**Tools to Enhance Internal Cybersecurity Awareness**

A Technical Report submitted to the Department of Computer Science

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

**Andrew Li**

Spring, 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Rosanne Vrugtman, Department of Computer Science

# Tools to Enhance Internal Cybersecurity Awareness

CS 4991 Capstone Report, 2023

Andrew Li
Computer Science
University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia, USA
al3qst@virginia.edu

## ABSTRACT

As data breaches and hacks become increasingly prevalent, enhancing cybersecurity awareness and defenses have become paramount for many companies. This is especially salient when a significant number of security breaches are the result of human error, often from an internal source. A comprehensive cybersecurity strategy requires many moving parts—including cybersecurity awareness among all employees. In light of its catastrophic data breach in 2019, Capital One's cyber department decided to develop an internal tool providing associates throughout the company with relevant and up-to-date information on cybersecurity events and threats. As a software engineering intern in summer of 2022, my project at Capital One aimed to augment existing infrastructure to provide useful tools and information to Capital One's employees in all roles. To build this project, I worked collaboratively with a number of other interns as well as full-time associates in my team in an Agile environment. In addition, I learned several new technologies to conform with Capital One's dynamic technology stack. The project achieved its goal by adding a new feature that allowed Capital One employees to quickly search and filter through a wealth of information about cybersecurity news, events, and knowledge. By improving overall access to important information, the feature will be a key part of bolstering Capital One's cybersecurity resilience.

## 1. INTRODUCTION

In July of 2019, Capital One suffered a devastating cyber breach that affected over 100 million Americans (Capital One, 2022). The breach was estimated to cost Capital One over $500 million even before harms to reputation, lost business, and regulatory backlash or punishment were considered (Lu, 2019). In February of 2022, a $190 million dollar settlement was reached after a class action lawsuit was filed against the company (Capital One, 2023). Capital One's case is only one of an increasing number of costly cyber breaches happening worldwide which is spurring growing investments in corporate cybersecurity.

Each cyber attack is different, making a comprehensive strategy critical. One major factor that often leads to the greatest vulnerabilities is curtailing sources of human error, especially from employees who are less knowledgeable about cybersecurity. Americans as a whole are concerned about cybersecurity but are overwhelmingly apathetic when it comes to their own actions (Pew Research, 2017). Thus, a cybersecurity strategy principally needs to promote ease of access to relevant cybersecurity knowledge that can benefit employees in different roles. This motivates the need for a simple but universal tool that can be used by all employees in an organization, regardless of their level of prior knowledge.

## 2. RELATED WORKS

The process of eliminating data fragmentation, called data unification, is a well-studied topic. Gibbs, et al. (2002) note that data fragmentation itself can cause security vulnerabilities due to the proliferation of vulnerable endpoints from which data can be accessed. In addition, data fragmentation also makes it difficult for organizations to remain in compliance with legislation in the US and Europe designed to protect consumer privacy. Data fragmentation also diminishes an organization's ability to detect and react to data breaches, potentially making the breaches much more severe (Threat Connect, n.d.)

The importance of human error in causing cybersecurity breaches is well-documented (Coffey, 2017). Because of this, cybersecurity strategies must put primary focus on people and coordinating human and digital responsibilities. Human-centered systems must focus not only on the capabilities of the technology itself, but also how different users will use and interpret the technology differently. Understanding this complexity is a key requirement for successful cybersecurity technology.

## 3. PROJECT DESIGN

Information and data are the most powerful tools for both those defending against and perpetrating hacks. Access to useful information can encourage best practices to prevent human sources of cyber vulnerability from weak passwords to social engineering. Data itself is also a critical resource, as effective monitoring can alert companies to hacks as they progress or even before they succeed. Combined with increasing regulatory scrutiny that requires detailed reporting after a cyber breach, many companies face challenges in maintaining and extracting useful data from their databases (Novaes Neto, et al., 2020).

As databases become increasingly large, complex, and distributed, the problem of data fragmentation arises. Data fragmentation occurs when data is stored in very disparate systems, called data silos, and disconnected from other data sources due to inconsistent storage formats and access methods. This data needs to be unified to allow companies to more efficiently and effectively extract value from the massive amounts of data they do collect, especially in contexts where that data can prevent or mitigate costly cyber-attacks. This issue can also be solved with a tool that can unify data and provide a powerful interface to access, filter, and sort that data.

### 3.1 Project Overview

This project, dubbed Universal Search, aimed to provide a powerful and user-friendly webapp for Capital One's diverse employees and departments to disseminate and access cybersecurity-related information and data. Universal Search unified a wealth of disparate data from various sources into a single portal accessible in Capital One's internal network and provides the infrastructure and extensibility to adapt to new requirements and use cases. In addition, users could use a sophisticated search tool to filter, identify, and share information that is useful to them. Advanced users could also access powerful tools to filter data with greater granularity. Strict access control systems were also in place to ensure that no user could access data beyond their privileges. The end goal of Universal Search was to increase accessibility and convenience for Capital One employees to get the information they need to stay up-to-date and prepared for cyber threats against the company.

### 3.2 Requirements

Universal Search is a complex tool that must serve the needs of a diverse group of clients. First, Universal Search must service existing administrators in Capital One's cybersecurity department which includes users on both the consumption and production ends of cybersecurity data. This includes associates

who digest a myriad of cybersecurity information directly relevant to their work tasks as well as associates who create and publish knowledge bases for other associates to consume. These users generally have some elevated privileges to access sensitive information and are very familiar with the variety of cybersecurity data Capital One collects. For these existing users, Universal Search must provide enough advanced functionality to augment the tools and processes they already use to access and publish information. This means advanced filtering options that allow these employees to quickly sift through a massive amount of data to find the information they are looking for. Additionally, their privileged access to certain information that is dependent on their specific role must be maintained and remain consistent with existing systems of access management. To enable this, Universal Search must implement a login system to verify the user's identity and privileges.

Universal Search must also expand access and accessibility of cybersecurity data of diverse types to Capital One employees who do not work directly with cybersecurity issues. These users may have been reluctant to seek out this information in the past due to ambivalence about its relevance to their work or because of barriers to access caused by complexity and data fragmentation. For these users, Universal Search must provide a user-friendly and intuitive design to increase adoption. One requirement was that Universal Search must be implemented as a widget on top of existing websites in Capital One's intranet, in addition to having its own, dedicated page. Across all of the web pages where Universal Search is implemented, it must maintain a consistent appearance and location on the page. Strict access controls must also be implemented for these users to ensure they cannot access confidential data beyond their privileges.

One important feature clients wanted was the ability to share search results with other employees by copying the URL of the search. Thus, each search using Universal Search must be exactly replicable using information stored in the URL. This would be a key feature in aiding the dissemination and sharing of data throughout Capital One.

Universal Search must also be built around existing infrastructure, meaning it must adhere to a specific technology stack. In this case, Universal Search was built using the MERN stack, composed of a backend database using MongoDB, routing using Express.js, frontend design in React, and a server environment using Node.js.

### 3.1 Key Components

The foundational component of Universal Search is its ability to unify data of different formats and types into one portal. Each data source could vary wildly in how it is stored and what types of values it can store. Universal Search needs to be able to process and understand each data source under a few common metrics, such as when each data record was created, or which fields in the record should be text-searchable. We solved this by developing an extensible repository of configuration files that store basic information about each data source. Because we were working in a MongoDB database, each configuration file just needed to hold the field name for a number of useful attributes such as title, date created, and category. Using these configuration files, a user can, for instance, filter data from many different data sources by date, despite the different ways in which each data source may store date and time information.

It was also important for us to make these configuration files transparent and easy to modify or extend to future-proof as data sources are added, removed, and modified. To do this, we made each configuration file independent of each other and wrote them in

TypeScript to maintain consistency and compatibility with the rest of the webapp.

Using these configuration files to unify the data also allowed us to implement filtering. On each metric for which we designed filtration capabilities, such as by keyword, date, and category, we stored the relevant field names inside the configuration file for that data source. On the backend, this allowed us to use simple regular expression queries using MongoDB to quickly search through every support data source at once and present them to the user in a consistent format.

Another requirement was that searches be reproducible using the search URL. This meant that any search terms and filters needed to be stored in a compact text format in the URL. The server must also be able to restore a search using only the information contained in the URL. This proved to be a challenge because filters could be complex and hierarchical, rapidly expanding the length of the URL even for relatively simple searches. To mitigate these issues, we compressed the search string as much as possible while still leaving it in a human-readable format.

Finally, we also implemented strict access controls to ensure data confidentiality. This was also achieved using the configuration files by recording the necessary user permissions to access each data source. These permissions would be automatically generated based on the user's role in the organization and could be easily modified by administrators if needed. A challenge we faced in access controls was ensuring that not only were unauthorized users not able to view restricted information, but they would also not know the information existed in the first place. Thus, we developed flexibility to hide certain filters and results based on the user's access level, rather than just prevent access.

## 4. RESULTS

Because the systems we worked with involved sensitive information, we extensively tested the system as we developed and were bound by a minimum requirement of 80% line coverage for tests. This ensured not only that there were no bugs that could cause a crash or break the user interface, but also ensured that access controls were being enforced correctly.

Still, further testing and code review were necessary before the changes were deployed officially. However, all of the main functionality specified in the requirements was successfully implemented in the short duration of the internship. Universal Search is a promising example of how data unification can help employees in all roles access and understand cybersecurity information and how it applies to their organization.

## 5. CONCLUSION

The proliferation of cyber attacks and costly data breaches in the past decade has motivated a comprehensive approach to prevention and defense. These increasingly complex systems are no longer limited to just cyber departments or technical employees. Cybersecurity touches upon every aspect of a business, and lack of a comprehensive strategy makes every aspect of a business vulnerable. Lack of knowledge about cybersecurity by non-technical employees has been extensively identified as a key point of weakness for organizations. Ameliorating this potential issue was the goal of my summer project. Universal Search was a powerful yet extensible tool that targeted one important aspect of cybersecurity—employee awareness. Universal Search is a useful addition to the arsenal of tools Capital One employs to prevent future data breaches and was also a valuable learning experience for me and other interns.\

## 6. FUTURE WORK

Future work to incrementally improve a project is always warranted. Because our project was under a tight timeframe, we were unable to unify every available data source under Universal Search. Future work would

expand the number and types of data available in the system. Because we had anticipated the need for future additions on this front, we designed the configuration file system to be as extensible and transparent as possible so that adding new data sources could be done relatively easily and quickly.

Additional filtering capabilities are also a task for future development. As we designed filtering to be as modular as possible, it is relatively easy to add new filtering features, such as filtering by the creator's name. In light of all these opportunities for improvement, our hope is that the project has developed a solid foundation that facilitates development in the long-term.

## 7. ACKNOWLEDGMENTS

I would like to acknowledge my full-time coworkers and managers at Capital One, without whom the project would not have been a success. I thank Matt Anderson, Dominic Ritchey, Luke Ordille, Juan Suhr, Bryan Benzinger, Stephen Tewes, Daniel Hazeley, and Danielle Horton. Additionally, I thank my fellow interns Lauren Brown, Bikram Kohli, Grace Liu, and Albert van Valkenberg, for their contributions that made the project a success.

## REFERENCES

Capital One. 2022. Information on the Capital One cyber incident. https://www.capitalone.com/digital/facts2019/

Capital One. 2023. Capital One Data Breach Class Action Settlement. https://www.capitalonesettlement.com/en

John W Coffey. 2017. Ameliorating Sources of Human Error in CyberSecurity: Technological and Human-Centered Approaches. In *The 8th International Multi-Conference on Complexity, Informatics and Cybernetics*. Pensacola.

Martin R. Gibbs, Graeme Shanks, and Reeva Lederman. 2002. Data Quality, Database Fragmentation and Information Privacy. *Surveillance & Society* 3, 1 (Sep 2002). https://doi.org/10.24908/ss.v3i1.3319

Jack Lu. 2019. Assessing The Cost, Legal Fallout Of Capital One Data Breach. (Aug 2019). https://doi.org/10.2139/ssrn.3438816

Nelson Novaes Neto, Stuart Madnick, Anchises Moraes G. de Paula, and Natasha Malara Borges. 2020. A Case Study of the Capital One Data Breach. (Mar 2020). https://doi.org/10.2139/ssrn.3570138

Pew Research. 2017. *Americans and Cybersecurity*. Pew Research Center.

Threat Connect. [n. d.]. *Fragmentation: The "Silent Killer" of Your Security Management Program*. ThreatConnect. https://threatconnect.com/wp-content/uploads/ThreatConnect-whitepaper-fragmentation.pdf