Visualization of Malware across Business Sectors

Emily Longman, Will Sims, and Taylor Kirkpatrick

Abstract— Businesses are targeted by cyber attacks every day, and there is no efficient way to visualize the industries that are most at risk and the types of attacks they are vulnerable to. Our research focuses on constructing a useful visualization for security professionals so they know what industries are most in need of their services. Currently, there are no interactive visualizations that solve this specific security problem. However, we built on previous research that looked at different problems in security visualization. The benefits of this visualization is to provide a tool that allows users to easily discover sectors most at risk so we can work towards keeping peoples information safe.

Index Terms—Security, malware

1 Introduction

Cybersecurity is a serious concern, or at least should be, for almost every industry. In the modern world relies heavily on computers of all forms, for basic tasks like checking the weather to heavy industrial computation. The data stored and processed by these systems can be very valuable and targeted for theft. Because of this there has been an ever-growing list of malware used to break into computer systems and gain whatever the attacker was looking for. Today, with nearly 40 years of enterprise computer usage, there is a lot of data available on the exploits, but it can be very technical and hard to sift through. With information visualization we can do a better job on analyzing these trends, which allows the creation of better defenses against future attacks.

2 Related Work

We looked at research in the field of cybersecurity and information visualization to see any problems people encountered and how we can avoid those challenges when creating our own visualization. We were also interested in seeing if there has been any visualizations on the same topic and if we could build on that work for our own project.

2.1 Challenges in Visualization for Cyber Security

Gates and Engle looked at why cyber security visualization has not been more effective in the past, how visualization can be utilized in cyber security, and how to evaluate cyber security visualization [1]. One of their main points is that you should not create a visualization for the sake of visualization and you should always have a clear problem that you would like to solve. Aesthetics should always come second to displaying information clearly and concisely. They also emphasize how the process should always be evaluated and case studies are sufficient for valuable feedback instead of large user studies. Best, Endert and Kidwell looked at seven key challenges when creating visualizations for cyber security [2]. The challenges that they encountered when developing a network security visualization that are most related to our problem had to do with their data sources. One of the issues in cyber security visualization is that there is so much data that only a small subset of data is compared when analyzing an event. This can be problematic because important information can be left out. Additionally, The diversity and quantity of security information provides challenges and it is important that users take into account the scope and validity of their data.

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

2.2 Evaluation Methods for Cyber Security Visualizations

It is critical to evaluate visualizations in order to determine whether or not the product is usable and solves the problem that is addressed. Langton and Baker explored various evaluation methods specifically in the domain of security and provided recommendations based on their test results [3]. It was found that design heuristics and user studies are effective for creating information visualizations in the field of cyber security. The methodology discussed for measuring the complexity of computer security visualization designs was developed by Suo, Zhu, and Owen [SOURCE] and they tested this method by evaluating TNV and RUMINT cyber visualization tools. Metrics such as color, shape and size were used to evaluate the complexity of the visualization. The user testing consisted of using common metrics such as task success, time on task, and errors.

2.3 Design in Information Visualization

According to Moere and Purchase, good design is a fine balance between utility, soundness, and attractiveness [SOURCE]. The requirement of attractiveness was investigated in and a model of the potential roles of design in information visualization was created. The model is made up of three distinct categories of visualization: practice, studies, and exploration. This model was based off the interaction design triangle developed by Fallman [SOURCE]. It was also recommended that pastel colors are used in visualizations to add less eye strain on the user. The journal article Fluid interaction for information visualization explored different types of interactions to consider when designing a visualization and created a set of guidelines for creating fluid interactions[SOURCE]. Some of the main guidelines consisted of using providing immediate visual feedback on interaction, minimize indirection in the interface, integrate user interface components in the visual representation. The guidelines closely followed Nielsens 10 usability heuristics [SOURCE]. Lastly, The design of our visualization was influenced by the stock market visualizations that organized all of the data by industry sector[SOURCE].

3 Method

3.1 Method Overview

3.2 Visualization Tasks

The question we are trying to answer is what industries are most vulnerable to cyber attacks and what types of attacks are most prevalent. This problem is important because it provides a way for cybersecurity companies to see which industries are most in need of their services. It also allows industries that are most affected by cyber attacks to see which types of threats they are most vulnerable to.

3.3 Data Sources

There are plenty of security companies that disclose information about cyber attacks. Popular cyber security companies such as Symantec, McAfee, Malwarebytes publish monthly and yearly threat reports about different kinds of attacks. Symantec organizes the monthly threat report into five different categories: Malware, Web Attacks, Spam, Phishing,

Mobile, and Social Media. They also give information about the number of threats per industry [6]. Malwarebytes publishes a state of malware report every year which contains information on the most common types of attacks with a focus on ransomware trends in particular. They give trends of different threats over the last three years, which provides a more thorough data set. [4] In the McAfee report, the volume of malicious samples cataloged per quarter ebbs and flows quarterly and annually. This data showed a decline during the past three quarters which mirrored the trend we observed at the start of 2015. A pattern of two to three quarters of growth followed by three quarters of decline has been consistent since 2013. [5]

3.4 Design Comparison

There were many iterations of design ideas that went into this visualization page, ranging from grand scheme design choices to small implementation tweaks. This basic framework stayed throughout the process, but the design of it evolved over time. A scroll bar for different months of data was one of the big things that was added, as well as a pie chart breaking down the data within each category upon clicking on one. Each of the industries has their own section, and these all contain data about which types of attacks are most prevalent.

A number of arrangement ideas for the data were compared, with the look eventually settling on one of differently sized and colored blocks. These help to display the distribution difference both with color and size, so that someone with some form of visual impairment is likely to still be able to tell the difference. There was much discussion over the use of a red to green scale, since a number of colorblind people may be affected, but it made the most sense for our data, and the size difference help to provide a backup for this.

4 Enhancements

5 Implementation

We will implement the ER diagram by putting the data from Symantec into a mySQL database that is hosted by OSU using PHP. We will have two tables, one that contains industries and the total number of attacks, along with threat IDs that correspond to a type of threat, the number of attacks, and the date associated with the report. We can compare this data by looking at different sectors and how they vary by the attacks that are most common.

6 Results and Performance

7 Applications Paper

8 Conclusions and Future Work

This topic has a wide variety of data sources which can be combined and compared in interesting ways to draw a variety of conclusions. With database manipulations and interesting front-end visualization this raw data can be made into something very interesting. The better the visualization, the better users can understand the information and gain valuable security knowledge.

References

[1] S. E. Carrie Gates. Reflecting on visualization for cyber security, 2013.

[2] D. K. Daniel M. Best, Alex Endert. 7 key challenges for visualization in cyber network defense, 2014.

[3] A. B. John T. Langton. Information visualization metrics and methods for cyber security evaluation, 2013.

[4] Malwarebytes. State of malware 2017.

[5] McAfee. Threat reports.

[6] Symantec. Symantec monthly report.