application fit into different local markets due to the language and culture differences between regions. Having a remarkable ability to handle the internationalization and localization for software becomes very crucial, which can help a company to achieve a higher customer satisfaction rate, more market share, and lower maintenance costs (Saito et al., 2017).

Internationalization in software development is a term that talks about how to develop software that can quickly adapt to other markets, i.e., other languages and cultures (Kockaert & Steurs, 2015, p. 451). Kockaert and Steurs (2015) also mentioned in their book that localization is the process of adapting a product to a local market. The localization process can include translation, date and time formatting, units converting, currency converting, and so forth.

Problem Statement
Implementing the internationalization and localization for web and mobile applications can cause a high cost during the development and maintenance process and a considerable performance impact. Therefore, managing the internationalization and localization for software can be very challenging since it will significantly increase the workload and cost due to multiple versions that may need to be created simultaneously.

Long delivering time is another problem because changing a new version may require changing every file containing text, symbols, images, videos, etc. Most importantly, adding a new feature will become more complex and time-consuming because multiple versions' software has to be maintained simultaneously. Moreover, the approach used to handle internationalization and localization may give an original system a significant performance impact due to more complexity.

Motivation
The first motivation is to find a way to allow the software to improve its user experience through internationalization and localization. Hau and Aparfcio (2014) mentioned that users always expect the software to show their languages, which can help raise productivity and significantly reduce mistakes.

The second motivation is to find a more effortless and cheaper solution for implementing and maintaining the internationalization and localization feature for web and mobile applications. According to Kidambi (2016), 60% to 80% of the total life-cycle costs for software is maintenance cost. Thus, how easy it is to maintain an application after adding the internationalization and localization solution becomes very important.

Approach
This paper evaluates how different frameworks handle the internationalization and localization problem and the non-framework way. We list the advantages and disadvantages of the existing approaches. We also compare them to find a way to improve. The ideal goal is to have a solution that can offer all the existing solutions' benefits without extra work and maintenance effort.

2. RELATED WORK

he best way to implement internationalization and localization for web and mobile applications have been discussed for a long time (Sugiura, 1986). There are many different solutions out there. Here are some popular industry solutions using front-end technologies in JavaScript:

- React applications with React-intl library (Facebook and Community)
- Angular applications (Google)
- Globalize library (jQuery Foundation)
- Android applications (Google)

React applications with React-intl library
React.js is a prevalent web application user interface library that can help developers to develop single-page web applications. It has many different libraries to help to handle internationalization and localization challenges.

The react-intl library is one of them. React-intl's (2019) official documentation can format message, date, time, number, and handle the plural issue. Developers can enable the functionality by wrapping the root component with the IntlProvider component, a higher-order component offered by the library.

A FormattedMessage component is used to tell the application to use the different messages based on the users' language setting. Another higher-order function inJectIntl is used to inject the intl object that contains format functions for the date, time, and number formatting. Using the higher-order function to wrap and inject functions makes this library very easy to use. Moreover, it also means this library will work with React library. Another downside is that the translation text files have to include the application itself, which requires republishing the application after adding a new language or updating some existing texts.

Angular applications
Angular is another popular web application framework that Google develops, used by over 1.9 million developers (2021). It also comes with its internationalization and localization solution. Angular's (2019) documentation can handle date, number, percentages, currencies, message, and plural forms of words. Moreover, Angular offers a Command Line Interface (CLI) tool to help developers generate necessary files for translators. It also can help to publish applications in multiple languages.

The following processes will be conducted after the internationalization is setup:

- Extracting localizable text for translation
- Building and serving the application with the translated message based on users' locale

- Creating multiple versions for different languages

The strength of this approach is that all of the necessary tools are included in the Angular framework, and developers can use them out of the box. It is easy to add new features with different languages since the CLI tool will extract the files automatically and allows translators to work on the text without touching any code. Moreover, this approach can be used with Angular applications since it is an internal tool for the Angular framework.

Globalize library

Globalize is a JavaScript library that aims to offer internationalization and localization capability to web applications. According to Rosa (2016), the Globalize library leverages the official Unicode Common Locale Data Repository (CLDR) JavaScript Object Notation (JSON) data, and very easy to have the latest CLDR data (CLDR, 2019). The features of the library include:

- Number formatting
- Date formatting
- Time formatting
- Currency formatting
- Message formatting
- Plural and unit formatting

The Globalize library's (2019) official website shows that using the library is very simple. After requiring the library and loading the CLDR data, developers need to call the different formatters such as currencyFormatter, numberFormatter, dateFormatter, and so forth.

The strength of using this approach is that it will work for all of the web applications and some of the mobile applications (using JavaScript technology such as React Native or progressive web app) since it is essentially a pure JavaScript function. Another advantage is that the latest CLDR data will always be used. The most significant disadvantage is that the message module needs to load a local JSON file that contains messages in all languages, which requires republishing the application whenever changing or adding words in the file.

Android applications

Android is another popular development platform with around 3.48 million mobile apps available in the Google Play app store by the first quarter of 2021 (Statista Research Department, 2021). It also officially supports the internationalization and localization functionality of its platform. According to the Android developers' documentation (2019), Android developers can use the resource framework to separate the localized aspects from core functionality code. Android applications will switch the resources such as static data, images, videos, sounds, logos, texts, and so on based on users' languages preference. Developers can just simply put the different localized resources into other folders with the correct language naming convention. For example, the message resource for English could be placed under the res/values-en/strings.xml when the French message resource could be put under the res/values-fr/strings.xml.

The advantage of this approach is that this is a build-in tool offered by Android, which makes the workflow very clean. It also can efficiently deal with all kinds of resources besides the text, such as images, sounds, and videos. The disadvantage is that this approach works for Android since it leverages Android resource loader to switch between different resources.

Summary

After we reviewed and evaluated several different current solutions, we summarize the findings as below:

- Most of the solutions are tied to specific frameworks or platforms.
- Offering a way to extract text for translation is very important.
- Adding new languages should not require republishing.
- Updating texts should not require republishing.
- All resources such as text, image, currency, etc., should be automatically switched to the correct format based on users' preference language.
- The approach should keep developers' extra work as little as possible.
- The approach should have the ability to handle text, image, audio, video, date, time, currency, and unit formatting.

Therefore, it is good to have the ability to update CLDR data to the latest version automatically.

3. APPROACH

Our approach described in this paper for solving the internationalization and localization issue includes five parts:

1. Use plain JavaScript to fit web and mobile development with all frameworks: Using the

plain JavaScript implementation can make sure the solution can be used by any frameworks such as Angular, React, Vue, and so on (Rauschmayer, 2019; Tackaberry, 2018). It can work without any framework as well (Osetskyi, 2019; Tin, 2018). The mobile applications that use JavaScript technology also work fine with this solution.

2. Separate the text content: All unrelated texts are extracted and stored in a separate file. This approach allows interpreters to work on only text files without touching the programming code.

3. Use the resource loader concept: Implementing a resource loader looks like a mechanism to allow the applications to load different images, videos, sounds, and CSS rule-based on users' languages and regions.

4. Leverage the CLDR rules: Automatically update the CLDR rules from the database to ensure the application uses the newest localization rules.

5. Use an independent cache layer to keep the resources: An in-memory cache layer is used to keep all localization-related resources such as text, images, videos, CSS rules, and so forth to reduce the application package size and allow end-users to download the necessary resources with low latency. The in-memory cache layer should also be easy to scale out with cluster mode when required.

How to integrate our system

We made our system a library and published it on a Node Package Management (NPM) system used to share code). Users can integrate and use it with the following steps:

1. Install the library into their existing system with the following comment:

```
npm install --save @kevinwang0316/i18n
```

```
// Define your dictionary for every language you want to support.
const dictionary = {
'en-US': { // Set the dictionary for the U.S. users
    login: 'login',
    confirm: 'confirm',
  },
  'es': { // Set the dictionary for Spanish users
    login: 'iniciar sesión',
    confirm: 'confirmar',
  },
  'zh-CN': { // Set the dictionary for Simplified Chinese users
    login: '登录',
    confirm: '确认',
  }
};
```

2. Add a dictionary file with all translated text content (this is stored in the cache layer after integrating with a cache system such as Redis or Memcached):

3. Import the translation text file and initialize the library (usually in the entry file):

```
// Set the dictionary to the I18n
I18n.setDictionary(dictionary);

// Optionally, you can set up a default language. If the user browser language is not found in the dictionary, this default language will be shown.
I18n.setDefaultLanguage('en-US');
```

4. Use the library in the place you need:

```
import I18n from '@kevinwang0316/i18n';

const YourComponent = () =>
<button>{I18n.get('login')}</button>;
```

How the integrated system works

The testing system has not integrated with the CLDR rules system and caching mechanism. After these two parts are done, the system works as:

1. Get the user's location setting config information from the browser.

2. One back-end call to fetch the newest translation and rule resource files comes from the cache.

3. Initialize the library with the resource files.

4. Swap the content based on the content in the resource files.

4. DATA COLLECTION

Since we design our solution as a whole system to make the internationalization and localization process easier for web and mobile applications, we collect the following data for the web application to measure the performance impact:

- Front-end CPU usage
- Page loading time
- Resources retrieving latency

All testing is conducted on the Macbook Pro 2014 version with a 2.2 GHz Quad-Core Intel Core i7 CPU and 16G memory.

Front-end CPU usage
The CPU usage was collected by using the Chrome DevTool profiling feature. The extra CPU usage that our system adds to the original system will be crucial for performance. If our system adds a considerable amount of CPU overhead, the functionalities of the original system may be impacted a great deal. The detailed data can be viewed in Figure 1.

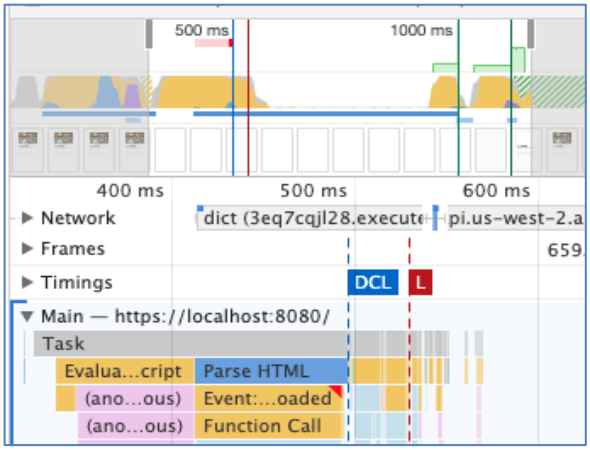


Figure 1. CPU usage

Page loading time
The page loading time can be increased significantly after using our solution since it requires loading extra resources from the Internet based on users' settings. Thus, collecting and monitoring this data is very important. It is collected by using the network module in the Chrome DevTool. The detailed data can be viewed in Figure 2.

Resources retrieving latency
Since our system can retrieve different resources such as text, video, audio, images, and so forth, the retrieving latency time should be considered vital data that has to be collected. This job can be done using AWS Cloud Watch and AWS X-Ray since our system will be integrated with AWS's services. Figures 3 and 4 show the data from AWS Cloud Watch and AWS X-Ray.

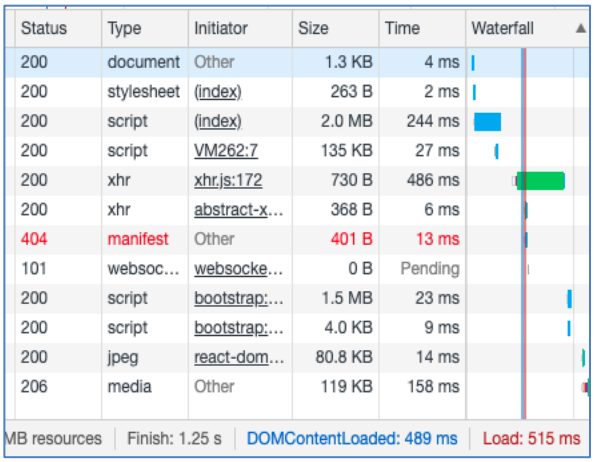


Figure 2. Page loading time

Figure 3. AWS Cloud Watch latency

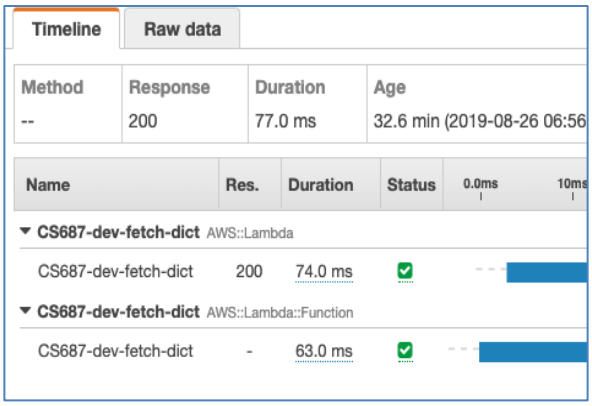


Figure 4. AWS X-Ray latency

5. DATA ANALYSIS

Three different kinds of data were collected for analysis purposes, which will be analyzed with various methods in this section.

Front-end CPU usage
The data in Figure 1 shows that our system has a shallow CPU footprint. After parsing the script, the execution phrase did not cause any high CPU usage and even finished before the HTML was parsed. Figure 1 also shows that the total script execution time is just a litter bit over 100ms.

Page loading time
Figure 2 shows that the total loading time is 515ms and the total finishing time is 1.25s. We collected the data by using the Chrome DevTool network panel. Because this test was run under the development environment that did not use a production build, the final loading time can be even lower since the production build will use multiple techniques such as minification, tree shaking, etc.

Resources retrieving latency
In the back-end code, a utility tool is written for collecting the data for a specific step. In this case, the latency of retrieving data from Redis is monitored by the utility tool and logged out to the AWS CloudWatch. Redis is an in-memory data structure store, used as a distributed, in-memory key-value database, cache, and message broker, with optional durability. Figure 3 shows that the time spends on retrieving a resource from the cache layer (Redis) is 42ms. Figure 4 shows the execution time for the whole back-end function (a warm Lambda function), 63ms.

6. FINDINGS

The finding will be shown in two parts to illustrate how our system impacts performance and whether this system is easy and cheap to use.

Performance Impact
- Low impact of CPU usage: The analysis in Section 6 shows that our system does not add any noticeable CPU impact to the original system. It means our system's impact on CPU usage is low.
- Fast page loading time: The page loading time analysis shows the whole page is loaded in 600ms. According to Google PageSpeed Insights (2019), the website will be considered fast if its First Contentful Paint (FCP) is under 1,000ms. Thus, the system does not harm the page loading time.
- Resources retrieving latency: The resource retrieving latency analysis shows latencies for resource retrieving from both the Redis and back-end function calls are very low, which will not significantly impact the original system.

Ease of Use
To use this system, we conduct the following three steps:
- Use a placeholder for all dynamic content instead of hard coding
- Create and fill out the resource template every time your system wants to add a new region support
- Add the resource template to the Redis server

Only these three steps need to be done to use the system, which is fairly to say it is straightforward to use. Additionally, adding and updating resources and other regions' support does not require any client-side or server-side code changing or redeploying, which causes the maintain cost very low.

7. CONCLUSION

Adding the internationalization and localization feature for web and mobile applications can cause a severe development and maintenance cost and a substantial negative performance impact. The system designed in this paper leverages the resource loader concept, cloud computing, and in-memory cache technology to balance developing cost, maintenance effort, and performance impact. The data collected and analyzed in the paper shows this system can help web and mobile applications handle the internationalization and localization functionality with several benefits such as a very low-performance impact in terms of CPU usage, loading page time, and resource retrieving time, a very low implementation and maintenance cost due to the ease of use.
We are not comparing the performance with other existing systems since this paper aims not to show how our system can improve the performance but to demonstrate that our system does not have a significant performance impact.

8. FUTURE WORK

There are three significant improvements to this system and could be done in future work. Firstly, make the data persistent and automatically load the data into Redis. All resource data are living in the Redis store, which is an in-memory database. More work should be done to make

the data persistent and allow Redis to load the data from the data source after a crash.

Secondly, remove the back-end layer. For the demo system, the AWS ElastiCache is not used to avoid the cost. The downside of this implementation is that a Lambda function has to be used to hide the Redis credentials from the front-end code. If the AWS ElastiCache is used, the system can take advantage of the AWS assumed permission mechanism to allow the front-end code to call the Redis store directly. In other words, the back-end code can be removed completely.

Lastly, offer a tool to generate the resource template based on the existing information in the Redis. All resource information is added to the Redis store manually using a JSON format for demonstration purposes. In the future, a tool should be offered to help users to generate a resource template or event offer interface based on the current information in the Redis. It can help non-technical people such as interpreters, UI designers and handle the localization process.

9. REFERENCE

Android Developers. (2019). Localize your app. Retrieved December 26, 2021 from https://developer.android.com/guide/topics/resources/localization.html#kotlin

Angular. (2021). Github Angular repository. Retrieved December 26, 2021 from https://github.com/angular/angular

Angular. (2019). Internationalization. Retrieved December 26, 2021 from https://angular.io/guide/i18n

CLDR. (2019). Unicode Common Locale Data Repository. Retrieved December 26, 2021 from http://cldr.unicode.org

Globalize. (2019). Globalize read me. Retrieved December 26, 2021 from https://github.com/globalizejs/globalize

Google PageSpeed Insights. (2019). About PageSpeed Insights. Retrieved December 26, 2021 from https://developers.google.com/speed/docs/insights/v5/about?hl=en-US&utm_source=PSI&utm_medium=incoming-link&utm_campaign=PSI

Hau, E., & Aparício, M. (2008, September). Software internationalization and localization in web-based ERP. *In Proceedings of the 26th annual ACM international conference on Design of communication* (pp. 175-180).

Kidambi, P. C. (2016). Maintenance issues in software engineering. Retrieved December 26, 2021 from http://www2.latech.edu/~box/ase/tp_2003/CS532Termpaper_Kidambi_Praveen%20Chandra.doc

Kockaert, H. J., & Steurs, F. (2015). Handbook of terminology (Vol. 1). John Benjamins Publishing Company.

Margaret, R. (2017). Web application. Retrieved December 26, 2021 from https://searchsoftwarequality.techtarget.com/definition/Web-application-Web-app

Margaret, R. (2019). Mobile app. Retrieved December 26, 2021 from https://whatis.techtarget.com/definition/mobile-app

Osetskyi, V. (2019). Web application architecture. Retrieved December 26, 2021 from https://medium.com/existek/web-application-architecture-da77ea0cb520

Rauschmayer, A. (2014). Speaking JavaScript: an in-depth guide for programmers. O'Reilly Media, Inc.

React-intl. (2019). React-intl gets started. Retrieved December 26, 2021 from https://github.com/formatjs/react-intl/blob/master/docs/Getting-Started.md

Rosa, A. (2016). How to Implement Internationalization (i18n) in JavaScript. Retrieved December 26, 2021 from https://www.sitepoint.com/how-to-implement-internationalization-i18n-in-javascript

Saito, O., Boafo, Y. A., Kranjac-Berisavljevic, G., Yeboah, R. W. N., Mensah, A., Gordon, C., & Takeuchi, K. (2018). Internationalization and Localization of the Ghana Model: Lessons Learned, Opportunities for Upscaling, and Future Directions. *In Strategies for Building Resilience against Climate and Ecosystem Changes in Sub-Saharan Africa* (pp. 333-343). Springer, Singapore.

Statista Research Department. (2021). The number of available apps in the Apple App Store from 1st quarter 2015 to 1st quarter 2021. Retrieved December 26, 2021 from https://www.statista.com/statistics/779768/number-of-available-apps-in-the-apple-app-store-quarter/

Tackaberry, A. (2018). How to set up Internationalization in React from start to finish. Retrieved December 26, 2021 from

https://www.freecodecamp.org/news/setting-up-internationalization-in-react-from-start-to-finish-6cb94a7af725

Tin, F. (2018). Automated software internationalization and localization. Retrieved December 26, 2021 from http://www.freepatentsonline.com/10078504.html

GIS for Democracy: Toward A Solution Against Gerrymandering

Peter Y. Wu
wu@rmu.edu

Diane A. Igoche
igoche@rmu.edu

Department of Computer Information Systems
Robert Morris University
6001 University Blvd,
Moon, PA 15108

Abstract

Political redistricting is periodically necessary to maintain and promote democracy with population growth and migration. The United States constitution establishes majority rule for democracy, but it also protects minority rights. There is provision that a minority group may form a political district so that the group can have representation in the government. Each state has the right to political redistricting accordingly. Since 1812, this has been referred to as gerrymandering. It was not easy to do and was not considered a serious issue. However, the Geographic Information Systems (GIS) today have made the task much easier, leading to the practice of extreme gerrymandering in the past decade. The practice is detrimental to the health of democracy, but it is difficult to legally disallow. We propose a scheme in which the GIS becomes part of the solution. The proposed scheme is to make the process of political redistricting public, to be scrutinized and debated, and perhaps voted for or against by the voting population. The politicians as well as concerned citizens will need to use the GIS. The paper calls for the promotion of GIS education for democracy, with the need for relevant data in redistricting to be publicly available.

Keywords: Gerrymandering, Political Redistricting, GIS, Geographic Information System.

1. INTRODUCTION

Gerrymandering is the practice of manipulating voting district boundaries for political gain (Griffith, 1907). Political redistricting however is necessary to account for the changes in the population, such as those reflected in the decennial census. It is also required for the protection of minority rights so that a minority group may have representation in the government (US Dept of Justice, 1965). The Constitution granted the authority of political redistricting to the states. That allows the party in power in the state government the legal right of gerrymandering. In the past, it was rarely done because the task was difficult and there was inaccurate demographic data to make the process effective. With geographic information systems (GIS) now available, and data easily accessible, gerrymandering can be done with ease (Wu, Deplato & Combs, 2020). The past decade has seen extreme cases of partisan gerrymandering, re-drawing voting districts into strange shapes for political gain (Crane & Grove, 2018; Forest, 2018). It is generally understood to be bad for democracy because it allows politicians to choose favorable voters to secure their elected positions. There have been attempts to disallow partisan gerrymandering but legally it requires proof of intent in the court

of law. We believe that the GIS can be part of the solution in this effort. This paper presents our proposal using the GIS, as well as the public knowledge of the GIS, to be our approach toward a solution.

The next section will review a brief history of gerrymandering and will explain its basic strategies: cracking and packing, and how to gain political advantage in redistricting. The section presents a simple description of how to use the GIS to simplify the gerrymandering process. Section 3 follows with a review of the effort to prevent gerrymandering. Given the context, section 4 presents the draft of our approach toward a solution, requiring plans of political redistricting to be made public, for scrutiny and debate. It requires the voting public to have access to use the GIS knowledgeably. It is therefore pertinent to promote GIS education. The last section closes with a summary and our conclusion.

2. REVIEW OF THE PRACTICE

Gerrymandering is the practice of manipulating voting district boundaries to gain political advantage in democratic voting. The term was coined in 1812 when Massachusetts governor Elbridge Gerry signed into state law to create a voting district in the shape of a salamander to include most of his supporters as majority (Griffith, 1907). It is legal since the political party in power has the privilege of drawing the map for redistricting. However, it was not a serious issue because it was difficult to execute, and accurate demographic data was not readily available for use. In the past decade, very strange shapes of voting districts emerged in political redistricting. We believe the common use of the GIS today and the ease of access to data has made the task relatively simple. Below, we will briefly explain the two basic strategies in gerrymandering: cracking and packing. Then we will describe how it is made easy using the GIS today.

Cracking

The strategy of cracking attempts to dilute the votes of the opposing party to suppress them from winning in any voting district. Cracking is the approach when the party has the majority. The voters for the minority party may be cracked in the redistricting, keeping them as minority in many voting districts. A hypothetical case is illustrated below in Figure 1. Party A of 55% majority exploits cracking in drawing five districts (in the 5 horizontal strips), distributing the 45% voters of opposing Party B evenly to win all five districts, therefore suppressing the minority party.

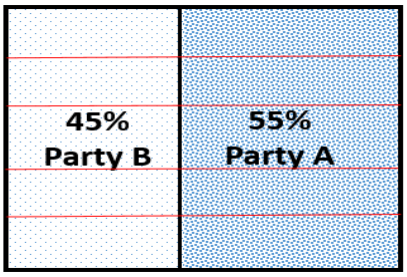


Fig.1 Cracking to Suppress the Minority

Packing

Packing attempts to concentrate the votes of the opposing party in one or a few districts to reduce the number of votes in the other districts. Packing is when the party in power is aware that they are in the minority. The redistricting will attempt to create one or a few districts packed with high percentage of voters for the opposing party. The voters not included in the packing are then distributed into the other districts so that they will not make majority, allowing the minority party to win these other districts. A hypothetical case is illustrated below in Figure 2. Party B has the 45% minority but is in power to do redistricting. One voting district shown as vertical to the right has Party A voters packed, of entirely Party A voters. The remaining Party A voters are distributed into the other four districts horizontal to the left. The result has the minority Party B winning these four districts.

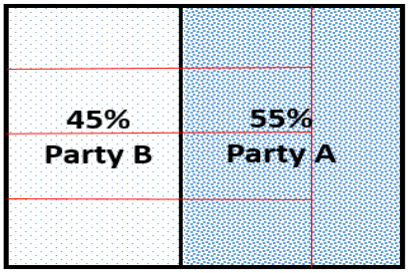


Fig.2 Packing to Limit the Majority

GIS For Gerrymandering

When data is available for use in the GIS, cracking and packing become much easier to do. Assume that we have gathered the addresses of the voters and which party they tend to vote. The GIS functionality known as *address geocoding,* uses an expert system to process the addresses to produce a point map (Wu & Rathswohl, 2010; Goldberg, 2016) as illustrated in Figure 3.

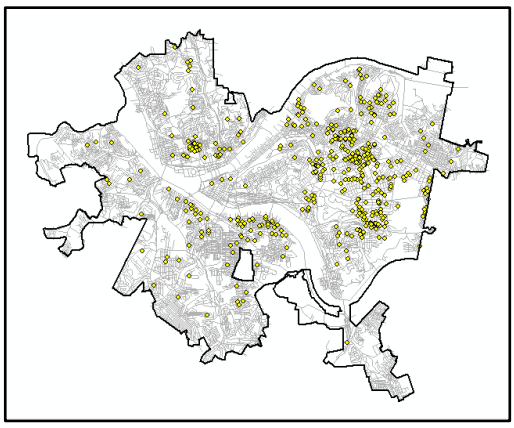


Fig.3 Point Map Showing Geocoded Locations

The map showing where the voters are located serves as our visual guide to draw the voting districts. With the point map as our base map, we can begin to draw voting districts one at a time, choosing to include or not to include areas where the voters are. Once we have drawn a district, the Spatial Join GIS function can readily verify the count of voters for or against the political party, verifying whether or not we are achieving our purpose in the effort. Figure 4 illustrates a voting district drawn to include where the voters are located.

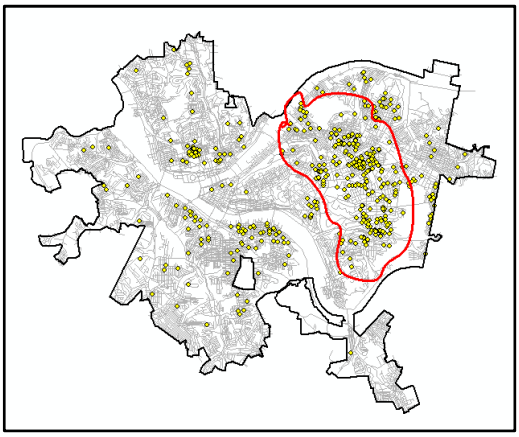


Fig.4 Drawing a Voting District

Thus, a redistricting plan can be constructed with relative ease, aided by the GIS. In the past decade, we have seen a rising number of cases of extreme gerrymandering (Crane & Grove 2018; Forest, 2018).

Wu, DePlato and Combs (2020) more thoroughly described cracking and packing, and the scheme of gerrymandering aided by a GIS. Noting the difficulties to objectively detect and therefore legally disallow gerrymandering, Wu et al called for further research in the area. This paper goes on to propose an approach toward a solution in which the GIS becomes essential.

3. TO PREVENT GERRYMANDERING

Gerrymandering is bad for democracy because it allows a politician to choose the voters by drawing the voting districts to his or her favor. This section presents the efforts attempted to prevent gerrymandering and the issues there.

To Count Total Popular Votes

Since 1824, the United States established the Winner-Take-all rule in having voting districts for presidential as well as local elections (McCarthy, 2012). The rule was originally designed to protect minority rights by allowing a minority population group to still have a voice in the democratic government. The Voting Rights Act of 1965 requires some states to have at least one district formed based on race, to ensure minority representation in the government (US Department of Justice 1965). Given that this Winner-Take-All Rule cannot be abolished, some states seek to revise it for appropriate adoption. Presently, Maine and Nebraska both practice a hybrid combination of statewide and district vote counts (McCarthy, 2012).

An Independent Commission

To prevent the political party in power from gerrymandering in redistricting, some have suggested to have a non-partisan commission in charge of redistricting. There would be no incentive to take political advantage for any party. But the problem is the same. The problem becomes: who should serve on the commission? The non-partisan commission will also have difficulty meeting the requirements of the 1965 Voting Rights Act. It is unlikely that the approach will remove gerrymandering since it only shifts the focus of the fight.

Computer Algorithms

From 1970s to 80s, founded strong in computer science, the field of computational geometry spawned many algorithms to process geometry represented in digital data (Forrest, 1971; Preparata & Shamos 1988). Much of the research work supplied for the GIS functionalities today. Using the GIS for gerrymandering became practicable and some attempted to automate the process (Li, Wang & Wang 2007; Yamada 2009; Siegel-Hawley 2013; Reitsma 2013). Yet automation of the process was hardly successful, though it might have become much easier when aided by the GIS. Realizing that partisan gerrymandering is unhealthy for democracy, many envisioned to

identify it (Niemi, Grofman, Carlucci & Hofeller 1990; Flint 2003; Chou & Li 2006; Ricca, Scozzari & Simeone 2008, Altman, Amos, McDonald & Smith 2015). If we can identify partisan gerrymandering objectively by a computer algorithm, we can contest it in court and disallow it legally. While many still call for research in the area (Crane & Grove, 2018; Grofman & Cervas, 2018; Forest, 2018), it proves to be more difficult than envisioned. A paper titled "An Impossibility Theorem for Gerrymandering" by two mathematicians (Alexeev & Mixon, 2018) perhaps was more telling in theoretical terms about the situation.

Automation of Redistricting

A definitive algorithmic solution to identify partisan gerrymandering may seem elusive. But that did not dampen the enthusiasm to automate the political redistricting process. If there is a computational process to generate political boundaries objectively based on acceptable criteria, such as population data only, we do not have to allow any attempt of gerrymandering, partisan or non-partisan. In 2014, Brian Olson, an avid programmer by trade, shared his automated solution to political redistricting, as reported in The Washington Post (Ingraham 2014). Olson's work was based on population data from census and required voting district boundaries to follow census block boundaries. Figures 5 and 6 respectively show the current congressional districts in Pennsylvania and those produced by Olson's algorithm. Also, the algorithm bypasses the issues of Voting Rights Act (US Dept of Justice 1965) which in some states requires majority-minority districts to be drawn. Olson then proceeded to start his Voting and Election Reform web site at bolson.org/voting/ to discuss possible adjustments to the criteria to apply to his algorithm.

Fig.5 Pennsylvania Congressional Districts

Without a satisfactory solution, the automated redistricting was also applied to produce redistricting maps as counter examples to argue against the cases of partisan gerrymandering in court (Magleby and Mosesson 2018; Krasno, Magelby, McDonald, Donahue and Best, 2019).

Fig.6 Pennsylvania Congressional Districts by Olson

Levin and Friedler (2019) published an experimental algorithm applying a divide-and-conquer strategy to recursively sub-divide an area in triangulation to construct political districts based on various demographic criteria. The process does need to follow census boundaries. The algorithm is much more promising, albeit computationally extremely expensive.

It was also noted that the application of artificial intelligence with machine learning may be applicable (Wu, DePlato & Combs, 2020). The suggested approach has not yet been explored.

4. GIS FOR DEMOCRACY

Our intention is that the GIS can be part of the solution against gerrymandering. In this section, we propose a potential solution. We trust that the people can determine what is good for democracy. If the GIS is available for everybody, the people will have a viable tool against gerrymandering. Our proposal has several facets. We discuss them in the following.

To Require Public Scrutiny

To prevent partisan gerrymandering, it is proposed that instead of allowing the majority party in the government, we should have an independent commission responsible for political redistricting. If the party in power decides who should be on the commission, the problem remains the same. The ideal of democracy should have the entire population serving in the commission. Our proposal therefore is to have any redistricting plan to be publicly scrutinized. A redistricting plan, along with all the relevant demographic data, has to be made available to the public. Reasons for redrawing a district must be stated to allow public discourse. We need to provide use of the GIS to the public so that

anyone wanting to review the redistricting proposal may study and analyze it in detail. Politicians and the citizens concerned about democracy will need to learn enough to use the GIS for the purpose. We put our trust in the people that they may have the discernment to see that a proposed redistricting plan is doing extreme gerrymandering, through public discourse and debate.

To Allow Alternative Proposals

If the GIS tool is made available to the public, we may also allow the minority party in the government to make opposing redistricting proposals which would have to face the same level of scrutiny. In fact, it is possible to set up appropriate regulations for other alternative redistricting proposals. Such a proposal may be sponsored by relevant elected members of the government. The feasibility of a proposal can be tested by the GIS and appropriate regulations may protect minority rights.

To Vote for The Right Proposal

When there are multiple legitimate redistricting proposals, voting can then be conducted to adopt one that is accepted by the majority of the electorate, not just the majority party in the government. This however will mean that sufficient knowledge and training need to be provided for the voting public.

Our conclusion, therefore, is that the GIS can be a critical part of the solution. To promote democracy, we need to promote GIS education. The call is for IS educators to make learning GIS accessible to a broader population, and for the GIS vendors to design the GIS with ease of use, and to provide reasonable learning tools to the public. The government can facilitate for the approach while providing the GIS learning and use along with relevant data for public use.

5. CONCLUSION AND SUMMARY

We presented our proposed approach to make political redistricting a public process, to be reviewed and debated by the voting public. Our approach can be implemented in three stages:

(1) To require the proposed redistricting plan to face public scrutiny.
(2) To allow alternative proposals by minority party, or any other individual or organization.
(3) To conduct voting by the public to decide which redistricting plan to be adopted.

The GIS is part of the solution since the politicians as well as the concerned citizens will need to be reasonably knowledgeable with using the GIS for redistricting. We believe that can be a viable solution against gerrymandering.

In summary, the paper began with the brief history of gerrymandering. The basic strategies of cracking and packing were illustrated. We also presented the steps of how using the GIS can make gerrymandering easy, leading us to the belief that the GIS has been the culprit of extreme gerrymandering. We then reviewed the various approaches attempted to possibly prevent gerrymandering. In the context that there seems to be no good solution, we propose to use the GIS to make the political redistricting process public. With the GIS available, any proposal for redistricting can be scrutinized and debated. The political party for the redistricting proposal will have to justify it publicly. We also suggest allowing opposing parties to make redistricting proposals. With appropriate regulations set up, legitimate proposals may be analyzed, debated, and finally voted for or against by voters. While the government needs to facilitate for the process, the GIS will require a better intuitive design for public use, and educators should be promoting GIS education, for democracy.

6. REFERENCES

Alexeev, B. and Mixon, D.G. (2018). An Impossibility Theorem for Gerrymandering. *The American Mathematical Monthly* 125(10), 878-884.
[https://doi.org/10.1080/00029890.2018.1517571]

Altman, M., Amos, B., McDonald, M.P. and Smith, D.A. (2015). Revealing Preferences: Why Gerrymanders Are Hard To Prove, And What To Do About It, Social Sciences Research Network (SSRN) Electronic Journal (Mar 22, 2015).
[https://dx.doi.org/10.2139/ssrn.2583528]

Chou, C.I. and Li, S.P. (2006). Taming the Gerrymander – Statistical Physics Approach to Political Districting Problem. *Physica A: Statistical Mechanics and Its Applications*, Elsevier.
[https://doi.org/10.1016/j.physa.2006.01.082]

Crane, N.J. and Grove, K. (2018). An Active Role For Political Geography In Our Current Conjuncture. *Geography Compass* 12(11). Wiley Online Library.
[https://doi.org/10.1111/gec3.12410]

Flint, C. (2003). Political Geography: Context and Agency in A Multiscalar Framework.

Progress in Human Geography 27(5), 627-636.
[https://doi.org/10.1191/0309132503ph453pr]

Forest, B. (2018). Electoral Geography: From Mapping Votes to Representing Power. *Geography Compass* 12(1). Wiley Online Library.
[https://doi.org/10.1111/gec3.12352]

Forrest, A.R. (1971). Computational Geometry. *Proceedings of Royal Society London* 321(4), 187-195.

Goldberg, D.W. (2016). Geocoding. *The International Encyclopedia of Geography*, D. Richardson (eds. et al). John Wiley & Sons.
[https://doi.org/10.1002/9781118786352.wbieg1051]

Griffith, E.C. (1907). *The Rise and Development Of The Gerrymander*. Scott, Foresman and Company, Chicago, IL.

Grofman, B. and Cervas, J.R. (2018). Can State Courts Cure Partisan Gerrymandering: Lessons from League of Women Voters V. Commonwealth of Pennsylvania. *Election Law Journal: Rules, Politics, and Policy* 17(4), 264-285. Mary Ann Liebert, Inc., publishers.
[https://doi.org/10.1089/elj.2018.0496]

Ingraham, C. (2014). This Computer Programmer Solved Gerrymandering In His Spare Time. *The Washington Post* (June 3, 2014).
[https://www.washingtonpost.com/news/wonk/wp/2014/06/03/this-computer-programmer-solved-gerrymandering-in-his-spare-time/]

Krasno, J., Magelby, D.B., McDonald, M.D., Donahue, S. and Best, R.E. (2019). Can Gerrymandering Be Detected? An Examination of Wisconsin's State Assembly. American Politics Research 47(5), pp.1162-1201.
[https://doi.org/10.1177/1532673X18767890]

Levin, H.A. and Friedler, S.A. (2019). Automated Congressional Redistricting. *ACM Journal of Experimental Algorithms 24(1), Article 1.10*.
[https://doi.org/10.1145/3316513]

Li, Z., Wang, R-S. and Wang, Y. (2007). A Quadratic Programming Model For Political Districting Problem. *The First International Symposium on Optimization and Systems Biology* (OSB'07), Beijing, China.

Magleby, D.B. and Mosesson, D.B. (2018). A New Approach for Developing Neutral Redistricting Plans. Political Analysis 26(2), April 2018, pp.147-167.
[https://www.cambridge.org/core/journals/political-analysis/article/new-approach-for-developing-neutral-redistricting-plans/31F8EB3FFB7A8F5B3F7C2171BE016D47]

Mccarthy, D. (2012). How the Electoral College Became Winner-Take-All. Published online by FairVote, Takoma Park, MD.
[https://www.fairvote.org/how-the-electoral-college-became-winner-take-all]

Niemi, R.G., Grofman, B., Carlucci, C. and Hofeller, T. (1990). Measuring Compactness and The Role of a Compactness Standard in a Test For Partisan and Racial Gerrymandering. *The Journal of Politics* 52(4), 1155-1181. University of Chicago Press.
[https://doi.org/10.2307/2131686]

Preparata, F.P. and Shamos, M.I. (1985). *Computational Geometry – An Introduction*. Springer-Verlag. ISBN 0-387-96131-3.

Reitsma, F. (2013). Revisting The 'Is GIScience a science?' debate (or quite possibly scientific gerrymandering). *International Journal of Geographical Information Science* 27(2), 211-221.
[https://doi.org/10.1080/13658816.2012.674529]

Ricca, F., Scozzari, A. and Simeone, B. (2008). Weighted Voronoi Region Algorithms For Political Districting. *Mathematical and Computer Modeling* 48(9-10), 1468-1477.
[https://doi.org/10.1016/j.mcm.2008.05.041]

Siegel-Hawley, G. (2013). Educational Gerrymandering? Race and Attendance Boundaries in a Demographically Changing Suburb. *Harvard Educational Review* 83(4), 580-612.
[https://doi.org/10.17763/haer.83.4.k385375245677131]

United States Department of Justice (1965). History of federal Voting Rights Laws: The Voting Rights Act of 1965.
[http://justice.gov/crt/history-federal-voting-rights-laws]

Wu, P.Y. and Rathswohl, E.J. (2010). Address Matching: An Expert System and Decision Support Application for GIS. *Proceedings of Information Systems Education Conference* (ISECON 2010). ISSN: 1542-7382, #1339, Nashville, TN.

[http://proc.isecon.org/2010/pdf/1399.pdf]

Wu, P.Y., Deplato, J.P. and Combs, A.B. (2020). Geographic Information System and Gerrymandering. *Journal of Information Systems Applied Research* 13(3) pp 4-10. http://JISAR.org/2020-3/ ISSN: 1946-1836. [http://jisar.org/2020-13/n3/JISARv13n3p4.html]

Yamada, T. (2009). A Mini-Max Spanning Forest Approach To The Political Districting Problem. *International Journal of System Science* 40(5), 471-477. [https://doi.org/10.1080/00207720802645246]

Determinants of Health Professionals' Intention to Adopt Electronic Health Record Systems

Jie Du
dujie@gvsu.edu

Jenna Sturgill
sturgije@gvsu.edu

School of Computing
Grand Valley State University
Allendale, MI 49456

Abstract

The purpose of this study is to understand health professionals' perception and intention towards Electronic Health Record (EHR) systems and how those intentions play a vital role in improving the adoption of EHR systems. We proposed a research model based on the unified theory of acceptance and use of technology and health belief model to investigate the impact of specific factors on health professionals' intentions of using EHR systems. The results showed that trust is a significant influencing factor to the adoption and acceptance of EHR systems by health professionals. This study then recommended that further investigation into the barriers and drivers of EHR adoption should be done. By identifying and understanding the determinants of adopting EHRs, interventions and education can be designed to improve the adoption of EHRs.

Keywords: Electronic health record, health care, adoption, trust, and survey.

1. INTRODUCTION

Due to the increasing cost of health care, rise of chronic disease, and a projected 10% less amount of healthcare workers by 2025, Electronic Health Record (EHR) systems are becoming increasingly popular (Tavares & Oliveira, 2018). EHR is a repository of patient data in a digital form that includes data such as medical history, medication and allergies, immunization status, laboratory test results, radiology images, vital signs, personal statistics, and billing information, all stored and exchanged securely (Gunter & Terry, 2005). The combination of an EHR system and a patient portal, increases a patient's ability to carry out self-management activities, making the use of the health care system more effective and sustainable as the job market declines (Tavares & Oliveira, 2018). Although the adoption rate for EHRs has been increased in recent years, many challenges and barriers still exist. To improve the adoption of EHRs, understanding the factors that impact the adoption of EHRs is the first step.

The purpose of this study is to understand health professional's perception and intention towards EHRs and how those intentions play a vital role in improving the adoption and implementation of EHRs. Our research question is: What factors are the determinants for the health professionals to adopt and use EHRs? We proposed a research model by combining the unified theory of acceptance and use of technology (UTAUT2) and health belief model (HBM) to investigate the barriers and drivers for EHR adoption. An electronic questionnaire was developed to gather insight from health information management (HIM) professionals,

who manage EHRs throughout the hospital setting and college students majoring in HIM who are privileged to EHR access. The results show that trust plays a significant role in EHR adoption. By identifying and understanding the determinants of adopting EHRs, interventions and education can be designed to improve the adoption of EHRs.

The remainder of this paper is organized as follows. Section 2 provides a literature review of studies that investigated factors that impact EHR adoption. Section 3 introduces our research model and hypotheses. The methodology including survey development and data collection is presented in Section 4 and the results are presented in Section 5. Discussions on the results and implications are presented in Section 6. Section 7 concludes the paper.

2. LITERATURE REVIEW

The Technology Acceptance Model (TAM) (Davis, 1989), the Unified Theory of Acceptance and Use of Technology (UTAUT) (Venkatesh, Morris, Davis, & Davis, 2003), and extensions of these models have been used to determine users' acceptance or adoption of technology in various scopes. In this section, these models and their extensions applied to the health care field were first reviewed. Works that added a factor associated with privacy risk or trust to variations of these models were reviewed next. How our study extends the literature is presented at the end.

Technology Acceptance Model

Vitari and Ologeanu-Taddei (2018) used variables of TAM, perceived ease of use (PEOU) and perceived usefulness (PU) to measure the intention of different occupational groups in the same hospital setting, to use the EHR system. PEOU is defined as "the degree to which a person believes that using a particular system would be free from effort" (p. 1); PU is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" (p. 1). Vitari and Ologeanu-Taddei (2018) sought to clarify the possible differences, in intention to use an EHR and its antecedents, existing between the different staff categories. They administered a survey to measure the medical staff's perceptions of EHR, using questions derived from a review of previous studies: PU, PEOU, misfit, data security, anxiety, self-efficacy, and trust. Each variable was measured using one question and each question was answered using a seven-point Likert scale, with one indicating "strongly disagree" and seven indicating "strongly agree." They found that secretaries' and assistants' perception of the ease of use of EHR does not influence their intention to use it nor could they be influenced by self-efficacy in the development of their perception of the ease of use of EHR. This finding can be explained because secretaries and assistants are required to follow more stringent rules and procedures for their work, including working with EHR, with less professional autonomy than healthcare professionals.

Another study that utilized TAM was (Beglaryan, Petrosyan, & Bunker, 2017) study on hospital-based physicians' perspective on EHR. The main objective of their work was to understand the barriers of implementation from the point of view of end users; identify major determinants of physicians' technology acceptance; and develop a deeper understanding of the various factors impacting implementation through development of an enhanced TAM. TAM and its numerous extensions are often criticized by researchers for its incomplete scope. In particular it is argued that these models ignore: a) group and social processes related to IT implementation; b) technology's organizational and social consequences. TAM models are said to leave a gap between an individual's reactions towards technology and their intentions of using technology. Specifically, TAM does not account for the motivations of acting and for how different reasons for acting interact together to emerge as intentions. Beglaryan et al. (2017) explored the implementation barriers from the perspective of end users, with a particular emphasis on the acceptance and post-acceptance stages of the implementation. All items were measured using a five-point Likert-scale, ranging from "strongly agree" to "strongly disagree." Their results suggested that the major barriers of EHR acceptance among physicians include group level clinical concerns, impact on job performance, required effort to utilize the system, personal characteristic of innovativeness, interference with patient-provider relationships, and resistance to change. However, perceived ease of use did not cast a significant direct effect on behavioral intention, which is aligned with previous studies reporting that a PEOU-behavioral intention (BI) link is often found as the weakest correlation in the core TAM. They also found that the main direct determinant of behavioral intention is projected collective usefulness (PCU), and that PU transmits its effect to behavioral intention through PCU. A limitation of this model was

there might be discrepancies between intentions and actual behavior as pointed out by several other studies.

The Unified Theory of Acceptance and Use of Technology

The UTAUT model has played a critical role in evaluating technology intention and EHR acceptance. Alazzam et al. (2016) used UTAUT2 (Venkatesh, Thong, & Xu, 2012), an extension of UTAUT to explore the antecedent factors of medical staff intentions to use an EHR system by conducting a review of studies that use the UTAUT2 model and involve trust in stored data. The aim of their study was to compare and combine results from different studies using the UTAUT2 model, in the hopes of identifying patterns among the studied results. They anticipated habit will directly affect the intention of medical staff to use EHRs. Thus, a high level of intention to use is likely to increase employee adoption of EHRs. To detect a set of determinants of acceptance of EHRs by medical staff, Alazzam et al. (2016) created a research model based on UTAUT2 but added new constructs to measure the trust medical staff have in stored data. Alazzam et al. (2016) termed the added set of constructs "e-health extension to UTAUT2."

The Health Belief Model

The health belief model was created in 1950s and is a psychological model that attempts to explain health preventative behaviors (Rosenstock, 1974). This model suggests that an individual's behavior is determined by threat perception and evaluation of the behaviors to resolve the threat. The threat perception depends on vulnerability and severity, and evaluation of the threat is determined as perceived benefits minus perceived barriers. Three other variables included in HBM are self-efficacy, cues to action, and general health orientation (Rosenstock, Strecher, & Becker, 1988).

Ng et al. (2009) used the health belief model to study user's computer security behaviors. To understand how security awareness programs influence a user's attitude and behavior to be more security-conscious, Ng et al. (2009) examined the influences for a user to use computer security at their organization. By identifying the determinants of computer security behavior, interventions can be constructed to change the user's behavior. In the perspective of the HBM (Rosenstock, 1974; Rosenstock et al., 1988), an individual's behavior is determined by the threat perception and what it takes to resolve the threat. Ng et al. (2009) found that perceived susceptibility, perceived benefits, and self-efficacy were all impactful determinants of a user's computer security practices. Self-efficacy was important because computer users must be confident and able to perform the necessary mitigation measures and it was the most strongly related to intention and behavior. Perceived barriers, cues to action, general security orientation, and perceived severity were all found to not have statistical significance.

Ng et al. (2009) extended the health belief model to a new area of research to help determine how to change user's behaviors. This can be applied to not just computer security behaviors but also EHR behaviors. Sher et al. (2017) used the health belief model to examine perspectives of HIM professionals on privacy effectiveness in EHRs. Their study administered a cross-sectional survey to determine HIM professional's intention to protect EHR privacy. Survey items were measured on a seven-point Likert-scale ranging from "strongly disagree" to "strongly agree," with multiple questions for each construct. The results found that perceived susceptibility and perceived severity were weak predictors of preventative behavior, which is opposite of what the HBM argues. However, the constructs perceived benefits, perceived barriers, self-efficacy, and cues to action were found to be significant predictors of intention to protect EHR privacy, as the HBM proclaims. Sher et al. (2017) also emphasized the importance of organizations to communicate the benefits certain practices have on the use of EHRs.

The Combinations of Theories

Tavares and Oliveira (2018) used an integrated model approach to understand the factors that drive electronic health record adoption. They used the combination of UTAUT2, the health belief model (HBM), and the diffusion of innovation theory (DOI) for their research model. HBM constructs, perceived health risk, and self-perception, were used to replace UTAUT2 construct hedonic motivation to better predict motivation to use. Data was collected using a mobile phone survey resulting in the constructs compatibility, performance expectancy, and habit playing significant roles on the dependent variable intention to recommend. The combination of the three theories were found to be a successful model because they each had constructs with statistically significant impact on explaining the adoption of EHRs. Performance expectancy, due to its effect on behavioral intention, suggested

that individuals care about the results and advantages that EHRs can bring to manage health more effectively. However, the social influence hypothesis was not supported. Based on their results, and the high impact of performance expectancy, Tavares, and Oliveira (2018) emphasized the importance of communicating the advantages that EHRs provide to users.

Trust in User's Acceptance

Researchers had added a factor associated with privacy risk or trust to variations of TAM-based models (Jang & Lee, 2018; Palos-Sanchez, Hernandez-Mogollon, & Campon-Cerro, 2017) and the Unified Theory of Acceptance and Use of Technology (UTAUT) model (Yun, Han, & C. Lee, 2013; Zhou, 2013) to examine the usage intention for location-based service. In the field of health care, trust has also been added into UTAUT2 (Alazzam et al., 2016) and TAM (Vitari & Ologeanu-Taddei, 2018) to explore users' acceptance or adoption of EHRs.

Previous literature has supported that the combination of UTAUT2 and HBM is a successful model (Tavares & Oliveira, 2018) to understand the factors that drive EHR adoption. However, the role of trust and privacy in a combination of UTAUT2 and HBM has received little attention in research to date. This study proposes a research model that (1) combines UTAUT2 and HBM, and (2) incorporates trust and privacy as factors that impact users' adoption of EHRs.

3. RESEARCH MODEL

Our research model is built upon the UTAUT2 (Venkatesh et al., 2012) and the HBM (Rosenstock, 1974; Rosenstock et al., 1988). There are seven constructs in our research model (see Fig. 1). The six independent variables are perceived benefits (BEN) (HealthIT.gov, 2019), perceived barriers (BAR) (Ng et al., 2009; Stanford_Medicine, 2018), privacy (PRI) (Sher et al., 2017), social influence (INF) (Tavares & Oliveira, 2018), self-efficacy (EFF) (Ng et al., 2009; Sher et al., 2017), and trust (TRU) (Alazzam et al., 2016). The one dependent variable in the research model is the subjects' self-reported attitude toward EHR adoption (BEH) (Tavares & Oliveira, 2018). The six hypotheses are posited:

H1 – Perceived benefits (BEN) of using EHRs are positively related to EHR adoption intention.

H2 – Perceived barriers (BAR) to using EHRs are negatively related to EHR adoption intention.

H3 – Privacy issues (PRI) of using EHRs are negatively related to EHR adoption intention.

H4 – Social influence (INF) to using EHRs are positively related to EHR adoption intention.

H5 – Self-efficacy (EFF) to using EHRs are positively related to EHR adoption intention.

H6 – Trust (TRU) to EHRs is positively related to EHR adoption intention.

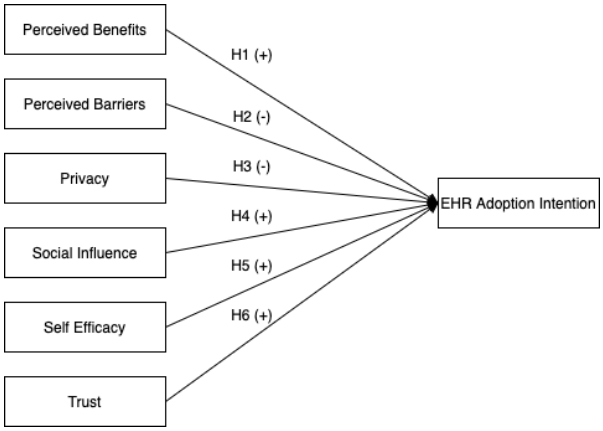


Fig. 1 Research Model

4. METHODOLOGY

Survey Development

An electronic survey was implemented to test the hypotheses. The survey questions were derived from (Alazzam et al., 2016; HealthIT.gov, 2019; Ng et al., 2009; Sher et al., 2017; Stanford_Medicine, 2018; Tavares & Oliveira, 2018). The survey questions are categorized into eight groups based on the constructs of our research model: demographics, perceived benefits, perceived barriers, privacy, social influence, self-efficacy, trust, and EHR adoption intentions. All items except the demographic items are scaled on a seven-point Likert scale: Strongly Disagree = 1, Somewhat Disagree = 2, Disagree = 3, Neutral = 4, Agree = 5, Somewhat Agree = 6, and Strongly Agree = 7.

Survey participants were health information management professionals. The survey was administered using the Qualtrics online survey platform. The survey consisted of seven demographic questions and 33 EHR questions with a target completion time of less than 15

minutes. All of study participants were informed about the research purpose, confidentiality protection, and the anonymity of the information collected, and each signed a consent form before participating.

Data Processing

A total of 51 responses were received, over a three-week period. After removing the 11 records of missing values, the data collection yielded 40 usable survey response sets. The table below summarizes the demographics of the sample.

Demographic	Category	Percentage
Age	Under 20 years old	0
	20-29 years old	12.5
	30-39 years old	25
	40-49 years old	30
	50-59 years old	20
	60 years or older	12.5
Gender	Male	30
	Female	70
Education	High school	5
	Some college	12.5
	Career training	7.5
	2-year degree	5
	4-year degree	45
	Master degree	17.5
	Professional degree	2.5
	Doctorate	5
Average experience		5.6 years

Table 1. Demographics of Participants

Data Analysis Steps

We conducted a two-step analysis to examine the effects of the key constructs on the EHR adoption intention dependent variable. First, an exploratory factor analysis (EFA) was done to extract the factors (latent variables) to validate our model constructs. Second, a multiple regression analysis was conducted using the SPSS calculated factor scores. The dependent variable was regressed on the six IVs to determine the main effects.

5. RESULTS

Construct Validity and Reliability

We conducted the factor analysis (using primary axis analysis) on the data set to extract the factors that influence HIM professionals' attitude toward EHR adoption. We use 0.5 as the recommended threshold (Hair, Anderson, Tatham, & Black, 1998). Five rounds were run before we arrived at a set of factors loading at 0.5 or above (BAR5 was removed in Round 4 due to unexpected loading on the TRU construct). Eight items (EFF2, EFF3, BAR7, BAR3, PRI2, BAR6, BAR10, and BAR5) having a factor loading lower than 0.5 were removed from further consideration.

The results of EFA resulted in eight factors being extracted from the data: TRU, BEN, INF, BAR_1, BEH, PRI, BAR_2, and EFF. Note that the BAR resulted in the splitting of the original BAR construct into two factors: BAR_1 and BAR_2. This unexpected result will be addressed in the discussions section later in this paper.

Cronbach Alpha coefficient was used to test the reliability of the items. The acceptable value of Cronbach Alpha should be at least 0.70 (Nunnally & Bernstein, 1994). However, for exploratory studies, a minimum Cronbach Alpha value of 0.5 is allowable (Hinton, McMurray, & Brownlow, 2004). Table 2 summarizes the factor loadings and Cronbach Alpha values for each item. The factor loadings for all items are greater than 0.5 and the Cronbach Alpha values for all factors are greater than 0.7 except BAR_2 with a .546. The Cronbach Alpha in BAR_2 is weak but allowable given the low number of questions (two questions) in that construct. Therefore, the factors loadings and the Cronbach Alpha coefficients show construct validity and reliability, allowing us to proceed with our regression analysis and hypothesis testing.

Construct	Item	Factor loadings	Cronbach Alpha
TRU			0.938
	TRU1	.575	
	TRU2	.800	
	TRU3	1.055	
	TRU4	.886	
	TRU5	.618	
	TRU6	.710	
	TRU7	.962	
	TRU8	.744	
BEN			0.885
	BEN1	.819	
	BEN2	.989	
	BEN3	.863	
	BEN4	.526	
INF			0.883
	INF1	.673	
	INF2	.950	
	INF3	1.011	
BAR_1			0.845
	BAR1	.781	
	BAR2	.879	
	BAR4	.831	
BEH			0.927
	BEH1	.828	
	BEH2	1.050	
	BEH3	.763	
PRI			0.844
	PRI1	.858	
	PRI3	.745	
	PRI4	.829	
BAR_2			0.546
	BAR8	.567	
	BAR9	.741	
EFF	EFF1	.739	

Table 2. Factor Loadings and Cronbach Alpha

Hypothesis Testing

To test the hypotheses, a multiple regression analysis was conducted using SPSS. The latent variable, trust has a significant coefficient as expected (p = 0.008). Thus, H6 was supported. Other variables were not significant. Therefore H1 - H5 were not supported.

Variables	Coefficient
TRU	.451**
BEN	.204
INF	.186
BAR_1	-.041
PRI	-.071
BAR_2	.208
EFF	.096
R^2	.450
Adjusted R^2	.330

Table 3: Regression Results

6. DISCUSSION

Discussion of Results

The results of this study show that trust is a significant determinant of the attitude toward adoption of EHR. Health care professionals who trust EHR systems have a more positive attitude toward adoption of EHR. Our findings indicate that perceived benefits, perceived barriers, social influence, privacy, and self-efficacy are not significant. Self-efficacy has been reported as a significant determinant in EHR adoption (Sher et al., 2017). However, in our study, only one question about self-efficacy was used in our data analysis, which might not be adequate to measure the respondents' self-efficacy. Social influence was not found as a significant determinant to EHR adoption in the previous literature (Tavares & Oliveira, 2018). Since all of our survey respondents have an average of five years in the health care field and 70% of them have a 4-year degree or a higher degree, they might not find many barriers in using EHR systems. Given the survey respondents' experience of using EHR, they might not likely be influenced by the other people regarding adopting an EHR system.

During the EFA, we found that one perceived barrier question (BAR5) was loaded in the construct of trust. BAR5 says physicians use other means as work arounds for EHR, which indicates certain barriers of using EHRs. In reality, physicians might choose using other means as work arounds for EHR due to personal preference, time limit, or other considerations. Also, BAR5 seems more concrete and observable than the other more abstract BAR questions. This might explain why BAR5 was not loaded in

the construct of perceived barriers. Some trust questions seem observable too. For instance, TRU3 (EHRs provides verification of user identity) is concrete and observable. This may help explain why BAR5 was loaded as trust.

The EFA analysis also resulted in the splitting of the perceived barrier factor (BAR) into two separate factors (BAR_1 and BAR_2). A simple look at the questions gives insight into why this may have been necessary (Table 4). The first three items (BAR1, BAR2, and BAR4) all highlight the physician role and look at the perceived barriers from the physician's perspective. These three questions were loaded as BAR_1 (barriers perceived by physicians). The other two items (BAR8 and BAR9) emphasize the barriers as time consuming or considerable investment of effort other than time. These two questions were loaded as BAR_2 (barriers perceived by general health care professionals). Given these differences, it seems at least logical that the perceived barriers factor needs to be split. The question as to how people perceive the barriers of using EHR is one that should be explored in the future.

Item	Question
BAR1	Using an EHR has increased the total number of hours physicians work on a daily basis.
BAR2	Using an EHR detracts from physicians' professional satisfaction.
BAR4	EHRs contribute greatly to physician burnout.
BAR8	Using an EHR is time-consuming.
BAR9	Using an EHR would require considerable investment of effort other than time.

Table 4. Questions of Perceived Barriers

Implications for Research and Practice

There are at least two implications of these findings for the research community. First, trust is a significant determinant to adoption of EHR. The results suggest that the more trust the users have on the EHR systems, the more likely they will adopt EHR. The trust can be built in the forms of EHR capturing, storing, and transferring patient medical records properly. Ensuring that adequate security mechanisms are put in place is an effective way to build trust in health care professionals when considering adopting EHRs. Second, we found that trust is the only significant determinant to adoption of EHR. The limited number of significant factors in the model could be an indicator that better models are needed.

Implications for practice focus around informing health care professionals about the security mechanisms implemented in EHRs so that they can trust the system and be more willing to adopt it. References and tutorials that explain how the patient medical data will be handled in the EHR will help the adoption of EHR. One of the best indicators for adoption of EHR is the decision maker's trust of the technology.

Limitations

There are two limitations that must be acknowledged regarding this research. One limitation of the research is that this was a small sample size. Future research could replicate this study using a larger sample size. Another limitation is that some questions of self-efficacy and perceived barriers were removed due to low factor loadings. This might indicate that future work is required to explore these concepts. For example, the perceived barriers could be measured based on the role of a health care professional in a health care setting. Self-efficacy was measured using one question in our study. There may be other self-efficacy questions, which might be significant in motivating a health care professional to adopt EHRs.

7. CONCLUSIONS

To better understand health care professionals' intention to adopt EHR systems, a survey was developed based on the unified theory of acceptance and use of technology model and the health belief model. The survey of seven demographics questions and 33 EHR questions, expecting to take less than 15 minutes, was administered, to 51 health information management professionals. After removing 11 records with missing values, 40 were considered in the results. The results showed that trust is a significant determinant of the attitude toward adoption of EHR. Perceived benefits, perceived barriers, social influence, privacy, and self-efficacy did not have significant impacts on the health care professionals' attitudes towards EHR. Questions on perceived barriers and self-efficacy should both be explored more extensively in the future. With the continued rise in use of EHR systems in the hospitals, this study hopes to help EHR developers and policy makers to better understand the motives and perspectives that

will affect the successfulness of a health care professional to adopt an EHR system.

8. REFERENCES

Alazzam, M., Basari, A. S., Shibghatullah, A., Ibrahim, Y., Ramli, M., & Naim, M. H. (2016). Trust in stored data in EHRs acceptance of medical staff: Using UTAUT2. 11, 2737-2748.

Beglaryan, M., Petrosyan, V., & Bunker, E. (2017). Development of a tripolar model of technology acceptance: Hospital-based physicians' perspective on EHR. International Journal of Medical Informatics, 102, 50-61. doi: 10.1016/j.ijmedinf.2017.02.013

Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly, 13(3), 319-340. doi: 10.2307/249008

Gunter, T. D., & Terry, N. P. (2005). The emergence of national electronic health record architectures in the United States and Australia: models, costs, and questions. Journal of medical Internet research, 7(1), e3-e3. doi: 10.2196/jmir.7.1.e3

Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C. (1998). Multivariate Data Analysis (5th ed.). Upper Saddle River, NJ: Prentice Hall, Inc.

HealthIT.gov. (2019). What are the advantages of electronic health records? Retrieved September 29, 2020, from https://www.healthit.gov/faq/what-are-advantages-electronic-health-records

Hinton, P. R., McMurray, I., & Brownlow, C. (2004). SPSS Explained: Routledge.

Jang, S. H., & Lee, C. W. (2018). The Impact of Location-Based Service Factors on Usage Intentions for Technology Acceptance: The Moderating Effect of Innovativeness. Sustainability, 10(6), 1876.

Ng, B.-Y., Kankanhalli, A., & Xu, Y. (2009). Studying users' computer security behavior: A health belief perspective. Decision Support Systems, 46(4), 815-825. doi: https://doi.org/10.1016/j.dss.2008.11.010

Nunnally, J., & Bernstein, I. H. (1994). Psychometric Theory. New York: McGraw-Hill Higher Ed. .

Palos-Sanchez, P. R., Hernandez-Mogollon, J. M., & Campon-Cerro, A. M. (2017). The Behavioral Response to Location Based Services: An Examination of the Influence of Social and Environmental Benefits, and Privacy. Sustainability, 9(11), 1988.

Rosenstock, I. M. (1974). The Health Belief Model and Preventive Health Behavior. Health Education Monographs, 2(4), 354-386. doi: 10.1177/109019817400200405

Rosenstock, I. M., Strecher, V. J., & Becker, M. H. (1988). Social Learning Theory and the Health Belief Model. Health Education Quarterly, 15(2), 175-183. doi: 10.1177/109019818801500203

Sher, M.-L., Talley, P. C., Cheng, T.-J., & Kuo, K.-M. (2017). How can hospitals better protect the privacy of electronic medical records? Perspectives from staff members of health information management departments. Health Information Management Journal, 46(2), 87-95. doi: 10.1177/1833358316671264

Stanford_Medicine. (2018). How Doctors Feel About Electronic Health Records Retrieved September 29, 2020, from https://med.stanford.edu/content/dam/sm/ehr/documents/EHR-Poll-Presentation.pdf

Tavares, J., & Oliveira, T. (2018). New Integrated Model Approach to Understand the Factors That Drive Electronic Health Record Portal Adoption: Cross-Sectional National Survey. J Med Internet Res, 20(11), e11032. doi: 10.2196/11032

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. MIS Quarterly, 27(3), 425-478. doi: 10.2307/30036540

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. MIS Quarterly, 36(1), 157-178. doi: 10.2307/41410412

Vitari, C., & Ologeanu-Taddei, R. (2018). The intention to use an electronic health record and its antecedents among three different categories of clinical staff. BMC Health Services Research, 18(1), 194. doi: 10.1186/s12913-018-3022-0

Yun, H., Han, D., & C. Lee, C. (2013). Understanding the use of location-based service applications: Do privacy concerns matter? (Vol. 14).

Zhou, T. (2013). An empirical examination of user adoption of location-based services. Electronic Commerce Research, 13(1), 25-39. doi: 10.1007/s10660-013-9106-3

Appendix - Survey Questions

The survey questions are categorized into eight groups based on the constructs of our research model: demographics, perceived benefits, perceived barriers, privacy, social influence, self-efficacy, trust, and behavior intentions.

All items except the demographic items are scaled on a seven-point Likert scale: Strongly Disagree = 1, Somewhat Disagree = 2, Disagree = 3, Neutral = 4, Agree = 5, Somewhat Agree = 6, and Strongly Agree = 7.

Demographics (DEM):
DEM1: Age verification - I verify that I am at least 18 years old (yes/no)
DEM2: What is your age? (<20, 20-29, 30-39 , 40-49, 50-59, 60+)
DEM3: What is your gender? (Female, male)
DEM4: What is your highest level of education? (Less than high school diploma, high school or equivalent, some college, career training, 2-year degree, 4-year degree, master's degree, doctorate degree, professional degree)
DEM5: What is your title in your current position?
DEM6: How many years of experience do you have in your field? (0, <1 year, 1-2 years, 3-4 years, 5+ years)
DEM7: Do you have experience with Electronic Health Record (EHR)? (yes/no)

Perceived benefits (BEN):
BEN1: Using an EHR improves the quality of health care I provide to my patients.
BEN2: Using an EHR improves the communications between my patients and me.
BEN3: Using an EHR fosters my patient engagement in their care.
BEN4: Using an EHR reduces medical errors for my patients.

Perceived barriers (BAR):
BAR1: Using an EHR has increased the total number of hours physicians work on a daily basis.
BAR2: Using an EHR detracts from physicians' professional satisfaction.
BAR3: Using an EHR detracts from physicians' clinical effectiveness.
BAR4: EHRs contribute greatly to physician burnout.
BAR5: Physicians often use other means (paper notes, scanning, faxing, etc.) as work arounds for EHR.

BAR6: There are more challenges to using EHRs than benefits.
BAR7: Using an EHR is inconvenient.
BAR8: Using an EHR is time-consuming.
BAR9: Using an EHR would require considerable investment of effort other than time.
BAR10: Using an EHR would require changing work habits, which is difficult.

Privacy (PRI):
PRI1: The chance that EHR privacy may be breached is high.
PRI2: There is a strong probability that EHR privacy breaches may lead to privacy issues.
PRI3: The use of EHR is likely to cause privacy issues.
PRI4: I am concerned for the privacy of my patient's personal information during data transmission among different EHR's.

Social influence (INF):
INF1: Most people who influence me think that electronic health records are helpful.
INF2: Most people who are important to me would use electronic health records.
INF3: Most people who are important to me believe that it is good to use electronic health records.

Self-efficacy (EFF):
EFF1: I am confident that I could complete a task using an EHR if I had seen someone else use it before trying it myself.
EFF2: I am confident that I could complete a task using an EHR if I could call someone for help if I got stuck.
EFF3: I am confident that I could complete a task using an EHR even if there was no one around to help me.

Trust (TRU):
TRU1: EHR are predictable and consistent regarding the usage of the information.
TRU2: EHR are honest with patients when it comes to using personal health information provided.
TRU3: EHRs provides verification of user identity.
TRU4: EHRs provide the actual identity of the user as claimed.
TRU5: EHRs provide authorization to access control of stored data according to the entity's privileges/rights of use.
TRU6: EHRs ensure the confidentiality of information accessibility.
TRU7: EHRs ensures that the data collected will be solely used for the intended purpose.
TRU8: EHRs ensures that stored data are protected from unauthorized manipulation/alteration.

Behavior intention (BEH):
BEH1: I intend to use EHRs.
BEN2: I intend to use EHRs in the next months.
BEN3: I plan to use EHRs frequently.