Davit Gabrielyan

LMC 4702

Final Thesis

Faculty Member 1: Sauvik Das

Faculty Member 2: John Stasko

TABLE OF CONTENTS

ABSTRACT

In this paper, my team endeavored to assist end-users in their cybersecurity struggles by constructing a security assistant. The security assistant would unobtrusively detect vulnerabilities on the end-users machines and notify the users in the most effective manner to ensure that they take the steps to mitigate the underlying causes of the vulnerability. The detection of the vulnerabilities is performed using machine learning and detection algorithms.

To ensure the effectiveness of our assistant we plan to conduct a rigorous study in order to assess the extent to which our security assistant is capable of encouraging end-users to have better cybersecurity practices. After obtaining the results of our study, we plan to further improve the capabilities of our security assistant so that the notifications it displays are more effective and informative. Ultimately, we strive to find techniques that would allow end-users to interactively communicate with our assistant about their security needs and have a safe virtual environment on their machines.

INTRODUCTION

The eminence of technology has undoubtedly skyrocketed over the past few decades. With that, the momentousness of cybersecurity has surged as well. We are entering an era when people start to worry about their cybersecurity to a very large degree [1]. Additionally, an increasing number of end-users are facing a problem when they are worried about their cybersecurity but do not necessarily have the expertise to protect themselves from attackers and conquer the sophisticated malware that can unobtrusively be installed on their machine [18]. For instance, over 70% of the U.S. population reports that they constantly fear that their credit card information and bank credentials will be compromised [1]. Likewise, over 65% of the U.S. population is anxious about their national identity being stolen and admits to not having the tools to alleviate the risks associated with protecting it [1, 18].

In the future, with ubiquitous, smart, connected devices this problem will only be exacerbated as the potential attack vectors attackers might use will increase in lockstep. With over 76% of security breaches being financially motivated [2] and an average data breach costing as much as $7.91 million [3], it appears to be obvious that immediate steps should be taken to address the current disregard for cybersecurity. If we are to prevent further attacks, we will need to find a way to motivate users to detect and act on their security and privacy vulnerabilities. Additionally, we will need to assist them in the most interactive, human-like manner.

More and more creative minds in the scientific community turn their attention towards cybersecurity and join the collective battle against hackers. As a result, there has been a lot of research done in the field of cybersecurity that mainly focuses on finding strategies to mitigate the vulnerabilities in the operating systems that get discovered by attackers [4]. However, after the formula to combat the vulnerability is found, there is no guarantee that a user is going to integrate it into their machine [5]. Particularly, this is evident from zero-day vulnerabilities. A zero-day vulnerability is a particular type of a susceptibility when attackers act fast on a recent, publicly released vulnerability and exploit unsuspecting users who are yet to integrate a patch onto their system (hence the name, as the attack is performed in zero days). Especially for end-users who are not aware of the severity of the vulnerabilities, it is common to neglect security updates and postpone their installation of the program patches which exposes them to critical risks.

Furthermore, there has not been substantial research done to assess the extent to which average users are aware of the dynamically progressing cybersecurity world. There is no evidence to conclude that the average end-user comprehends what is going on behind the scenes of any security patches or software updates. This raises the question of whether an average user can be more resilient to cyber-attacks if they comprehend the severity of the currently existing exploits and vulnerabilities.

Therefore, our team looked at different approaches to create a security assistant that would help the user in their security struggles. We looked at different paths that we could take to achieve a secure environment for the user. Our definition of a secure environment is an environment where the user can freely go about their daily tasks in the cyberspace without worrying about their information being compromised or their machine becoming infected. Since that would require the security assistant to access some personal and private information of the user, we aim to achieve our goals without sacrificing the user’s trust. In order to achieve that, we strived to make our steps, actions and findings transparent to the user by educating them about potential vulnerabilities and what the security assistant is actually accessing in order to keep the user's machine secure. For that purpose, we looked into strategies to make our user interface as informative as possible while keeping it concise. We predict that such an approach can effectively educate the user and make them more aware of the severity of the situation without overwhelming them with technical details. Simultaneously, we predict that this will reduce the probability of the user ignoring the security notifications.

Additionally, we aimed to create an interactive experience for the user when dealing with the assistant. We investigated different AI techniques to achieve this, however, most of the work is still in progress. We are currently investigating deep learning models that would effectively deliver the information through our user interface to the user without interfering with their daily tasks. Ultimately, we aim to conduct a comprehensive user study with a working prototype of our security assistant and assess the extent to which it is actually able to aid end-users in their cybersecurity practices. Our user study will be constructed around a certain cybersecurity scenario or a few scenarios that users find most critical. This will be identified based on a survey that will be conducted both in the USA and Russia to ensure diversity in the collected data sample. Our main goal will be to investigate the extent to which our assistant helps protect the user and how effective our assistant’s notifications are.

LITERATURE REVIEW

The field of cybersecurity has recently seen a steep increase in the number of attacks that have malicious motives to exploit people and compromise their information

[6]. It is reported that in 2018 alone the volume of cybercrime increased by over 40% [6]. Every hour cybercriminals devise more and more sophisticated strategies to bypass the current cybersecurity measures by misleading the current antivirus software in end user's machines. The newly devised attacks go far beyond what experts expected to deal with. For instance, a pixel-based attack software has recently been launched that virtually has no countermeasures [7]. As a result, no user or corporation can protect themselves when they are infected with the software since this type of attacks are novel to the cybersecurity community.

Given these dire circumstances, a lot of users have a desperate need to protect themselves. Notwithstanding, they do not necessarily have the expertise to combat the infinitely big spectrum of cyberspace where an exploit can be waiting at any corner and can even be installed unobtrusively on an unsuspecting user's machine [8]. Therefore, there has been a lot of research done to help mitigate these ramifications [9, 10]. Despite that, the most novel attacks seem to fly under the radar of the current detection software, moreover, after a solution was found to a specific cyber threat, there is no guarantee that users comprehend the consequences and are going to manually patch the vulnerabilities on their machines. Therefore, it is crucial that we evaluate the risks and make them transparent to the user. As discussed previously, we genuinely believe that increasing user awareness can significantly benefit their overall cybersecurity behavior. Additionally, it is vital for us to assist the end-user in mitigating those risks.

Our team, deeply inspired by the creators of virtual assistants like Almond [11], decided to upgrade cybersecurity to a whole new level and create an interactive security assistant for an average user. Almond is a natural language processing-based security assistant designed specifically for the Internet of Things (IoT) and online services [11]. It largely uses Android-based frameworks in its architecture. Furthermore, it incorporates natural language processing and is claimed to be the only virtual assistant that lets users specify trigger-action commands in a natural language. While our goals differ from Almond's and we aim to cover a much broader spectrum of cybersecurity, this is a very good reference point for us to monitor our progress and draw extra inspiration if needed.

The implementation process of Almond revealed to our team the extent to which an interactive security agent can actually communicate with the user. Deliberating on the results we saw in the research papers surrounding Almond and other assistants, we have made the decision to implement a chatbot interface that will use natural language processing to interact with the user as it was proven to be the most convenient through extensive testing by researchers [12]. By the same token, there has been an extension proposed about including a voice-based assistant to interact with the user on a more personal level. For this case scenario, we need to carefully weigh the risks of such an implementation because as we can see in various research papers, there are novel risks associated with voice assistants [13].

Lastly, we have observed that there have been few studies that researched the extent to which users actually utilize the security assistants available to them. Therefore, we plan to conduct a full, controlled experiment to recognize the degree to which users use/comply with the directions of our security assistant and to what extent they comprehend the assistant's suggestions [14]. In order to conduct our study in the most beneficial manner, we investigated some research papers to understand the conventions of conducting a user study [15, 16]. For example, we came across a research project where a survey style analysis was proven to be very helpful [16]. Survey-style assessments help to draw the conclusion that security, privacy, usability, and reputation have positive effects on consumer trust on a web site in the online banking context and that trust has a positive effect on consumer commitment [16, 17]. One significant limitation, however, is how researchers collected their data. Their data set was collected only from English speaking customers, hence, the findings cannot be generalized to other language speaking customers [16].

We have conducted our own survey to find out what scares users the most in the field of cyberattacks. In order to incorporate diversity into our survey, we conducted this study in two different countries (USA and Russia). We have found that the most feared scenario was the thievery of bank credentials and credit card information. Therefore, using the findings from Dr. Casaló's research [16], and with a similar approach, we can consolidate our argument about the imminent need for measures to protect users when they are on online banking websites. For that purpose, we will proceed to build countermeasure software to protect the user's credit card information and can use these tools in our user study to assess how effective and convenient they are for the user.

METHODOLOGY

Overview

The most important security scenario was chosen to be the credit card information scenario. In this scenario, we aimed to protect the users from giving out their credit card information while on an unsecured network. Therefore, our user study will be constructed around that scenario with the utilization of the architecture we have already implemented. Our main goal will be to investigate the extent to which our security assistant helps protect the user and how effective its notifications are.

Materials

We concluded that our security assistant, which we named Penny, is going to need a browser extension to unobtrusively detect when users are about to insert their credit card numbers. The detection is performed using a rigorous algorithm involving pattern matching and pixel analysis. To develop the algorithm, we leveraged some traits that we observed in regular pages that require credit card information. The main details of the algorithm have to stay confidential in order to preserve the integrity of the detection process. Once Penny detects that the user is on an unsecured network and is about to insert sensitive information, it needs to pass the information down to the operating system in order for the user to be notified. It is crucial to do this step securely because the user is connected to an unsecured network and we cannot use it to transmit any alerts or warnings. To accomplish this, we decided to have a local server communicate with the extension. A local server is one that provides a service by running an application that is on the same machine as the client application, hence, not utilizing any external networks. This would mean that we do not require to transmit data over potentially insecure networks that can temper with our alerts and warnings.

Lastly, to interact with the user, we implemented a UI (User Interface) to be able to visually connect with the user. We chose a desktop application for this purpose to deliver the alerts in the most appropriate manner directly to the user. We believe that visually connecting with the user is absolutely critical to cultivate their understanding on the subject and make them aware of the vital security flaws that put them at risk. To help better illustrate how the Penny architecture works, we created a simple diagram to illustrate the coordination between the separate components of Penny (figure 1).

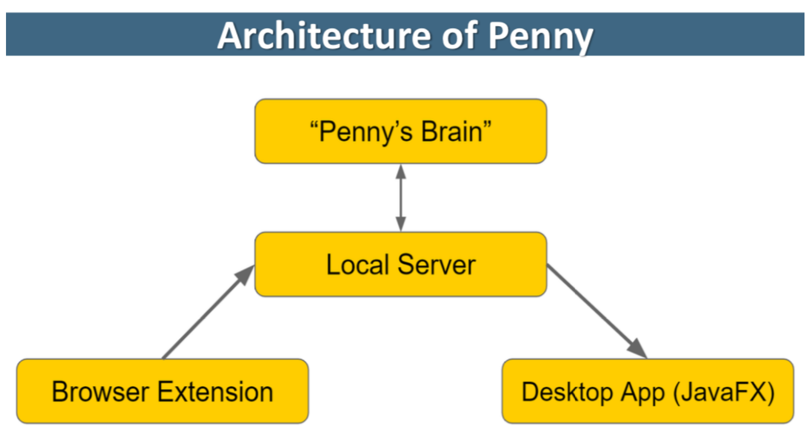


Figure 1

Participants

Once the architecture is completed, we will move on to conduct our rigorous research study with various participants that will be interacting with Penny. For this purpose, we plan to recruit 5 - 7 participants who will be assigned to a specific machine in the lab that is equipped with Penny software and will be browsing the web for 1 - 2 hours twice a week. We will conduct a survey before confirming participants to ensure demographic diversity in our participant sample as we believe it is a crucial step to understanding the common user behavior in cyberspace.

During the study information on how many times warnings from Penny were ignored or were effective in preventing users from entering their sensitive information will be collected. This data will be collected using a program that will be installed on the machine alongside Penny. The program will increment predefined variables whenever a notification is ignored, dismissed or acted upon. After the study, the compiled data will be analyzed and observations will be made by our team. Combining all the observations, a conclusion will be deduced regarding the effectiveness of Penny and the data will be used to support our claims. We believe that regardless of the outcome of the study, it will be a vital piece of evidence to suggest whether further research and development needs to be done in this area or should other approaches be considered to ensure user security.

DISCUSSION AND CONCLUSION

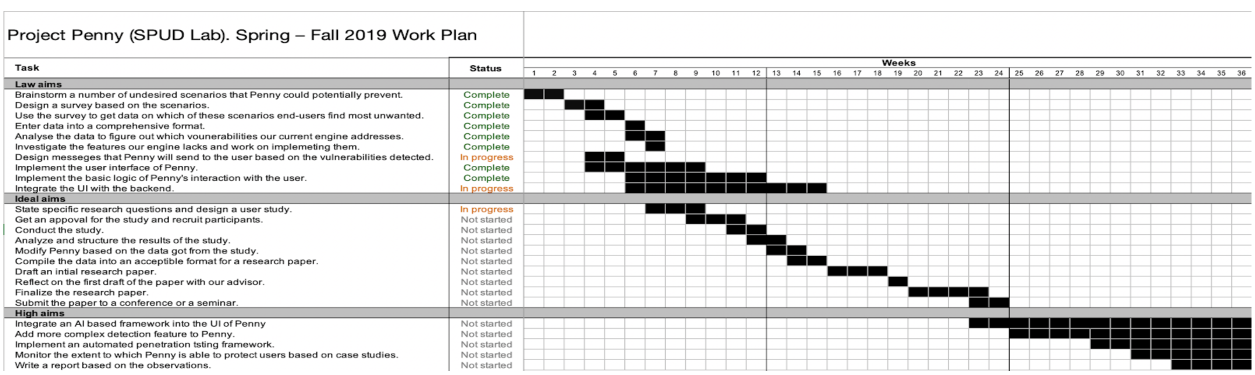
Our main goals when starting this research were to analyze the extent to which end-users are aware of the critical risks that exist in cyberspace and the strategies we can use to educate them and motivate them to alter their cybersecurity behaviors for the better. Ultimately, we created a security assistant that unobtrusively detects vulnerabilities on the user's machine and brings them to their attention which allows us to statistically analyze the frequency with which users will neglect the notifications or take action. We predict that after completing our user study we will be able to answer the most vital question of whether our security assistant helps improve the cybersecurity behavior of end-users. Additionally, the data we collect will open up endless horizons for us to take Penny to the next level.

While researching about the cybersecurity behaviors of the population and the ways we can develop Penny to better assist the average user, we always had the objective to make cybersecurity accessible for the general population. In addition to helping individual end-users, we can further develop Penny to act as an assistant on various networks and systems where cybersecurity might not be of primary concern and hence can be neglected. We envisioned that the eventual applications of Penny can be truly worthwhile both inside and outside of the tech world. Within the tech world, Penny can be used as an additive framework to assist or work alongside already existing security software. Outside of the tech world, Penny can be implemented into the systems of high schools, hospitals and non-profits. It can be tremendously helpful for these particular types of communities as they are usually very tight on the availability of finances and cybersecurity might not be the immediate area of investment. Additionally, once on the networks of those organizations, Penny can collect various data about the particular types of attacks that are more prominent for the systems and networks in those communities. The collection of such data can aid the development of more robust techniques to prevent more complex cyber-attacks.

As for the further research directions, Penny can be utilized and expanded to answer much more sophisticated questions that arise within the cybersecurity community. For instance, it would be helpful to investigate the particular times of the day when users are more likely to ignore the security notifications. A model can then be constructed which would predict the probability of the user not neglecting a security notification at a particular time. This can potentially help maximize the effectiveness of the notifications as those will be sent out at times with the highest probabilities derived from the model with the least chance of outer interference.

In conclusion, we had initially discovered that there is a huge information gap between certain end-users and cybersecurity and strived to assist these users in their struggles in our research. We constructed a security assistant that utilized some of the most advanced machine learning and detection algorithms to unobtrusively identify vulnerabilities and exploits on the end-users' machines. Moreover, we explored various techniques of how to notify the user in the most convenient and effective manner and developed our user interface to achieve those techniques. While we are still in the process of conducting our user study to assess the extent to which our security assistant is capable of aiding end-users, we sincerely believe that it has the potential to radically redefine cybersecurity practices of average end-users. After collecting and meticulously analyzing the results from our study, we will be able to move forward and improve our security assistant to better serve the end-users. Particularly, we will be focusing on aspects of our security assistant that ensure a more secure environment for the user while simultaneously encouraging better cybersecurity behavior and awareness.

WORK PLAN



REFERENCES

1. Cybercrimes Remain Most Worrisome To Americans Gallup, Inc - https://news.gallup.com/poll/244676/cybercrimes-remain-worrisome-americans.aspx
2. 76% Of It Security Breaches Are Motivated By Money First Louis Columbus - https://www.forbes.com/sites/louiscolumbus/2018/05/15/76-of-it-security-breaches-are-motivated-by-money-first/#429cda32199e
3. The Average Cost Of a Data Breach 2018 Sydny ShepardJul 17 - https://securitytoday.com/articles/2018/07/17/the-average-cost-of-a-data-breach.aspx
4. Yue, On-Ching. "Cyber security." *Technology in Society* 25.4 (2003): 565-569.
5. Fagan, M., Khan, M. M. H., & Buck, R. (2015). A study of users' experiences and beliefs about software update messages. Computers in Human Behavior, 51, 504-519.
6. Parker, R. D., & Farkas, C. (2011). Modeling Estimated Risk for Cyber Attacks: Merging Public Health and Cyber Security. Information Assurance and Security Letters, 2, 32-36.
7. Dixon, M., & Fogarty, J. (2010, April). Prefab: implementing advanced behaviors using pixel-based reverse engineering of interface structure. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1525-1534). ACM.

8. Thompson, H. H., Whittaker, J. A., & Mottay, F. E. (2002, March). Software security vulnerability testing in hostile environments. In Proceedings of the 2002 ACM symposium on Applied computing (pp. 260-264). ACM.
9. Cranor, L. F., Reagle, J., & Ackerman, M. S. (2000). Beyond concern: Understanding net users' attitudes about online privacy. The Internet upheaval: raising questions, seeking answers in communications policy, 47-70.
10. Jøsang, A., Zomai, M. A., & Suriadi, S. (2007, January). Usability and privacy in identity management architectures. In Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68 (pp. 143-152). Australian Computer Society, Inc..
11. Campagna, G., Ramesh, R., Xu, S., Fischer, M., & Lam, M. S. (2017, April). Almond: The architecture of an open, crowdsourced, privacy-preserving, programmable virtual assistant. In *Proceedings of the 26th International Conference on World Wide Web* (pp. 341-350). International World Wide Web Conferences Steering Committee.
12. Jendricke, U., & tom Markotten, D. G. (2000, December). Usability meets security-the Identity-Manager as your personal security assistant for the Internet. In acsac (p. 344). IEEE.
13. Diao, W., Liu, X., Zhou, Z., & Zhang, K. (2014, November). Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices (pp. 63-74). ACM.

14. Felt, A. P., Ha, E., Egelman, S., Haney, A., Chin, E., & Wagner, D. (2012, July). Android permissions: User attention, comprehension, and behavior. In Proceedings of the eighth symposium on usable privacy and security (p. 3). ACM.
15. Yee, K. P. (2004). Aligning security and usability. IEEE Security & Privacy, 2(5), 48-55.
16. Casaló, L. V., Flavián, C., & Guinalíu, M. (2007). The role of security, privacy, usability and reputation in the development of online banking. Online Information Review, 31(5), 583-603.
17. Das, S., Kramer, A. D., Dabbish, L. A., & Hong, J. I. (2014, November). Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security* (pp. 739-749). ACM.
18. Balkin, J., Grimmelmann, J., Katz, E., Kozlovski, N., Wagman, S., & Zarsky, T. (Eds.). (2007). Cybercrime: digital cops in a networked environment (Vol. 4). NYU Press.