Visualization of Malware across Business Sectors

Emily Longman, Will Sims, and Taylor Kirkpatrick

Index Terms—Security, malware

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.

1 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.

2 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. [3] 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. [1] 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. [2]

3 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.

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

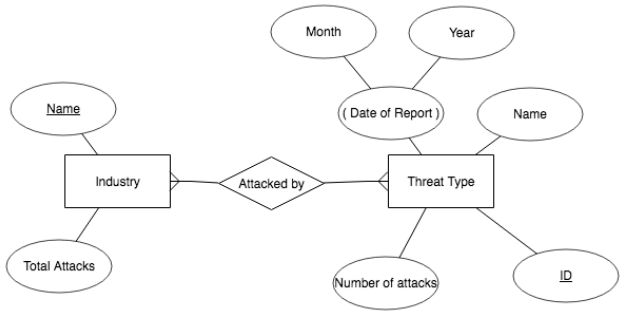

Fig. 1. Layout of how different sources are related

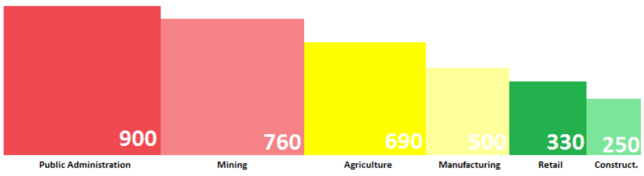

Fig. 2. Representation of attacks within different industries, utilizing color and size to represent the data visually

4 ER Diagram

The diagram for this project can be seen in Figure 1.

5 Design Ideas

This is the main part of the design, which can be seen in Figure 2. Each square represents an industry, labeled below it. The squares get larger and shift from green to yellow to red the more prevalent cyber attacks are in an industry. At a glance it is easy to tell which industry has the biggest problem with cyber attacks. Each box is labeled below, and inside the box is the total number of attacks per month (this is not real data). Something we put a lot of thought into while designing is the possibility of colorblind users, as they are quite common. In this part of the design, and in the rest, it should be noted that the design does not rely on color, and has an alternative way of conveying this information.

When hovering over any box with ones mouse, say, the Agriculture box: You can see in Figure 3 that a pie chart has appeared above the square, with slices colored and sized based on cyber attack type. This shows the user a breakdown of not only what industry is subjected to attacks, but the distribution of attacks as well. You might note the small size of the light green slice, and how there is only a number, not a label inside it. The is because the name cannot fit in the slice. The last part of the design can be seen by hovering over a slice to learn more about it.

For example, hovering over the small green slice: This tab in figure 4 shows the detailed information we have about the attack, such as its name, the id we assign it, the occurrences per month (displayed on the slice) the total occurrences, and recent reports of the attack. This is a bit of a wall of text, but there isnt much more we can do there, as this is

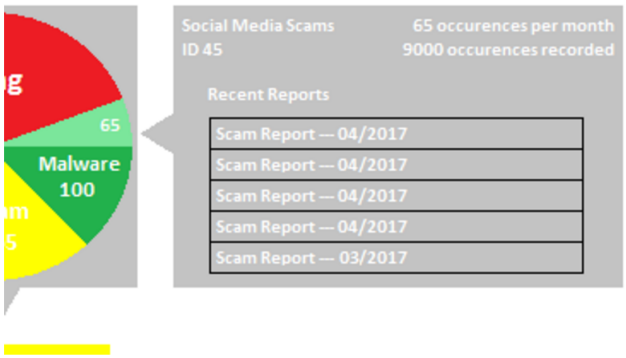

Fig. 3. Provides a breakdown of the prevalence of specific attack types within an industry

detailed information that does not translate as well into a fancy graphic. This tab always opens to the right, to avoid running into taller industry squares.

6 Conclusion

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] Malwarebytes. State of malware 2017.
[2] McAfee. Threat reports.
[3] Symantec. Symantec monthly report.

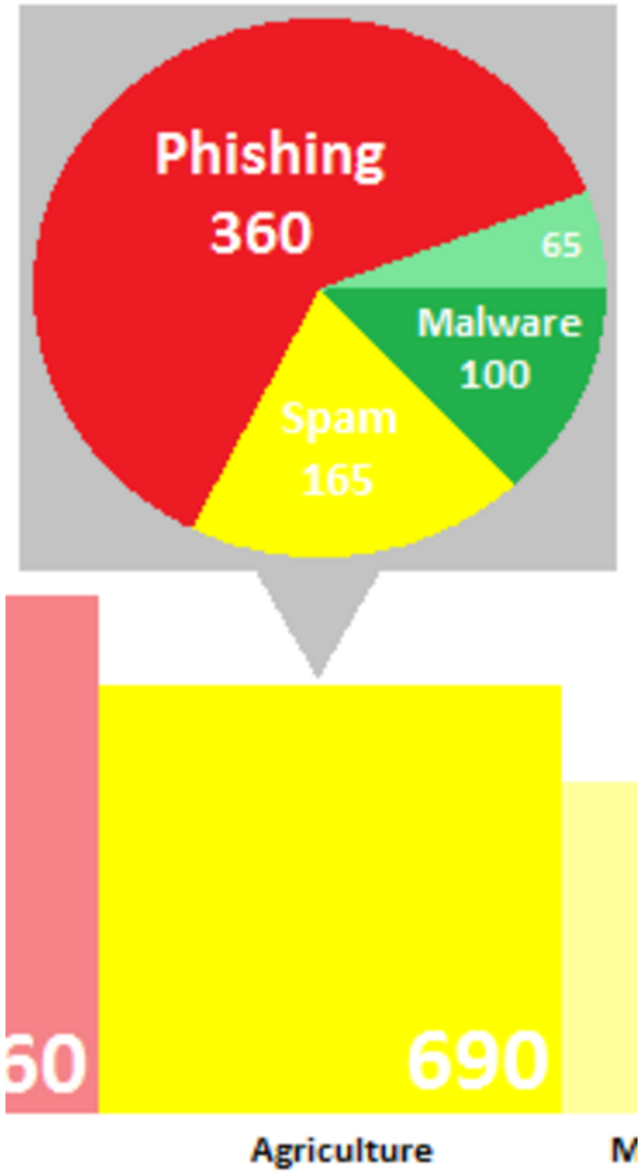

Fig. 4. A further breakdown, providing source details for each attack type