META-STUDY ON DETECTING ILLICIT CRYPTOCURRENCY MINING

A Research Paper submitted to the Department of Computer Science
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By

Joseph Davidson

April 27, 2022

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.

ADVISOR
Daniel Graham, Department of Computer Science

Meta-study on Detecting Illicit Cryptocurrency Mining

CS4991 Capstone Report, 2022

Joseph Davidson
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia, USA
Jmd8rb@virginia.edu

Abstract
As cryptocurrencies have gained value and adoption, they have increasingly become the target of cyberattacks, whether as a source of ransom or being directly mined from a host computer. As a result, reliable means of detecting cryptojacking are necessary in order to prevent this form of malware and the costs victims must shoulder.

In order to better understand the existing technologies and potential areas of improvement, I identified and analyzed a variety of these cryptojacking detection methods. Some examples include those used in traditional malware identification, such as fragment analysis. Additionally, I examined machine learning methods, network analysis methods, as well as methods that examine resource use and allocation in devices. I found that each of these methods has its advantages and disadvantages and that each, to varying degrees, has relevance in a robust detection system.

1 Introduction

As recently as the summer of 2021, the US witnessed one of the most disruptive cyberattacks in history. The Colonial Pipeline, which transports huge volumes of gas and jet fuel to the Southeast, was shut down due to a ransomware attack that demanded cryptocurrency as payment [1]. This is just one of many increasingly common examples of cyberattacks using cryptocurrencies as a convenient means of receiving ransoms or financial benefit.

As cryptocurrencies have become more valuable and widely used, they have increasingly become the target of cyberattacks, whether as a source of ransom or directly being mined from a host computer [2]. In particular, cryptojacking, when bad actors use a victim's computer processing power without permission to mine for cryptocurrencies, has become widespread. Cryptojacking incurs many costs on victims, including sluggish performance, additional electricity usage, and component damage [3]. In order to prevent these negative outcomes related to cryptojacking, we must develop methods for detecting and removing the offending malware.

2 Review of Research

A variety of methods currently exist to detect when cryptojacking is occurring. One of the most common ways of identifying malware uses fragment analysis, which is also a very common approach for detecting cryptojacking [4]. While this method cannot identify all variations of these attacks it serves as a good baseline to build on.

In addition to fragment analysis, there are machine learning approaches, which seek to build models that can effectively detect cryptojacking through learned pattern recognition [5]. There are approaches that

analyze WebSocket payloads to detect potential communication between attackers and a victim's device [6]. Other methods focus on behavior-based decision tree analysis in order to identify the common patterns of cryptojacking [7].

3 Meta-Study Analysis

As discussed in the previous section there are a variety of methods that currently exist within academic literature for detecting cryptojacking. The first method I will discuss is fragment analysis, sometimes referred to as semantic analysis. This type of malware detection is not new and has been utilized in many different settings for detecting a variety of viruses and other cybersecurity threats. Fragment analysis, as the name implies, relies on using known fragments of viruses and other malware. Romano et al. [2020] used a form of fragment analysis that focused on detected the use of certain hashing functions within WebAssembly by detecting the patterns that typical hashing functions use. Focusing on the hash functions is helpful as all cryptocurrencies and therefore all cryptojacking requires the use of hashing. This method is not affected by the use of various language to implement cryptojacking. Additionally, within their tests, this method provided to be particularly effective.

Using machine learning techniques for cryptojacking detection is another popular method. Within their paper Handaya, et al. [2020] discuss the various steps needed to implement a functioning machine learning-based approach to detect cryptojacking. They explain the need to have an adequately large and varied dataset, and an effective way to abstract features from malware in order to train and implement a machine learning model. One of the primary goals of their paper was to develop a technique that would effectively to detect cryptojacking when day-zero vulnerabilities are discovered. The approach they settled on was to use a few different techniques such as random forest, SVM, and k-nearest neighbor in order to construct the models.

In addition to fragment analysis and machine learning methods, some papers have analyzed network connections and traffic to identify when cryptojacking is occurring. In their paper Saad et al. [2019] demonstrate the methodology for using WebSocket payloads to identify cryptocurrency mining, and therefore cryptojacking, on web applications. Their method works due to the relatively unique pattern of payload size that cryptojacking websites send to servers. From this pattern the authors were able to accurately identify instances of cryptojacking using a web browser extension they developed.

The last methodology is the use of a behavior-based decision tree for identifying cryptojacking. In Tanana and Tanana's [2020] paper they use a combination of CPU usage as well as other factors such as network usage and calls to common crypto-mining libraries in order to accurately identify when cryptojacking was occurring. This method uses the various levels of usage and their consistence and concurrency in order to identify the behavior as being similar to cryptojacking and therefore in need of being flagged and handed.

4 Meta-Study Findings

There are a variety of effective methods for detecting cryptojacking. While these methods do differ in their flexibility and generalizability, given the correct conditions all of them can be very effective in their job of identification. That being said, in a "wild" environment, condition cannot be guaranteed and often times attackers intentionally

obfuscate their attacks to make detection more difficult.

With these factors in mind, it is important to use more generalized methods, such as WebSocket payloads analyses and behavioral analyses, to ensure initial detection of new approaches to cryptojacking. Additionally, methods such as machine learning and fragment analysis, should be implemented once a sufficiently large sample of attacks have been identified.

5 Future Work

Cryptocurrencies are a relatively new invention and the ways in which they are used in attacks continues to change and evolve. Given these dynamic circumstances it is important that new methodologies continue to be developed and refined in response. Particularly, in the area of cryptojacking it is important for more work to be done in order to identify and counter new approaches that bad actors will develop to avoid current detection techniques.

In addition to the technical challenges that cryptocurrencies have presented to the cybersecurity field, it is also important to maintain an accurate and up-to-date understanding of the social, economic, and governmental aspects of cryptocurrencies. Along these lines, more research and attention should be given to the effects that cryptocurrencies cause in our world in each of the many areas they impact. With better understanding hopefully our businesses, governments, and societies will be able to gain the greatest benefit from the technical innovations that have come with cryptocurrencies while also minimizing the harm that this new technology has the potential to create.

References

[1] William Turton and Kartikay Mehrotra. 2021. Hackers Breach Colonial Pipeline Using compromised Password. (June 2021). Retrieved 2021 from https://www.bloomberg.com/news/articles/2021-06-04/hackers-breached-colonial-pipeline-using-compromised-password

[2] Radhesh Krishnan Konoth et al. 2018. Minesweeper. *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security* (2018). DOI:http://dx.doi.org/10.1145/3243734.3243858

[3] Tom Leithauser. 2019. Report: 2M Cyber Incidents Caused $45B in Losses in 2018. (July 2019). Retrieved 2021 from https://www.proquest.com/docview/2266935077/LXHADICJ3853LT7R/?accountid=133485

[4] Alan Romano, Yunhui Zheng, and Weihang Wang. 2020. Minerray. *Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering* (2020). DOI:http://dx.doi.org/10.1145/3324884.3416580

[5] W.B. Handaya, M.N. Yusoff, and A. Jantan. 2020. Machine learning approach for detection of fileless cryptocurrency mining malware. *Journal of Physics: Conference Series* 1450, 1 (2020), 012075. DOI:http://dx.doi.org/10.1088/1742-6596/1450/1/012075

[6] Muhammad Saad, Aminollah Khormali, and Aziz Mohaisen. 2019. Dine and dash: Static, dynamic, and economic analysis of in-browser cryptojacking. *2019 APWG Symposium on Electronic Crime Research (eCrime)* (2019). DOI:http://dx.doi.org/10.1109/ecrime47957.2019.9037576

[7] Dmitry Tanana and Galina Tanana. 2020. Advanced behavior-based technique for Cryptojacking malware detection. *2020 14th International Conference on Signal Processing and Communication Systems (ICSPCS)* (2020). DOI:http://dx.doi.org/10.1109/icspcs50536.2020.9310048