Smart Analysis and Detection System for New Host-Based Cryptojacking Malware Dataset

Hadeel Almurshid

Computer Science Department, Prince Sultan University, Riyadh 11586, Saudi Arabia,
221421254@psu.edu.sa

Abstract

Cryptocurrency is a quickly growing technology in the finance industry, with the first cryptocurrency, Bitcoin, being created in 2009. Each cryptocurrency has its own unique hash value, and cryptocurrency mining involves participating in a guessing competition to release a unique hash into circulation, with the winner receiving a modest bonus in the form of bitcoin. However, as more bitcoins are discovered, it becomes increasingly difficult to obtain more, resulting in a need for extra computer resources and power. Consequently, the increasing popularity of cryptocurrency has led to a rise in cryptojacking malware, which secretly uses victims' computing resources to mine cryptocurrency. This malware can be either web-based or host-based, with similar execution and goals but differing in implementation and injection. Cryptojacking has affected numerous devices worldwide, but few studies have been carried out to detect it, especially the host-based type. Furthermore, the current studies on cryptojacking have limited datasets, which are often outdated or small, and the prediction models developed from these datasets may not be accurate. To address this gap, we conducted a thorough analysis of cryptojacking's behavior, lifecycle, impact, implementations, and possible detection methods. Additionally, we created an up-to-date dataset consisting of 114,985 samples, with 57,948 categorized as benign and 57,037 as cryptojacking. The dataset was used to build a smart cryptojacking detection system, with 5 different convolutional neural network models trained and evaluated against a subset of the dataset. The best performing model achieved an accuracy of 98.4%, an F1-Score of 98.3%, a precision of 98.4%, and a recall of 98.4%. Our proposed method, which involves running Windows executables in an isolated environment and closely monitoring their CPU usage, provides a thorough understanding of cryptojacking malware behavior and enables detection of the malware. The comprehensive dataset collected facilitates efficient detection model development. Additionally, evaluating the dataset with 5 different CNN algorithms and assessing their performance using established evaluation metrics ensures the effectiveness of our proposed method and dataset.

Keywords: Cryptocurrency; Cryptomining; Blockchain; Cryptojacking Malware; Host-based; Machine Learning; Deep Learning; Convolutional Neural Network (CNN); Dataset.

Introduction

Cryptocurrency has emerged as one of the fastest-evolving technologies in finance. Money has taken various forms throughout history, such as goods, cowrie shells, metal, banknotes, and more. However, the true revolution of money came in 2009 when an anonymous person or group named Satoshi Nakamoto introduced Bitcoin, the first-ever cryptocurrency [1]. Bitcoin's primary aim is to eliminate the need for intermediaries to control transactions. Instead, it relies on a computer algorithm called the blockchain [2] [3] [4], which operates based on

trust in the algorithm rather than any individual or institution.

Each cryptocurrency has a unique hash, and cryptocurrency mining involves solving a complex mathematical problem to guess the hash and release it into circulation. The winner of the mining competition receives a small amount of cryptocurrency as an incentive, encouraging users to support the peer-to-peer network [1]. However, as more bitcoins are discovered, obtaining more becomes increasingly challenging, requiring more computer resources and electricity.

As mining becomes more difficult and resource-intensive, attackers have developed an illegal mining technique called cryptojacking, which allows them to receive the mining reward without using their resources. Cryptojacking malware is a new type of malware that can be classified into web-based and host-based cryptojacking [5]. In web-based cryptojacking, the attacker takes advantage of client-side web scripting technologies, such as JavaScript and WebAssembly, to inject the cryptomining code into a legitimate website to force visitors' devices to mine cryptocurrency for them. In host-based cryptojacking, the attacker first needs to download the cryptominer into the victim's device and then run the miner to force the device to mine cryptocurrency for them without the victim's knowledge.

Research Problem and Motivation

In recent times, cryptojacking has become a widespread issue affecting numerous devices worldwide [5]. This is due to the high resource requirements of mining, which can cause the victim's device to overheat and experience performance issues. Despite this, there has been a limited amount of research on detecting cryptojacking. Most of the research conducted on cryptojacking detection has focused on web-based cryptojacking malware [6] [7] [8] [9] [10] [11], with little research conducted on host-based cryptojacking malware in 2022 and 2023 so far. Additionally, some studies have used outdated or insufficient datasets to build prediction models, further highlighting the need for more comprehensive research.

In this study, we aim to provide a thorough analysis of cryptojacking malware, with a specific focus on host-based cryptojacking malware. Our study includes examining its behavior, lifecycle, implementations, and impact, as well as proposing potential detection solutions. We also introduce an up-to-date dataset of host-based cryptojacking malware that can be utilized to train and test machine learning and deep learning models, evaluate intrusion detection systems, and develop new mitigation strategies. Additionally, we can analyze this dataset to provide insights into the tactics and techniques used by attackers, helping to better defend against this growing threat. Overall, our dataset represents a valuable contribution to the cybersecurity field, and we hope it will inspire further research and innovation in the fight against this emerging malware.

Contributions and Paper Structure

As a result of the identified shortcomings

and gaps in research, particularly with regard to detecting host-based cryptojacking malware, we have primarily made the following contributions:

- Conducting a comprehensive study of the behavior and implementation of cryptojacking malware, as well as its impact.
- Analyzing and comparing various current methods for identifying cryptojacking malware.
- Introducing a large and up-to-date dataset of host-based cryptojacking malware.
- Evaluating the effectiveness of our dataset by testing and examining a range of deep learning algorithms.

Therefore, our contributions are significant in advancing cybersecurity by intensely studying and analyzing cryptojacking malware, providing a new dataset of host-based cryptojacking malware, conducting a thorough dataset analysis, and evaluating the performance of different CNN models for detecting and classifying this type of malware. The remainder of this work is structured as follows. Section 2 provides a brief overview and background of cryptocurrency, cryptomining, and cryptojacking malware categories. Section 3 summarizes and discusses the state-of-the-art cryptojacking malware detection techniques and highlights their limitations. Section 4 describes the research methodology of building the new dataset of host-based cryptojacking malware in detail. It also explains the malware cryptojacking analysis, sources of the samples, the data collection process, the dataset's composition, and the dataset's evaluation using different CNN models. Finally, section 5 summarizes the paper's conclusions and discusses the implications of the findings. Also, it highlights the study's limitations and suggests future research directions.

Background

This section provides background about cryptocurrency, cryptomining, and cryptojacking malware. To fulfill our first research objective, we conducted a comprehensive analysis of both types of cryptojacking malware, examining how they operate, are created and distributed, and the consequences they have on infected systems.

Cryptocurrency

Many people may ask: Why do people use cryptocurrency? What makes it valuable and trusted? What is cryptojacking and how is it used to make money? These questions could be answered by first understanding the history of money and how it has evolved up until today. By learning about the origins and development of money, we can gain a better understanding of the potential benefits and risks of using cryptocurrency, as well as the impact of cryptojacking on the cryptocurrency ecosystem.

Money has always been an essential need for humankind, dating back to the earliest records of history. In the past, people relied on direct bartering to exchange goods and services [12]. However, this system had several drawbacks. For instance, goods were not always divisible, making it challenging

to make exchanges. For example, a person wanting to buy a water bottle with a fully-grown cow would be out of luck. Additionally, determining the value of different goods could be challenging. For example, a farmer wanting to buy a diamond ring might have to offer several cars to match its price. To overcome these issues, our ancestors developed a more generic form of money.

Initially, cowrie shells were used as a form of currency [13], and they were used extensively and for an extended period, compared to other forms of primitive money. For many people, cowrie shells were considered ideal since they were durable, easy to count and clean, and difficult to counterfeit. However, as trade increased, the use of cowrie shells became more prevalent, leading to an oversupply and consequent depreciation. Other less common forms of primitive money included whale teeth and Rai stones.

Subsequently, human society transitioned from primitive money to coin money. Initially, coins were made from various metals, including copper, bronze, iron, aluminum, gold, and silver. Metals thus formed the basis of the transition from primitive money to coined money [14]. Initially, metal chunks were used as money, based on their weight. Subsequently, they were stamped to create coins, marking the first step in the shift from weighted to counted money.

As civilization evolved, new machines were invented for minting and printing, leading to the emergence of paper money [15]. These two forms of currency have endured over time and continue to be widely used today. Nowadays, when most people hear the term 'money,' they typically think of banknotes and coins.

Initially, banknotes and coins were backed by valuable objects, particularly gold. However, due to the perceived constraints of this system, bankers eliminated the gold standard in 1971, as declared by former U.S. president Richard Nixon [16]. This decision marked the introduction of what is now known as fiat currency. Unlike previous forms of money, fiat currency derives its value solely from the trust placed in the governing authorities [2], without any direct backing by precious metals or other tangible assets.

Money has continued to evolve with the advent of digitalization. Nowadays, a significant amount of money exists in digital form, with bank account balances being represented by digital numbers that are monitored and controlled by banks. Instead of relying solely on cash, many individuals now use credit cards for their purchases. The combination of the shift towards digital transactions and the concept of fiat currency has laid the groundwork for the development of cryptocurrency, which represents a true evolution in the realm of money.

In 2009, an individual or group operating under the pseudonym Satoshi Nakamoto introduced the world to Bitcoin, the first cryptocurrency. Bitcoin operates on the premise that it does not require trust in any particular entity or institution, but rather in the integrity of a computer algorithm

known as the blockchain [2]. It is decentralized and not managed by any government or central bank; instead, it is overseen by a vast network of computers around the globe. The more people trust and use Bitcoin or other cryptocurrencies, the more valuable they become. This is demonstrated by Bitcoin's price, which surged from $500 in 2015 to over $20,000 in 2022 [17].

Cryptomining

The process of obtaining a Bitcoin can be accomplished by either purchasing or mining it. Bitcoin is essentially a unique hash. Mining Bitcoins entails attempting to solve a complex mathematical problem to predict a hash of a Bitcoin that has not yet been released into circulation. This procedure is analogous to gold mining [1], where miners must expand their territory and use more resources to acquire more gold. In cryptocurrency mining, also known as cryptomining, the more Bitcoins that are discovered, the more challenging it becomes to obtain more of them. This complexity is reflected in the need for additional computer resources and electricity [1]. The mining function is a critical component of cryptocurrency because there is no authority to add new coins to circulation, and this is the only method for doing so.

Cryptojacking Malware

Cryptojacking is the illegal mining of cryptocurrency. As the mining process becomes very expensive, attackers invented a new type of malware that is called cryptojacking malware to force the victim's device to mine cryptocurrency for them. The malware can be found in two types, web-based and host-based.

Web-based Cryptojacking Malware

The interactive technologies of web browsers, including JavaScript and WebAssembly (Wasm), have frequently been exploited for malicious purposes [5]. In the context of cryptojacking, a harmful mining code is injected into a website to force visitors' devices to mine cryptocurrency. The trend of this attack began in 2017, when a service provider named Coinhive introduced a mining script that could be directly inserted into a web page to mine a particular cryptocurrency, Monero [10]. The company's objective was to provide website owners with a way to generate revenue as a substitute for using advertisements. However, hackers misused the service by illegally injecting the script into legitimate websites to mine cryptocurrency. Figure 1 depicts the lifecycle of web-based cryptojacking malware.

The illustrated lifecycle begins with the attacker signing up with a service provider, who then provides a unique API key to the attacker. This API key is used in a script that the attacker injects into a public website. Whenever a user visits the infected website, the malicious code runs in the background and mines cryptocurrency. Since the code includes the attacker's API key, the attacker will receive the revenue generated by the mining operation.

Host-based Cryptojacking Malware

In contrast, host-based cryptojacking malware requires direct installation onto the

victim's machine for cryptocurrency mining such as xmrig, T-Rex, ccminer, etc. [5]. The attacker may disguise the malicious mining code as seemingly harmless software, which can be downloaded by the user. Infected software is often located on online data-sharing platforms like Torrents and public clouds [18].

With attackers increasingly targeting more powerful machines, animators and video gamers have become primary targets [19]. Crackonosh, a miner malware, was found in multiple popular games, as per a report released by Avast [20]. The attacker uses a tool called InnoSetup [21] to inject miner software into a legitimate video game installer, allowing them to combine the miner installation with legitimate software. In essence, the miner and legitimate software are installed and operated simultaneously during installation. Figure 2 illustrates the lifecycle of the host-based cryptojacking malware.

The lifecycle of the host-based cryptojacking malware begins with the attacker using a tool like InnoSetup to merge legitimate software, such as a game, with a miner. A specific type of malware known as Crackanosh is an example of this. The attacker then uploads the infected file to a publicly available sharing platform. Once the victim downloads and installs the infected files, they become infected with the cryptojacking malware, and the attacker can generate revenue through their API key, which is included in the miner executable.

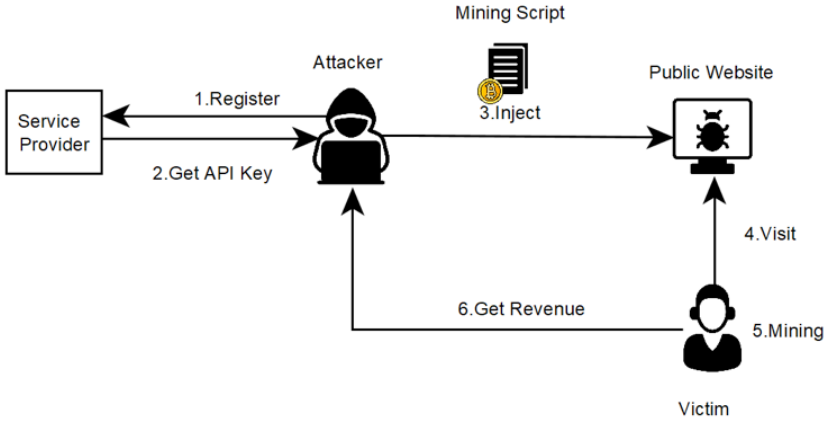

Fig. 1. Web-based cryptojacking malware lifecycle

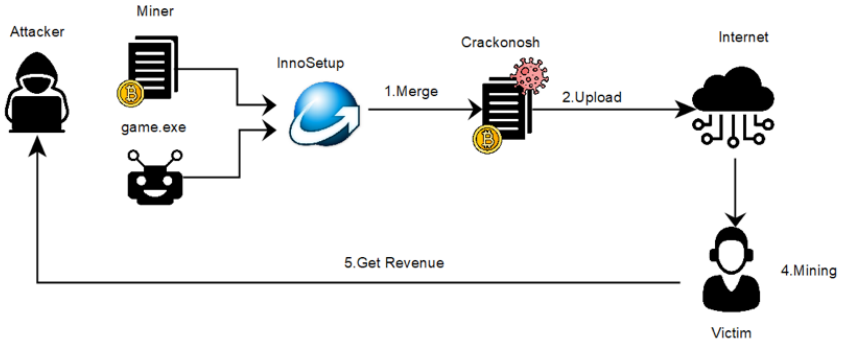

Fig. 2. Host-based cryptojacking malware lifecycle

Literature Review

As we mentioned earlier, Cryptojacking refers to the unauthorized mining of cryptocurrency using malware known as Cryptojacking Malware. It has significantly impacted both individuals and organizations, leading to increased costs, slow or non-functional devices, and higher electricity bills due to the computing power required. However, it is important to highlight that the effect of cryptojacking on the cryptocurrency and blockchain industry has not been thoroughly studied, providing a chance to explore new research directions [5].

In this section, to accomplish our second research objective, we aim to enhance the understanding of current research directions by analyzing and comparing various methods used to identify cryptojacking malware. This malware is classified into two categories: Web-based Cryptojacking Malware and Host-based Cryptojacking Malware [5]. Below is a recent review study of both types.

Web-based Cryptojacking Malware

In [6], Faraz Naseem et al. presented MINOS, a lightweight detector for web-based cryptojacking malware that aims to detect mining activities on websites utilizing the WASM programming language. The detector utilizes deep learning and implements a convolutional neural network (CNN) model that has been trained on a dataset of 300 samples, half of which are malicious. MINOS achieved an accuracy of 98.97% while consuming only 4% and 6.5% of CPU and RAM, respectively. It should be noted that MINOS is only compatible with the Google Chrome web browser.

In [7], Franco Tommasi et al. put forward MinerAlert, a new technique that detects web-based cryptomining in real-time using a web browser extension. The approach uses a Support Vector Machine (SVM) model that classifies web pages based on the analysis of hardware resources performance like CPU and web page behavior. The researchers gathered data from 604 websites, including 130 malicious ones. By running the model on a combination of several features, they achieved a maximum accuracy of 99.59%.

In [8], Aldo Hernandez-Suarez et al. presented a new approach in which they used a deep dense neural network (DDNN) to classify websites based on their network traffic and hardware performance. The dataset consisted of 8000 benign sites and 8156 cryptojacking websites, and 18 different features were used as input to the model (such as total processor idle time, disk reading/sec, disk writing/sec, number of subprocesses, etc.). Their deep learning model achieved a precision of 99.41%, a recall of 99.10%, and an F1 score of 99.25%.

In [9], HYUNJI HONG et al. introduced CIRCUIT, a novel approach for detecting cryptojacking websites by monitoring the memory heap. This technique involves generating heap graphs that depict the behavior of the JavaScript code and extracting reference flows to identify the call flow of JavaScript objects and detect

cryptojacking behavior. To aid in detection, CIRCUIT stores signatures for cryptojacking websites. The authors noted that CIRCUIT is particularly effective at detecting obfuscated cryptojacking scripts. Their approach successfully detected 1813 cryptojacking websites among the top 306K websites, which includes the Alexa top 100K, Majestic top 200K, and Alexa category top websites. However, CIRCUIT has limitations, such as difficulties in handling abnormally obfuscated mining scripts resulting in long reference flows, which in turn requires significant editing when comparing signatures. It may also face issues when dealing with very short reference flows.

In [10], Min-Hao Wu et al. have presented a solution based on artificial neural network (ANN) named MinerGuard. The system aims to identify the existence of web-based cryptojacking malware by relying heavily on monitoring the CPU usage of the webpage. The authors used a dataset of 850 websites, with 350 classified as miners, and obtained an accuracy rate of 99%, including the detection of zero-day attacks. However, it should be noted that MinerGuard is only compatible with the development version of Google Chrome and relies on the chrome.processes API

In [11], Khan Abbasi et al. have proposed a hybrid approach to detect and prevent web-based cryptojacking malware. The approach aims to detect malware in both WASM and non-WASM scripts and comprises three phases: 1) comparing suspicious URLs to a blacklist, 2) analyzing malware based on static signatures, and 3) conducting dynamic analysis of malware. The authors evaluated their proposed solution on a dataset of 1000 samples, 30 of which were malicious, gathered from Alexa and PublicWWW. The approach achieved an accuracy of 99.6%.

Host-based Cryptojacking Malware

In [22], Hamid Darabian et al. conducted a study on the effectiveness of machine learning algorithms in detecting cryptojacking malware. The study utilized a dataset of 1500 Windows Portable Executables that were registered in VirusTotal in 2018. Both static and dynamic analyses were performed on the malware samples. For static analysis, the authors employed Long Short-Term Memory (LSTM), Attention-based LSTM, and CNN algorithms. These ML models were implemented into the opcodes of the cryptojacking malware samples, and a 95% accuracy rate was achieved. In dynamic analysis, the malware samples were executed in a sandbox environment, and system call sequences were captured, resulting in a success rate of 99%.

In [23], Dmitry Tanana proposed a detection technique that relies on the analysis of CPU and RAM consumption to detect the presence of cryptojacking, regardless of whether it is web-based or host-based. The technique employs a decision tree algorithm that compares the CPU and RAM consumption of an application to a predefined threshold. The author tested the technique on a dataset comprising 20 web-based cryptojacking and 5 host-based

cryptojacking malware samples and validated it on 40 browser-based and 10 host-based cryptojacking malware samples, all of which were collected from VirusShare in 2019. The technique achieved an 82% success rate.

In [24], Ganapathy Mani et al. introduced a novel framework named DeCrypto Pro, which is designed to select the appropriate machine learning algorithm, such as Random Forest, k-Nearest Neighbor, or LSTM, based on the computing resources of the system being investigated. DeCrypto Pro utilizes performance counters, such as CPU usage, to classify an application as benign or malicious. The authors collected and monitored a dataset of performance counters by running benign software, including 7-Zip, SecureZip, PeaZip, WinRAR, WinZip, and Freemake, as well as malicious software such as XMRig, XMRStak, Coinhive, Computta, and GUIminer. The proposed framework achieved an F1-score of 89.99%, 97.62%, and 95.5% for k-Nearest neighbor, Random Forest, and LSTM classifiers, respectively.

In [25], Gilberto Gomes et al. presented the CryingJackpot intrusion detection system (IDS), an unsupervised approach to detecting cryptojacking. The proposed solution utilizes system events and network flow data of an application as features for clustering, using K-means, Agglomerative, DBSCAN, and ensemble machine learning algorithms. The system was evaluated on a public dataset (CSECIC-IDS2018) and achieved an F1-Score of 82%. Additionally, the authors evaluated the system on a separate dataset they created and obtained an F1-Score of 97%. CryingJackpot is capable of detecting both types of cryptojacking malware.

In November 2021, a systematic study of 128 research papers was conducted by Ege Tekiner et al. [5] Two datasets were collected for the purpose of analyzing cryptojacking malware. The first dataset was obtained from VirusTotal using academic access and contained 20,200 cryptojacking samples. The second dataset was obtained from PublicWWW and contained 6,269 URLs. Among the key findings, only 7 out of the 128 research papers examined host-based cryptojacking malware, and Monero was identified as the most targeted cryptocurrency.

According to the 2022 mid-year update cyber threat report from SONICWALL[26], although cryptocurrency prices had decreased, the volume of cryptojacking malware had still increased. However, our research shows that there has been very little focus on web-based cryptojacking malware in 2022 and almost no contributions toward host-based malware. This is likely due to limited access to host-based malware samples, such as those from VirusTotal and VirusShare, in comparison to publicly available website samples.

Attackers are increasingly targeting devices with more processing power for faster profit, making host-based cryptojacking malware the new trend [5]. The European Union Agency for Cybersecurity (ENISA) cryptojacking report [27] states that host-based cryptojacking botnets can generate

$750K in a month, while web-based botnets can only generate $30K.

The main objective of our study is to enhance the field of cryptojacking detection by generating an extensive and up-to-date dataset of over 100,000 samples and building multiple detection models to demonstrate its efficiency. The effectiveness of these models will be assessed using standard metrics such as accuracy, recall, precision, and F1 score to determine the best model. Table 1 summarizes the previous research in this area.

Table 1. Summary of Previous Studies

Work	Year	Dataset	ML-Based	Classifier	Type	Performance
Minos [6]	2021	Benign: 150 Malicious: 150	Yes	CNN	Web-based	Accuracy=98.97%
MinerAlert [7]	2022	Benign: 474 Malicious: 130	Yes	SVM	Web-based	Accuracy=99.59%
[8]	2022	Benign: 8000 Malicious: 8156	Yes	DDNN	Web-based	Precision=99.41% Recall=99.10% F1-score=99.25%
CIRCUIT [9]	2022	306K most popular websites	No		Web-based	1813 cryptojacking websites detected
MinerGuard [10]	2022	Benign: 500 Malicious: 350	Yes	ANN	Web-based	Accuracy=99%
[11]	2023	Benign: 970 Malicious: 30	No	-	Web-based	Accuracy=99.6%
[22]	2020	Malicious: 1500	Yes	LSTM, Attention-based LSTM, CNN	Host-based	Static: Accuracy=95% Dynamic: Accuracy=99%
[23]	2020	Malicious: Web-based: 60 Host-based: 15	Yes	DT	Both	Success rate=82%
DeCrypto Pro [24]	2020	Benign: 6 Malicious: 5	Yes	RF, KNN, LSTM	Host-based	F1-score: KNN=89.99% RF=97.62% LSTM=95.5%
CryingJackpot [25]	2020	Dataset1: CSECIC-IDS2018 Dataset2: Malicious: Web-based: 5 Host-based: 5	Yes	K-means, Agglomerative, DBSCAN, ensemble ML	Both	Dataset1: F1-score=82% Dataset2: F1-score=97%
Ours (Proposed)	2023	Benign: 20,000 Malicious: 20,000	Yes	5 CNN models Best performed: Color-based Scratch model [28]	Host-based	Accuracy=98.4% F1-score=98.3% Precision=98.4% Recall=98.4%

Research Methodology

We initiated this study by analyzing and explaining the characteristics of cryptojacking malware, including its behavior when impacting a system. Subsequently, we presented an up-to-date dataset of host-

based cryptojacking malware comprising 114,985 sample files. We proved the effectiveness of this dataset by training and evaluating 5 distinct CNN models using a subset of 40K balanced samples consisting of 20,000 benign and 20,000 cryptojacking samples.

Cryptojacking Malware Analysis

As noted before, cryptojacking malware does not cause direct harm like other kinds of malware (e.g., ransomware). At first glance, it may not seem as harmful; however, it causes severe consequences the longer it resides in the infected system. It greedily drains the infected system's resources, resulting in system deficiencies and high electricity consumption.

To enrich our first objective and gain a better understanding of the behavior of the malware, we executed multiple Windows executables in an isolated environment and closely monitored their CPU usage. The graphical representation in Figure 3 illustrates the typical CPU usage of the machine (7%), while Figure 4 depicts the CPU consumption when the cryptojacking malware was running (100%).

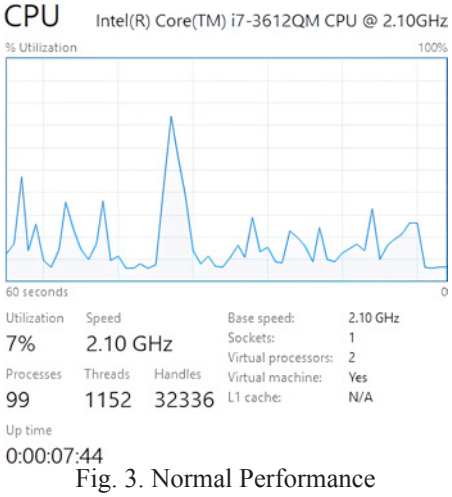

Fig. 3. Normal Performance

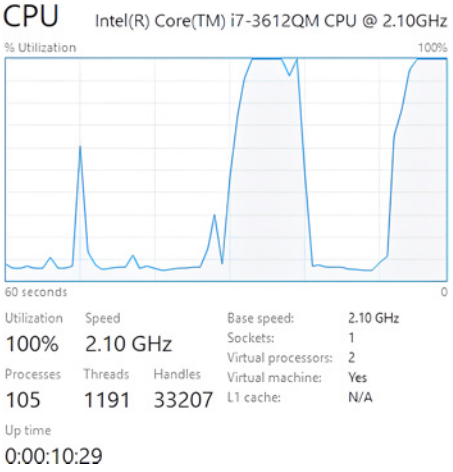

Fig. 4. Performance with Cryptojacking Malware Running

It is not surprising that the CPU usage increases significantly when running cryptojacking malware, as its primary function is to compute unique hashes for cryptocurrency mining using cryptography libraries and APIs. As a result, the performance of the system is affected due to the resource-intensive nature of these activities.

Dataset Collection

In this work, we aim to address the current gap in the literature regarding host-based cryptojacking malware detection [5]. To achieve our third research objective, we gathered a comprehensive dataset of 114,985 samples to facilitate the development of efficient detection models. The dataset was generated recently and collected in January 2023 using a premium VirusTotal [29] account. It includes executable files in a variety of formats, such as exe, pe, elf, apk, among others.

We used VirusTotal to search for both benign and cryptojacking samples and stored their hashes in two separate text files. The initial output was 57,037 hash values for executable files of cryptojacking malware

samples and 57,948 hash values for executable files of benign software samples. To further process the data, we divided each text file into several chunks of 5000 hash values. We then utilized VirusTotal to download the actual executable files for each chunk file. After this, we carefully cleaned the dataset by removing any files that were irrelevant or damaged. The resulting dataset contains a total of 114,985 samples, which is our final output.

Cryptojacking CNN-based Detection Analysis: A Case Study

Our final objective of evaluating the effectiveness of our dataset was achieved through testing it with 5 different CNN algorithms: Scratch [28], VGG16, ResNet50, VGG19, and DenseNet121.

To conduct our testing, we employed a balanced dataset of 40,000 samples. The executable files were converted to color images, and subsequently to grayscale images. We trained each algorithm twice - once with color images and once with grayscale images. We assessed the performance of each trained model using the evaluation metrics described in [30] [31]:

$$\text{Accuracy} = \frac{TN + TP}{FP + TP + FN + TN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3)$$

$$\text{F1 Score} = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN} \quad (4)$$

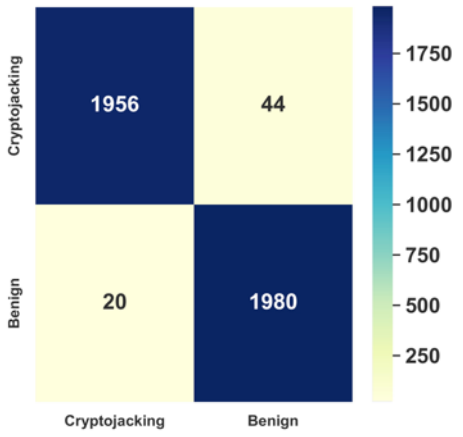

Fig. 5. Scratch Model Confusion Matrix

where TP is true positive score, FP is false positive score, TN is true negative score, and FN is false negative score [32] [33].

After conducting a comprehensive analysis and evaluation of all CNN models, the Scratch model [28] demonstrated the best performance in detecting cryptojacking malware when color images were used. It resulted in an accuracy of 98.4%, an F1-score of 98.3%, a precision of 98.4%, and a recall of 98.4%. Figure 5 shows the confusion matrix, and Figure 6 shows the loss and accuracy curve of the Scratch model's performance.

On the other hand, ResNet50 showed the best performance in detecting cryptojacking malware when grayscale images were used. It achieved an accuracy of 97.8%, an F1-score of 97.8%, a precision of 97.8%, and a recall of 97.8%. Figure 7 displays the confusion matrix, and Figure 8 displays the loss and accuracy curve of the ResNet50 model's performance.

All the tested models exhibited promising performance in terms of all the examined assessment detection parameters. The evaluation results for each model using

both color and grayscale images are presented in Table 2.

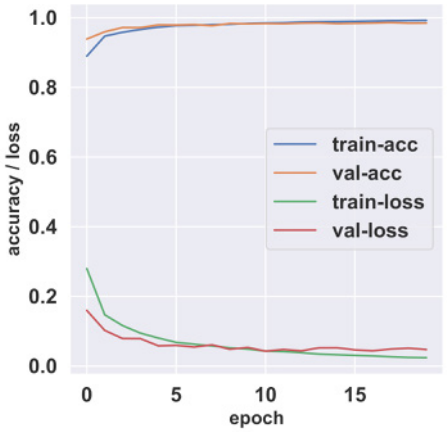

Fig.6. Scratch Model Loss and Accuracy Curve

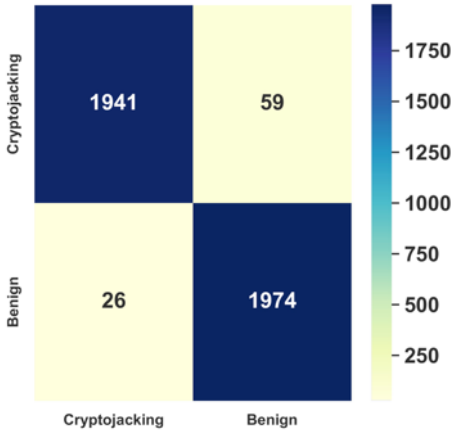

Fig. 7. ResNet50 Model Confusion Matrix

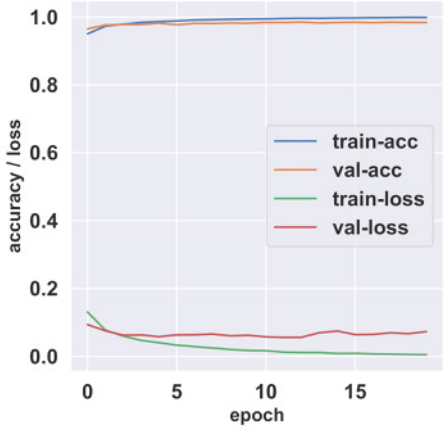

Fig. 8. ResNet50 Model Loss and Accuracy Curve

Table 2. Results of Evaluation Analysis

Model	Format	Accuracy (%)	F1 Score (%)	Precision (%)	Recall (%)
Scratch	color	98.4	98.3	98.4	98.4
	gray	97.7	97.7	97.8	97.8
VGG16	color	98.1	98.1	98.1	98.1
	gray	97.7	97.7	97.8	97.8
ResNet50	color	98.1	98.1	98.1	98.1
	gray	97.8	97.8	97.8	97.8
VGG19	color	97.8	97.8	97.8	97.8
	gray	97.2	97.2	97.2	97.2
DenseNet21	color	97.2	97.1	97.3	97.2
	gray	96.7	96.7	96.8	96.7

Conclusion and Future Work

To sum up, the use of cryptojacking malware poses a substantial danger to both individuals and organizations. This malicious software can covertly seize a victim's computer resources to mine cryp-

tocurrency without their awareness or consent. Despite being a relatively recent phenomenon, cryptojacking malware has already inflicted significant harm on numerous victims.

Our study is very important as it provides valuable insights into how cryptojacking malware works, is injected, implemented, and executed. This will serve further research into understating cryptojacking malware to help in creating better detection and prevention strategies. Additionally, we surveyed the current studies in the cryptojacking malware, thus helping in introducing current gaps and limitations to encourage more efficient solutions. Furthermore, we introduced a large up-to-date dataset of 114,985 samples for host-based cryptojacking malware. We demonstrated the efficacy of our dataset by training 5 convolutional neural network models against a subset of it, using both color and grayscale images. The Scratch model using color images performed the best, with an accuracy of 98.4%, an F1-Score of 98.3%, a precision of 98.4%, and a recall of 98.4% This research emphasizes the need for ongoing investigation into cryptojacking malware and the implementation of robust cybersecurity measures to safeguard against this growing threat. As a result, it is crucial for both individuals and organizations to remain vigilant and take appropriate measures to protect their systems and data against cryptojacking attacks.

While our study has yielded valuable insights into the features and actions of cryptojacking malware, it is important to acknowledge its limitations. These include:

- The dataset has been obtained from a single source, and the cryptojacking samples included in the dataset were restricted to those identified by antivirus software in VirusTotal.
- Our models were trained on a portion of the dataset, rather than the complete dataset.
- Restricting model training to only 5 models
- The malware was not statically analyzed
- Furthermore, there are numerous opportunities for further research that can broaden and enhance the findings of this study. The following are potential areas for future investigation:
- Increase the number of ML and DL models trained on the complete dataset
- Collect host-based cryptojacking malware datasets from multiple sources
- Conduct research to comprehensively analyze the impact of cryptojacking malware on the cryptocurrency industry
- Explore and analyze the different targets of cryptojacking malware.
- Examine and analyze the existing techniques used by organizations to detect and prevent cryptojacking malware
- Examine and analyze the existing techniques used by organizations to hinder the impact of the cryptojacking malware

References

[1] S. Nakamoto, “Bitcoin: a peer-to-peer electronic cash system, october 2008,” Metzdowd mailing list, 2008.

[2] E. Prasad, The Future of Money: How the Digital Revolution Is Transforming Currencies and Finance, 1st ed. Belknap Press: An Imprint of Harvard University Press, 9 2021.

[3] F. Jamil, O. Cheikhrouhou, H. Jamil, A. Koubaa, A. Derhab, and M. A. Ferrag, "Petroblock: A blockchain-based payment mechanism for fueling smart vehicles," Applied Sciences, vol. 11, no. 7, p. 3055, 2021.

[4] A. Allouch, O. Cheikhrouhou, A.Koubˆaa, K. Toumi, M. Khalgui, and T. Nguyen Gia, "Utm-chain: blockchain-based secure unmanned traffic management for internet of drones," Sensors, vol. 21, no. 9, p. 3049, 2021.

[5] E. Tekiner, A. Acar, A. S. Uluagac, E. Kirda, and A. A. Selcuk, "Sok: Cryptojacking malware," in 2021 IEEE European Symposium on Security and Privacy (EuroSP). IEEE, 2021, pp. 120–139.

[6] F.Naseem, A. Aris, L. Babun, E. Tekiner, and A. S. Uluagac, "Minos*: Alightweight real-time cryptojacking detection system." Network and Distributed Systems Security (NDSS) Symposium 2021, 2 2021.

[7] F. Tommasi, C. Catalano, U. Corvaglia, and I. Taurino, "Mineralert: an hybrid approach for web mining detection," Journal of Computer Virology and Hacking Techniques, vol. 18, p. 333–346, 2022.

[8] A. Hernandez-Suarez, G. Sanchez-Perez, L. K. Toscano-Medina, J. Olivares-Mercado, J. Portillo-Portilo, J. G.Avalos, and L. J. G.Villalba, "Detecting cryptojacking web threats: An approach with autoencoders and deep dense neural networks," Applied Sciences (Switzerland), vol. 12, no. 7, 2022.

[9] H. Hong, S. Woo, S. Park, J. Lee, and H. Lee, "Circuit: A javascript memory heap-based approach for precisely detecting cryptojacking websites," IEEE Access, vol. 10, pp. 95 356–95 368, 2022.

[10] M.-H. Wu, Y.-J. Lai, Y.-L. Hwang, T.-C. Chang, and F.-H. Hsu, "Minerguard: A solution to detect browser-based cryptocurrency mining through machinelearning," Applied Sciences, vol. 12, no. 19, 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/19/9838

[11] M. H. Khan Abbasi, S. Ullah, T. Ahmad, and A. Buriro, "A real-time hybrid approach to combat in-browser cryptojacking malware," Applied Sciences, vol. 13, no. 4, 2023. [Online]. Available: https://www.mdpi.com/2076-3417/13/4/2039

[12] G. Davies, A History of Money: From Ancient Times to the Present Day, 3rd ed. University of Wales Press, 11 2002.

[13] P. Patel, A. Dalvi, and I. Siddavatam, "Exploiting honeypot for cryptojacking: The other side of the story of honeypot deployment," in 2022 6th International Conference On Computing, Communication, Control And Automation (ICCUBEA. IEEE, 2022, pp. 1–5.

[14] L. P. Krishnan, I.Vakilinia, S. Reddivari, and S. Ahuja, "Scams and solutions in cryptocurrencies— a survey analyzing

existing machine learning models," Information, vol. 14, no. 3, p. 171, 2023.

[15] Z. Mineau, D. Hoffman, J. Lor, and N. Choudhury, "Cryptocurrency: Is it the future of payments?" in Cybersecurity for Smart Cities. Springer, 2023, pp. 169–183.

[16] T. I. TEAM, "Fiat vs. representative money: What's the difference?" https://www.investopedia.com/ask/answers/041615/what-difference-betweenfiat- money-and-representative-money.asp, 10 2022.

[17] "Coinmarketcap - bitcoin," https://coinmarketcap.com/currencies/bitcoin/, 10 2022.

[18] M. Abdelrahim, B. Omonayajo, A. S. Mubarak, and F. Al-Turjman, "Crypto currency cloud mining," in 2022 International Conference on Artificial Intelligence in Everything (AIE). IEEE, 2022, pp. 488–492.

[19] A. Mozo, A´ . Gonza´lez-Prieto, A. Pastor, S. Go´mez-Canaval, and E. Talavera, "Synthetic flow-based cryptomining attack generation through generative adversarial networks," Scientific Reports, vol. 12, no. 1, p. 2091, 2022.

[20] "avastreport," https://decoded.avast.io/danielbenes/crack-onosh-a-new-malwaredistributed- in-cracked-software/, 3 2023.

[21] "Innosetup," https://jrsoftware.org/isinfo.php, 3 2023.

[22] H. Darabian, S. Homayounoot, A. Dehghantanha, S. Hashemi, H. Karimipour, R. M. Parizi, and K. K. R. Choo, Journal of Grid Computing.

[23] D. Tanana, "Behavior-based detection of cryptojacking malware," in 2020 Ural Symposium on Biomedical Engineering, Radioelectronics and Information Technology (USBEREIT). IEEE, 2020, pp. 0543–0545.

[24] G. Mani, V. Pasumarti, B. Bhargava, F. T. Vora, J. Macdonald, J. King, and J. Kobes, "Decrypto pro: Deep learning based cryptomining malware detection using performance counters." IEEE, 2020, pp. 109–118.

[25] G. Gomes, L. Dias, and M. Correia, "Cryingjackpot: Network flows and performance counters against cryptojacking," in 2020 IEEE 19th International Symposium on Network Computing and Applications (NCA). IEEE, 2020, pp. 1–10.

[26] "Min-year update 2022 sonicwall cyber threat report," pp. 31–33, 6 2022.

[27] "Cryptojacking enisa threat landscape report," 2020.

[28] I. Almomani, M. Ahmed, andW. El-Shafai, "Android malware analysis in a nutshell," Plos one, vol. 17, no. 7, p. e0270647, 2022.

[29] "Virustotal," https://www.virustotal.com/gui/home/upload, 2 2023.

[30] A. Ammar, A. Koubaa, and B. Benjdira, "Deep-learning-based automated palm tree counting and geolocation in large farms from aerial geotagged images," Agronomy, vol. 11, no. 8, p. 1458, 2021.

[31] A. Noor, Y. Zhao, A. Koubˆaa, L.

Wu, R. Khan, and F. Y. Abdalla, "Automated sheep facial expression classification using deep transfer learning," Computers and Electronics in Agriculture, vol. 175, p. 105528, 2020.

[32] I. Almomani, A. Alkhayer, and W. El-Shafai, "An automated vision-based deep learning model for efficient detection of android malware attacks," IEEE Access, vol. 10, pp. 2700–2720, 2022.

[33] W. El-Shafai, I. Almomani, and A. AlKhayer, "Visualized malware multiclassification framework using fine-tuned cnn-based transfer learning models," Applied Sciences, vol. 11, no. 14, p. 6446, 2021.