Behavioral-Based Malware Detection in Cybersecurity

(Technical Paper)

Examining the Ethical Implications of Behavioral Detection on Society

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree
Bachelor of Science, School of Engineering

Benny Bigler-Wang

Fall 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

Signature Benny Bigler-Wang Date 11/12/2023
Benny Bigler-Wang

Approved ____ Date __________

Capstone/Technical Advisor Name, Department of Choose Department

Approved ____ Date __________

STS Advisor: Richard D. Jacques, Ph.D., Department of Engineering & Society

Introduction

Behavioral-based malware detection is an emerging approach in the field of cybersecurity research and development that focuses on analyzing user behavior to identify and prevent malware attacks. With the increase in sophisticated cyber threats, this form of detection holds significant importance in safeguarding individuals, organizations, and society as a whole from potential harm. It is estimated that cybersecurity crime cost the global economy a total of $7 trillion in the year 2022 and that the average cost of a cyber breach was $4.35 million (Aag IT Support [2023], n.d.). This not only indicates the severity that cyber-attacks have nowadays but also the increase in the frequency of these attacks. These attacks not only result in financial losses but also have far-reaching consequences, including the compromise of sensitive information, disruption of services, and reputational damage to companies.

During the summer of 2023, I interned at a Linux cybersecurity startup based in Charlottesville called Vali Cyber which focuses on providing Linux security tools that utilize state-of-the-art behavioral and AI/ML techniques to detect and stop threats. A major part of the tool that they are developing is based on behavioral detection, a cybersecurity technique that monitors patterns in software and network traffic to flag suspicious activity. This detection method is relatively new and is quite different from the traditional signature-based detection that most cybersecurity companies implement. There are major upsides to using behavioral-based detection such as increased adaptability to new threats, earlier detection times, and dynamic analysis. However, there also are a couple of downsides such as it being more resource-intensive and having a higher false positive rate. This was my first introduction to behavioral-based detection, and I found it to be a very intriguing approach to identifying new malware strains. However, the implementation of behavioral-based malware detection raises several ethical concerns and challenges that need to be addressed. These include a potential invasion of privacy due to the collection and analysis of use behavior data, concerns regarding data security and misuse, and the need to balance cybersecurity needs with individual privacy rights.

The technical topic outlined in this prospectus aims to outline the benefits and drawbacks of behavioral-based malware detection and furthermore provide potential solutions to these drawbacks. The STS research topic will explore the privacy concerns and potential biases that may arise when implementing such detection systems. By examining these issues, I aim to contribute to the responsible and ethical development of cybersecurity practices.

Technical Topic

Behavioral detection was a concept that was emphasized at Vali Cyber as a better alternative to the outdated detection methods that are prevalent in the cybersecurity industry. Unlike the traditional forms of malware detection that rely on signature-based or heuristic-based methods, behavioral-based detection focuses on identifying unusual patterns or deviations in user behavior that may indicate the presence of malware. Security companies are slowly adopting behavioral detection techniques while maintaining their signature-based defense systems because both methods have their own strengths and weaknesses. As a result, in order to create a comprehensive defense strategy, signature-based detection is most effective right now when used in conjunction with other dynamic and behavior-based security measures. New research in the past couple of years shows several use cases of researchers developing novel state-of-the-art security techniques that leverage emerging technologies such as artificial intelligence and machine learning to mitigate some of the downfalls of behavioral-based detection. I believe conducting further research in this area is essential to ensure the continuous improvement and effectiveness of defensive security measures in the face of evolving cyber-attacks.

Both signature-based malware detection and behavioral-based malware detection are two key approaches in the realm of cybersecurity. Signature-based malware detection relies on known patterns or bit sequences of previous malicious code to identify and block threats. Its strengths lie in its effectiveness against recognized malware, providing a quick and reliable method to detect and neutralize known threats. This approach is resource-efficient and generally has a low rate of false positives, as the signatures are specific to known malicious entities. However, signature-based detection has notable weaknesses. It is ineffective against new or evolving malware strains, commonly referred to as zero-day threats, as it relies on predefined signatures. Attackers can easily evade detection by modifying the code or using obfuscation techniques (Scott, J. [2017] n.d.). Additionally, the increasing use of polymorphic and fileless malware poses challenges, as these variants change their signatures dynamically. Behavioral-based malware detection is a proactive approach that focuses on identifying malicious activities based on the behavior of software or processes rather than relying on known signatures. Its strengths lie in its ability to detect previously unknown or zero-day threats by analyzing patterns of behavior that deviate from normal system activity. This approach is particularly effective against polymorphic malware and other sophisticated threats that may change their characteristics to evade traditional detection methods. Behavioral analysis can offer insights into the tactics and techniques employed by malware, aiding in early detection and response. However, this method is not without its weaknesses. False positives can be a challenge, as legitimate software or user behavior might trigger alerts if it deviates from established baselines. Behavioral-based detection can also be resource-intensive, especially when conducting real-time analysis on a large scale. Additionally, sophisticated malware may employ new evasion techniques to hide malicious behavior, posing a constant challenge to behavioral-based detection systems. As shown in (L. Caviglione et al. IEEE 2023), advanced threats will deploy anti-emulation techniques to change their behavior based on whether or not they are running in virtual or real environments. This leaves machines vulnerable to attacks if their behavioral-based detection system runs in a sandbox environment.

At Vali Cyber, I got to evaluate their existing behavioral detection engine through offensive penetration testing. This consisted of evaluating several distinct types of malware or attack vectors on a virtual machine that had its Linux security agent, ZeroLock, installed on it. If I was successful in an attack on the machine, I would suggest certain modifications to the existing rule set that the engine used. Behavioral detection techniques such as these are a relatively new concept that is usually paired with machine learning as previously mentioned to identify deviations from normal activity on a computer. I got to look under the hood of one of the machine learning models that Vali Cyber developed to detect cryptojacking attacks specifically. Cryptojacking is the unauthorized use of other people's devices to mine for cryptocurrency, essentially stealing other people's processing power. They created a feature set of the distribution of bitwise operators used in the assembly code of certain cryptojacking malware strains and then trained an SVM model to recognize the pattern behind these specific attempts at cryptojacking. After training the model, they were able to achieve a 99.98% accuracy rate at detecting cryptojacking on future testing data. This application of machine learning for behavioral detection was extremely effective because crytpojacking produces a very unique distribution of bitwise operators, specifically a lot of ANDs in the assembly code. Figuring out other similar, niche cyber-attacking techniques that also have distinct behaviors or patterns in their assembly code might lead to a very effective, novel behavioral detection system and is something that I would like to do further research on.

Combining recent developments in security research with updated machine learning techniques and computational resources to help combat these weaknesses will allow for a more comprehensive, effective behavioral-based detection system. My proposed methodology will involve a combination of data collection, analysis techniques, and evaluation processes. To collect relevant data, a variety of sources will be utilized, including real-world malware samples, simulated environments, and network traffic logs. This diverse range of data will provide a comprehensive understanding of behavioral patterns exhibited by different types of malware. The analysis techniques will involve leveraging machine learning algorithms, specifically deep learning models and SVMs, to identify and classify behavioral characteristics of malware (Gopinath M., Sibi Chakkaravarthy Sethuraman 2023). Machine learning techniques that are utilized for monitoring network traffic that could be related to ransomware activities can be identified during file reading and overwriting which does not require large computational power (Eduardo Berrueta et al. 2022). Training a model to detect specific malware behavior will be evaluated by its false positive rate which will determine the usability of the model because too many false positives make the security tool unusable in a real environment. Additionally, the effectiveness of behavioral-based malware detection will be evaluated through comparison studies with traditional signature-based methods, statistical analysis, and benchmarking against known malware datasets. I would like to test several different datasets that correlate to several different unique topics of malware such as ransomware, trojans, keyloggers, cryptojacking, etc.

STS Topic

The rapid acceleration of technological advancements in recent years has transformed how we interact with the world and will only continue to grow. As IoT devices, cloud computing, and other emerging technologies expand the surface area of technologies needing defensive security measures, the need for adaptive methods will become greater (Zheng et. al. 2022). This digital revolution, however, has resulted in an escalating amount of cybersecurity challenges that have made it difficult for the security industry to keep up with. Implementing new defense techniques such as behavioral-based malware detection will probably not significantly impact people socially or economically, but it is critical that developers in the future do their due diligence to consider outside factors and control algorithmic bias in their implementation of such security systems. One of the potential concerns is the potential invasion of privacy that comes with collecting and analyzing user behavior data. By monitoring and analyzing user activities, behavioral-based detection systems may have access to highly sensitive information, raising questions regarding the extent of data collection, storage, and retention. Additionally, there are concerns about the security of this user information and the potential for its misuse or unauthorized access. Balancing the need for robust cybersecurity measures with the protection of individual privacy rights is another ethical consideration that requires careful thought and consideration. Behavioral detection works around an understanding of malicious behavior. To learn this behavior, security analysts and developers manually tune this understanding or use machine learning to derive an interpretable model. This can introduce algorithmic bias, which stems from cultural, social, or demographic factors of how a user behaves, in their detection method as it is not entirely objective as to what is considered authorized or normal behavior. From an STS perspective, signature-based detection has less potential for negative societal impact as it simply utilizes a matching algorithm to search for known malware strains. As soon as we introduce more applications of machine learning and algorithms there is more potential for algorithmic bias to creep into our security systems. No research that I could find has been conducted to consider the ethical implications and societal ramifications of the integration of behavioral-based detection into our security systems.

To ensure the validity of behavioral-based malware detection systems and address the ethical concerns associated with them, several measures can be implemented. First, privacy concerns can be addressed by adopting measures such as data encryption, anonymization, or data aggregation techniques. These techniques can help protect user privacy by minimizing the risk of unauthorized access or misuse of personal information. Additionally, rigorous testing and evaluation of the detection algorithms are paramount to identify and minimize potential biases. By conducting thorough testing, researchers can ensure that the algorithms are accurate and free from bias, thus increasing the reliability of the system. Ultimately, a combination of advanced privacy protection measures and rigorous evaluation can enhance the validity and effectiveness of behavioral-based malware detection systems.

Conclusion

It is important to realize the need for cybersecurity advancements such as behavioral-based malware detection in order to combat the increasingly more prevalent cyber-attacks. Behavioral-based detection has its strengths and weaknesses when compared with signature-based detection, but an effective and comprehensive defense strategy should include a combination of both of these detection methods. Further research and the examination of Vali Cyber's Linux security tools and machine learning models showcase the potential of behavioral-based detection in identifying and mitigating cyber threats. However, the weaknesses, such as resource intensity and potential false positives, highlight the necessity for continuous improvement. The proposed methodology involves leveraging machine learning techniques and diverse datasets to enhance the adaptability and effectiveness of behavioral-based detection. The STS deliverable focuses on the societal and ethical considerations involved with implementing behavioral-based detection systems. Adequate attention needs to be paid to acknowledging the potential invasion of privacy, data security concerns, and algorithmic bias associated with behavioral-based detection. As technological advancements expand the scope of defensive security measures, it becomes imperative to address these concerns responsibly. Privacy protection measures, such as data encryption and anonymization, coupled with rigorous testing and evaluation to identify and mitigate biases, can enhance the validity and societal acceptance of behavioral-based malware detection systems. Therefore, this research will not only contribute to more effective implementations of behavioral-based detection but also highlight the significance of responsible and ethical development in the cybersecurity field. By addressing both the technical challenges and ethical considerations, we can contribute to the responsible and effective development of cybersecurity measures, safeguarding both individual privacy and collective security in our rapidly evolving digital landscape.

Word Count: 2185 words

Citations

1. Zheng, Y., Li, Z., Xu, X., & Zhao, Q. (2022). Dynamic defenses in cyber security: Techniques,
   methods and challenges. Digital Communications and Networks, 8(4), 422-435.
   https://doi.org/10.1016/j.dcan.2021.07.006

2. Singh, J., & Singh, J. (2021). A survey on machine learning-based malware detection in executable files. Journal of Systems Architecture, 112, 101861.
   https://doi.org/10.1016/j.sysarc.2020.101861

3. A. M. Lungana-Niculescu, A. Colesa and C. Oprisa, "False Positive Mitigation in Behavioral Malware Detection Using Deep Learning," 2018 IEEE 14th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2018, pp. 197-203, doi: 10.1109/ICCP.2018.8516611.

4. L. Caviglione et al., "Tight Arms Race: Overview of Current Malware Threats and Trends in Their Detection," in IEEE Access, vol. 9, pp. 5371-5396, 2021,
   doi: 10.1109/ACCESS.2020.3048319.

5. Gopinath M., Sibi Chakkaravarthy Sethuraman,
   A comprehensive survey on deep learning based malware detection techniques,
   Computer Science Review,
   Volume 47,
   2023,
   100529,
   ISSN 1574-0137,
   https://doi.org/10.1016/j.cosrev.2022.100529.

6. Eduardo Berrueta, Daniel Morato, Eduardo Magaña, Mikel Izal,
   Crypto-ransomware detection using machine learning models in file-sharing network scenarios with encrypted traffic,
   Expert Systems with Applications,
   Volume 209,
   2022,
   118299,
   ISSN 0957-4174,
   https://doi.org/10.1016/j.eswa.2022.118299.

7. The latest Cyber Crime Statistics (updated October 2023): Aag it support. (2023). Retrieved from https://aag-it.com/the-latest-cyber-crime-

statistics/#:~:text=The%20growing%20cost%20of%20cyber%20crime&text=The%20average%20cost%20of%20a,to%20%2410.5%20trillion%20by%202025.

8. Malek, Zakiyabanu & Trivedi, Bhushan. (2018). User Behaviour based Intrusion Detection System Overview. Engineering, Technology and Applied Science Research. 6. 149-156. https://www.researchgate.net/publication/343166387_User_Behaviour_based_Intrusion_Detection_System_Overview

9. Scott, J. (n.d.). Retrieved from https://informationsecurity.report/Resources/Whitepapers/920fbb41-8dc9-4053-bd01-72f961db24d9_ICIT-Analysis-Signature-Based-Malware-Detection-is-Dead.pdf