An investigation into the current state of web based cryptominers and cryptojacking

Submitted in partial fulfilment of the requirements for the degree of Master of Science of Rhodes University

Robert Len

June 2020

Abstract

The aim of this research was to conduct a review of the current state and extent of surreptitious crypto mining software and its prevalence as a means for income generation. Income is generated through the use of a viewer's browser to execute custom JavaScript code to mine cryptocurrencies such as Monero and Bitcoin. The research aimed to measure the prevalence of illicit mining scripts being utilised for "in-browser" cryptojacking while further analysing the ecosystems that support the cryptomining environment. The extent of the research covers aspects such as the content (or type) of the sites hosting malicious "in-browser" cryptomining software as well as the occurrences of currencies utilised in the cryptographic mining and the analysis of cryptographic mining code samples. This research aims to compare the results of previous work with the current state of affairs since the closure of Coinhive in March 2018. Coinhive were at the time the market leader in such web based mining services.

Beyond the analysis of the prevalence of cryptomining on the web today, research into the methodologies and techniques used to detect and counteract cryptomining are also conducted. This includes the most recent developments in malicious JavaScript de-obfuscation as well as cryptomining signature creation and detection. Methodologies for heuristic JavaScript behaviour identification and subsequent identification of potential malicious out-liars are also included within the research of the countermeasure analysis.

The research revealed that although no longer functional, Coinhive remained as the most prevalent script being used for "in-browser" cryptomining services. While remaining the most prevalent, there was however a significant decline in overall occurrences compared to when `coinhive.com` was operational. Analysis of the ecosystem hosting "in-browser" mining websites was found to be distributed both geographically as well as in terms of domain categorisations.

Acknowledgements

I would like to express my gratitude to my supervisor Prof. Barry Irwin for providing constant support and feedback throughout the entire research process, as well as through the delivery of this paper. Further thanks must go out to the Computer Science department at Rhodes University for all their assistance and support.

Last but not least, thank you Tony and Izzy for keeping me company and sane during the late nights and early mornings.

Thanks to MaxMind Geolite2 for allowing the use of their IP geolocation databases for academic research as well as to VirusTotal for allowing me access to their API for research purposes.

Contents

Abstract i

Acknowledgements ii

1 Introduction 1

1.1 Background . . . 1
1.2 Problem statement . . . 3
1.3 Research Goals . . . 4
1.4 Research Limitations . . . 4
1.5 Document Conventions . . . 5
1.6 Document Structure . . . 5

2 Literature Review 6

2.1 Introduction . . . 6
2.2 Cryptocurrency Mining Basics . . . 6
2.2.1 Mining Pools . . . 7
2.2.2 Monero . . . 8
2.3 The Monero and Coinhive infrastructure . . . 9
2.4 Ethical Discussions . . . 10
2.5 Threat Modelling . . . 11
2.5.1 Malicious mining infrastructure . . . 12
2.6 Browser Mining Prevalence . . . 13
2.6.1 The Coinhive Captcha service . . . 14
2.6.2 The Coinhive short link forwarding service . . . 14
2.7 Detecting and Countering Web Based Cryptomining . . . 17
2.7.1 Static Analysis . . . 17
2.7.2 Dynamic Analysis . . . 17
2.7.3 Deeper Analysis with CMTracker . . . 20

2.7.3.1 The Hash based profiler . . . 21
2.7.3.2 The Stack based profiler . . . 21
2.7.4 Sample Collection - for CMTracker . . . 22
2.7.5 Evasion Techniques . . . 23
2.8 Summary . . . 24

3 Data Collection and Enrichment **25**
3.1 Research Method . . . 25
3.2 Initial Dataset Aquisition . . . 26
3.3 Data Enrichment . . . 31
3.4 Summary . . . 34

4 Data Analysis **35**
4.1 Geographic analysis by mining variant . . . 35
4.1.1 `feesocrald.com` . . . 35
4.1.2 `jsecoin.com` . . . 35
4.1.3 `freecontent.date` . . . 36
4.2 Domain Categorisation Analysis . . . 38
4.3 ISP and subnet Analysis . . . 40
4.4 Graph Network Analysis . . . 43
4.4.1 Case 1 - 213.232.126.134 . . . 44
4.4.2 Case 2 - 66.96.161.196 . . . 45
4.4.3 Case 3 - 81.231.232.61 . . . 47
4.4.4 Case 4 - `ghs.google.com` . . . 48
4.4.5 Case 5 - `blogspot.l.googleusercontent.com` . . . 50
4.4.6 Case 6 - 62.210.16.62 . . . 51
4.5 Domain Categorisation Graph Analysis . . . 52
4.6 VirusTotal Graph Analysis . . . 54
4.7 JSECoin Analysis . . . 57
4.7.1 JSECoin financial analysis . . . 60
4.8 Discussion of Results . . . 63

5 Conclusion **66**
5.1 Introduction . . . 66
5.2 Summary of Previous Chapters . . . 66
5.3 Review of Research Objectives . . . 66

5.4 Closing Statement 67
5.5 Future Work 67

List of Figures

2.1 Monero blockchain and "proof of work" with mining input 9
2.2 Coinhive prevalence in browser mining 14
2.3 Tokens by number of links 15

3.1 Example of content from enriched dataset 34

4.1 ForcePoint domain categorisations with uncategorised removed 38
4.2 BitDefender domain categorisations 40
4.3 213.232.126.134 endpoint analysis 45
4.4 IP node analysis 45
4.5 Mining variant node analysis 46
4.6 66.96.161.196 endpoint analysis 47
4.7 81.231.232.61 endpoint analysis 48
4.8 `ghs.google.com endpoint Analysis` 49
4.9 `blogspot.l.googleusercontent.com endpoint analysis` 50
4.10 62.210.16.62 endpoint analysis 51
4.11 206.189.153.135 domain categorisation analysis 52
4.12 109.108.145.100 domain categorisation analysis 53
4.13 Detected VirusTotal communicating samples from the top 10 IP addresses 55
4.14 Detected VirusTotal downloaded samples from the top 10 IP addresses - 62.210.16.62 56
4.15 Detected VirusTotal downloaded samples from the top 10 IP addresses - Remaining top 9 hosts 57
4.16 81.231.232.16 landing page example 60

List of Tables

2.1 Destination URL classifications 16
2.2 Cryptojacking domains based on website categorisations 23

3.1 Top 10 encountered mining scripts 28
3.2 Top 10 Alexa rank sites with cryptomining scripts 29
3.3 Top 10 counties hosting cryptomining scripts by IP address count 30
3.4 IP address geolocation data compilation 30
3.5 Data points for enrichment 34

4.1 Top 10 countries hosting `feesocrald.com` cryptomining scripts 36
4.2 Top 10 countries hosting `jsecoin.com` cryptomining scripts 37
4.3 Top 10 countries hosting `freecontent.date` cryptomining scripts 37
4.4 Domain categorisations - ForcePoint 39
4.5 ForcePoint and BitDefender correlations 39
4.6 Top 10 endpoints of cryptomining domains 41
4.7 Top 10 IPS's hosting cryptomining domains 42
4.8 Top 10 most prevalent subnets (/24) 43
4.9 Assessed VirusTotal IP Addresses 56
4.10 Top 5 JSECoin user account prevalence 59
4.11 Top 5 JSECoin user account prevalence with balances 62
4.12 Top 10 JSECoin user account balances 63
4.13 Ranking of sites associated with top 10 user accounts by value 63

Chapter 1

Introduction

1.1 Background

Since the emergence of the Bitcoin open source cryptocurrency project in 2009, the cryptocurrency sector had collectively capitalised over 500 billion US dollars by 2017 (Eskandari *et al.*, 2018). The process of mining cryptocurrencies is a technical process used to incentivize participating nodes to assist in transaction verification as the transactions are recorded in the blockchain (Nakamoto, 2008). During the early years of Bitcoin (2010-2011), users could efficiently mine the cryptocurrency on personal computers using the CPU. Although technically mining Bitcoin using CPU power is still possible, it is no longer efficient (Reddy, 2019).

The second wave of mining utilised improved technology whereby users supplemented computer CPU power with that of GPU processing capability. The usage of JavaScript code on websites that utilised the CPU's of visitors to the site for the purpose of cryptomining was first noted by (Kroll *et al.*, 2013). Cryptomining is the process of solving computational math problems for cryptographic currency reward. This computational requirement is known as the "proof of work". The process is responsible for ensuring information authenticity each time a cryptocurrency transaction is processed. The actual process of mining involves competing with other cryptocurrency miners to solve the computational problems around cryptographic hash functions associated with the relevant block containing the actual cryptocurrency transaction data. This large increase in the processing power required to mine Bitcoin was mainly due to the emergence of ASICs (Application-Specific Integrated Circuits) as well as the increase in collective mining pools (Narayanan *et al.*, 2016). Occurrences of cryptomining scripts on websites were often implemented without the knowledge of the site visitors. This is made possible using embedded JavaScript within the websites source code that utilise the browser's resources

to mine cryptocurrency. Both the usage of cryptomining JavaScript embedded on sites as well as the second wave of mining capability using personal computing hardware became unfeasible in profit generation due to the rapid growth in processing power required to mine Bitcoin cryptocurrency. This is primarily due to the large increase in mining participants, increasing the mining difficulty. This grew to over 10^{21} hashes per second (Rauchberger *et al.*, 2018). This rate is the measurement of mining difficulty.

The first recorded mined Bitcoin block occurred on the 18th of July 2010 by the user ArtForz (BitcoinWiki, 2009). This block was deemed to have been mined by code developed by the aforementioned user. By mid 2011 various open source GPU based mining tools such as `guiminer`[1] were released and were being widely utilised by individuals. Such tools drastically increased the efficiency of cryptocurrency mining due to the hashing power of GPU's and in particular to the introduction of mining rigs. These rigs take advantage of the massively increased parallelizing made possible with multiple GPUs. By 2012 ASIC's specifically designed for the mining of Bitcoin were being manufactured and sold. This transition from GPU to ASIC based mining is where the current state of cryptocurrency mining remains today. As a consequence to the advancements in the state of Bitcoin mining, the hashing power of the Bitcoin network subsequently increased as did the mining difficulty (Huang *et al.*, 2014). As the technology utilised for mining advanced, so did the emergence of mining pools. A mining pool is a collective and collaborative group of individual cryptocurrency miners. Each participant receives a pro rata reward based on the proportion of work performed. The first recorded pool mined Bitcoin block was from the 16th of December 2010 (Rosenfeld, 2011). Mining pools do not amplify earnings, but they provide a steady stream of income as opposed to the ad-hoc large dumps of earnings from individual mining.

The concept of in-browser cryptocurrency mining started with Bitcoin in its first few years of adoption. This rise in JavaScript based Bitcoin miners revealed the use of the `JSMiner`[2] and `MineCrunch`[3] miners (Kauthamy *et al.*, 2017). Although optimised for JavaScript, `MineCrunch` was claimed to be one and a half times slower than contemporary CPU based mining applications (Kroll *et al.*, 2013). Although CPU mining had since become no longer competitive due to advances in GPU and ASIC based mining, it remains prevalent for botnet operators with vast swathes of CPU processing power at their disposal (Tahir *et al.*, 2017). As well as being relatively

[1] GuiMiner - GUI Miner for Bitcoin. `https://guiminer.org/`

[2] JSminer - JavaScript Bitcoin miner. `https://github.com/jwhitehorn/jsMiner`

[3] MineCrunch - Configurable cryptocurrency JavaScript miner. `https://github.com/Kukunin/webminer`

unprofitable, in-browser mining was adjudicated as being illegal by the New Jersey Attorney General's office in May 2015. This ruling was passed in accordance with the developers of the `Tidbit` browser based Bitcoin miner. The attorney general stated that no website can legally utilise an individuals computational resources without the basis for opting out (Hoffman, 2015). `Tidbit` agreed to cease operation via the terms of the settlement.

As various cryptocurrencies such as Monero, Ethereum and Bitcoin emerged as leaders within the market, browser based mining as a concept became less and less prevalent. The most common method of acquiring cryptocurrencies became purchasing them. There has since been a revival in the noted occurrences of browser based cryptomining of Monero and Bitcoin since 2017. This was due to increases in the value of various cryptocurrencies from mid 2017 into 2018. The Pirate Bay torrent search engine was seen to have experimented with browser based cryptomining as an alternate revenue stream (Hruska, 2017). The website for Showtime entertainment, `showtime.com`, was also discovered to be using browser based cryptomining JavaScript code. Showtime claimed the code was injected via a third party advertisement provider (Liao, 2017).

The term cryptojacking is used to describe this technique of surreptitiously mining cryptocurrency via a user's browser while that user visits a website. This is achieved with JavaScript code that is embedded within the websites source code that utilises the user's computational resources. Browser based cryptomining can cause noticeable computational performance degradation by utilising between 25 and 100 percent of the user's CPU. Other terms for this are coinjacking and drive-by mining. Browser based mining is technically a subset of cryptojacking (Dev, 2014). Most uses of cryptojacking apply to mining via an unwitting users browser, however, cryptojacking also applies to binary malware that mines a particular cryptocurrency. Such scenarios are indicative on compromised machines that have cryptomining malware unknowingly installed. The malware utilises the machine's CPU or GPU hardware for the purposes of mining cryptocurrency (Pastrana and Suarez-Tangil, 2019). The research conducted by Eskandari *et al.* (2018) in the the published paper "A first First Look at Browser-Based Cryptojacking" is related research and was used as a starting point for many of the research goals.

1.2 Problem statement

The rise of cryptocurrency value coupled with its intended design to isolate itself from the global financial system has given rise in attempts at accumulating it through illegitimate means. In-

browser cryptojacking (see Section 1.3) is one such means that has been utilised by malicious actors to obtain cryptocurrencies illegitimately. Assessing the extent of web pages utilising illicit cryptomining scripts will provide insight into the prevalence and variances of cryptominig scripts being utilised to accumulate cryptocurrencies. Such analysis will provide insight into the countries and Internet Service Providers that host the majority of websites containing cryptomining scripts, as well as the data around domain categorisation classifications of the identified websites.

1.3 Research Goals

The popularity of adblocking extensions installed in user's web browsers is indicative of the dissatisfaction with both the extent and intrusiveness of website advertising. The ease of its monetisation and invisibility has led to cryptomining being used as an alternative to advertising. These factors have further contributed to abusive implementations of cryptomining on websites, "cryptojacking" where users do not give explicit permission for their resources to be utilised for mining while visiting the website in question. Due to the lack of large-scale detection mechanism for cryptojacking and its infrastructure, technical characteristics and proliferation rate are not widely known (Caviglione *et al.*, 2016). As so little is actually known about the threat, the following are questions that require deeper analysis:

- How prevalent is illicit cryptomining throughout the Internet ?
- Which cryptocurrencies are preferred for cryptojacking ?
- Which Internet Service Providers and countries contain the most servers hosting cryptomining websites ?

1.4 Research Limitations

As a relatively new and rapidly evolving topic, a limited number of academic papers on the current state of cryptojacking prevalence were discovered. As such, there is a tendency towards web based references utilised in the research. It must be noted that the work done is at a given snapshot in time.

1.5 Document Conventions

The conventions that are adhered to within the remainder of the body of this document are as follows:

- Where mention is made of an application or service the URL for the associated website shall be noted at the bottom of the page as a footnote.
- Where any of the following variant names are quoted, they shall be referenced in the set font.
  - Filenames
  - Filetypes
  - DNS names
  - System paths
  - System commands
  - System components
  - System permissions
  - Malware variant names
  - Referenced strings

1.6 Document Structure

The remainder of the document is comprised of the following chapters:

- **Chapter 2 (Literature Review)** examines previous research related to the cryptojacking ecosystem, various documented cryptojacking attacks and previous research around its prevalence on the Internet.
- **Chapter 3 (Data Collection)** provides detail on the data collection process via the use of a source code repository, accessed via an API.
- **Chapter 4 (Data Enrichment and Analysis)** details the method used for enriching and analysing the research data. This study used various online repositories with Python Pandas and Graph Database tools to analyse and illustrate the research.
- **Chapter 5 (Conclusions)** concludes the paper with a summary of findings in relation to the research goals

Chapter 2

Literature Review

2.1 Introduction

This chapter provides the relevant background and insight into the current state of cryptojacking. The technical fundamentals of cryptocurrency mining are explained with particular regard to the illicit cryptomining threat. This includes context around JavaScript, Monero and the Coinhive cryptomining infrastructure and how it fits into the current threat model for both end users and web site administrators. Prior research into the prevalence of sites hosting JavaScript mining scripts is reviewed and discussed. The cryptomining threat for mobile Android devices is analysed as well as current detection strategies for web based cryptomining.

2.2 Cryptocurrency Mining Basics

Cryptocurrencies that are blockchain based rely on the embedding of transactions in series of blocks that are are both public and tamper proof. The continued working of the system requires new blocks to be constantly generated and appended in order to store transactions that are pending. This process of generating new blocks is mining. Those performing the mining are miners. The task of the miners is to solve a cryptographic puzzle. This puzzle solving is known as a "proof of work". The difficulty of a "proof of work" is adjusted dynamically in order to ensure that new blocks are produced at a constant rate (Romano and Schmid, 2017). This constant block rate ensures both tamper resistance and predictability. Due to this requirement, the difficulty for mining and solving the puzzles increases as more miners participate and compete with each other for finding blocks. When the "proof of work" meets the required difficulty, the newly mined block is linked to the previous block, resulting in a cryptocurrency reward to the miner in exchange for the computational power contribution (Anjum *et al.*, 2017).

The recent rise in popularity with regard to cryptocurrencies has led to a significant increase in mining difficulty. The consequence of the difficulty increase is a need for even faster hardware to effectively mine blocks. The economic challenge of mining profitability requires analysis of not only the hardware expense but also the energy costs required to run the hardware (Raikos, 2019). Increases to the hardware requirements for earning capability has led to various hardware solutions being utilised in order to increase speed and computational power, these include (Draghicescu *et al.*, 2018) GPUs, FPGAs and ASICs. An alternative consensus mechanism to "proof of work" is "proof of state". The "proof of stake" implementation addresses the issue of resource intensive mining requirements by aligning mining capacity to the proportion of coins held by the proposed forger. As no block reward is awarded in "proof of state", the transaction fee is awarded to the successfull validator (Bentov *et al.*, 2014).

2.2.1 Mining Pools

Combining the mining resources of individuals into mining pools has also proved to be a popular method to reduce expenses by sharing the hardware and its running costs and consequently sharing the revenue earned for each block that has been newly mined (Berecz and Czibula, 2019).

The computational power of computers running web browsers can be utilised to perform cryptocurrency mining. This provides an alternate stream for web site owners to monetise visits to their site. Via the inclusion of embedded code within the pages of a website, the site visitors CPU resources can be utilised for mining upon visiting the site. This mechanism for an alternate revenue stream for site owners can be implemented either with or without the consent of the user visiting the site. The later being illicit cryptomining or "cryptojacking" (O'Gorman, 2018). Bitcoin browser miners do exist, however, the large imbalance from a performance perspective results in Bitcoin based browser mining as being highly inefficient. This is mainly due to the large disparity between ASICs, CPUs and GPUs in terms of mining. Browser based cryptocurrency mining requires currencies with "proof of work" functions that are computable on CPUs and not GPUs or ASICs (Krishnan *et al.*, 2015). These are memory-hard "proof of work" computations and require frequent reading and writing from memory and are therefore suited for low latency memory on chip traditional CPU's and not GPUs or ASICs. These memory-hard "proof of work" computations are well suited for the micro computations that take place in a web browser and therefore are ideal for "in-browser" cryptojacking.

2.2.2 Monero

The privacy centric cryptocurrency Monero is an ideal candidate for browser based mining. Its "proof of work" is by design ASIC resistant due its periodic redesign and intensive memory requirements, thus perfect for CPU's where RAM is available. Its mining is thus CPU enabled, as is the requirement for browser based mining. Using the "Cryptonight"[4] hash function in its "proof of work", a new block is mined with a two minute block rate average (Bijmans *et al.*, 2019). The following steps illustrate the Monero blockchain and "proof of work" with mining input.

1. A Merkle tree[5] of the transactions that are to be included in the new block are constructed by the miner. It requires at a minimum the "Coinbase" transaction that is paying the miner with the block reward.

2. Nodes within the tree are constructed of the hashes of the child nodes with the hashes of the transactions making up leaves on the tree.

3. The root links of the tree, the final block and the transactions to the "proof of work" are included in the tree leaves.

4. The goal of the miner is then to locate a nonce[6] where the "proof of work" output matches the global difficulty.

5. The miner is required to re-draw a new nonce and continue recomputing the hash until the goal is achieved.

6. Verification of the proof can be verified by the network with a single round of hashing.

7. By the block being included in the blockchain, the miner is rewarded via the Coinbase transaction with the block reward.

This process is illustrated in figure 2.1. When mining pools are being utilised, the pool pushes out jobs that contain the "proof of work" inputs while requesting that participating miners find

[4]CryptoNight is a "proof of work" algorithm. It is designed to be suitable for ordinary PC CPUs. `https://en.bitcoin.it/wiki/CryptoNight/`

[5]Merkle tree is a tree in which every leaf node is labelled with the hash of a data block, and every non-leaf node is labelled with the cryptographic hash of the labels of its child nodes. `https://en.wikipedia.org/wiki/Merkle\_tree/`

[6]Nonce is an arbitrary number that can be used just once in a cryptographic communication. `https://en.wikipedia.org/wiki/Cryptographic\_nonce`

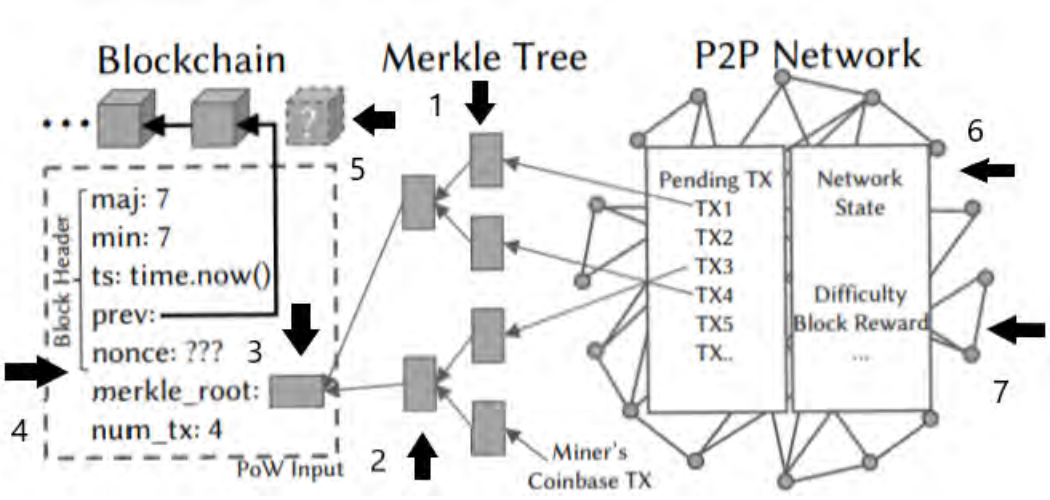

Figure 2.1: Monero blockchain and "proof of work" with mining input (Rüth *et al.*, 2018)

a nonce that satisfies a difficulty lower than the total network's. Once this lower difficulty is satisfied, the share is awarded to the miner. The share being of the final block reward.

2.3 The Monero and Coinhive infrastructure

Coinhive's mining services provided the platform for the mining of Monero[7]. As such, it is of particular relevance in this research. Part of Monero's allure is the supposed increased privacy facilitated by the total obfuscation of all participants in a transaction and the transaction amounts. This extended privacy capability differs from the other leading cryptocurrencies such as Bitcoin and Ethereum where a complete transactional graph can be reconstructed via the public blockchain (Musch *et al.*, 2018). Monero continued to grow and gain popularity during its early years after launch. This growth lead to the redevelopment of browser mining amongst the community of developers. September 2017 saw the launch of Coinhive[8] (Meland *et al.*, 2019). Coinhive established itself as the market leader for both illegitimate and legitimate browser based mining services by releasing additional services such as short-links and CAPTCHAs used to prevent illicit use while mining Monero. Various other competitors have joined the market as legitimate API providers to developers for the purpose of monetising browser based CPU mining for users who visit their sites. A percentage of the mined cryptocurrency is kept by the API provider while the rest is kept by the website owner (Kshetri and Voas, 2017). Without explicit user consent, in-browser mining is considered abuse.

[7]Open source cryptocurrency that focuses on privacy and decentralisation. `https://www.getmonero.org/`

[8]JavaScript miner for the Monero Blockchain. `https://coinhive.com/`

2.4 Ethical Discussions

The concept of cryptojacking and its variety of deployment scenarios raises various ethical questions around its use. The deployment of in-browser cryptojacking scripts on a web site is most likely due to one of three different scenarios (Razali and Shariff, 2019):

1. The deployment of cryptojacking scripts following a breach of the system.
2. Deployment by the webmaster without any consent requests to the users of the site
3. Deployment by the webmaster followed by in-browsing consent request to users of the site

While item number one is certainly unethical, point three is debatable from an ethical perspective as even though a user may consent to the mining, they may not understand the implications of their consent. It is extremely complicated to deduce whether or not a user understands or not what exactly he or she is consenting to. Furthermore, it is as unclear if the user then understands what exactly they receive in return and whether or not it is therefore a fair exchange. The following are examples of what users may receive in return for allowing in-browser cryptomining are (Zhao *et al.*, 2014):

- Video streams with improve quality, such as High Definition
- Content that would otherwise only be available to Premium subscribers
- The removal of advertisements

Research conducted by the technology web site Bleeping Computer in 2017 (Cimpanu, 2017) revealed that a large percentage of users accept web sites using their resources to mine cryptocurrency in the background if advertisements are no longer displayed to them. The Pirate Bay[9] torrent search engine was caught out using cryptomining scripts without any amendments to their privacy policy. After these details were publicly released, site administrators ultimately removed the code with the caveat response of "do you want ads to display or do you want to give up a few CPU cycles?" (Hruska, 2017). In both the keyword and auction based models of online advertising, the advertisement publisher is paid by the advertiser to distribute the advertising content. The owner of the website on which the content was displayed is paid a fee by the advertising publisher. As a replacement monetization strategy, in-browser mining is a more direct compensation system with fewer intermediaries, thus potentially benefiting site owners as well as end users (Edelman *et al.*, 2005).

[9]ThePirateBay `https://www.thepiratebay.org`

The trade off in terms of not being subjected to advertisements on the website versus harm to the end users taking part in cryptojacking include the following (Darabian *et al.*, 2020):

- Poor system performance
- Slow substandard web experience caused by the heavy utilisation of system resources diverted to cryptocurrency mining
- Higher electricity bills
- Accelerated hardware degradation from continued high utilisation during cryptocurrency mining

Even if consent is obtained for in-browser cryptomining from the end-user, it is unclear if the user is exactly aware of how they paying. Although, the same can be questioned with regard to users consenting to tracking cookies via a display banner (European Commission, 2011).

2.5 Threat Modelling

The attack surface to implement such abuse is vast due to the varying attack vectors that could implement mining scripts within hosted web sites (Hajri *et al.*, 2019). These five attack vectors are summarised as the following:

1. **Third Party Services**: Third party providers serve JavaScript to many websites in the form of tracking tools, analytic services and advertisement syndication. These third parties could knowingly inject mining JavaScript code or themselves be targeted and breached with the intention of deploying such mining code across their clients. Various instances of such mining has occurred across clients such as Movistar, Youtube and others (Liu and Chen, 2018).
2. **Breaches**: Should an attacker be able to breach the servers hosting an Internet facing web site, they could inject cryptomining scripts that mine a cryptocurrency such as Monero from users who visit the site. Such attacks were noted on the websites for both Tesla and the Los Angeles times (Dunn, 2018).
3. **Man-in-the-middle**: Any clear text traffic routed through intermediaries such as wireless routers or upstream network hardware could have malicious JavaScript injected into the non-HTTPS traffic. This was discovered at certain free Wi-Fi access points at Starbucks in Argentina as well as at the Marriot Courtyard hotel in Times Square New York (Hollister, 2012).

4. **Webmaster initiated**: An insider attack whereby as a website administrator, a mining script could be added to the web site without the knowledge or consent of the site's users. This is often implemented to supplement revenue streams.

5. **Browser extensions**: The concept of cryptojacking is not limited to websites. A chrome extension named Archive Poster surreptitiously mined cryptocurrency from thousands of their user's who had installed the extension (Hackett, 2018).

2.5.1 Malicious mining infrastructure

The research by Hong *et al.* (2018) noted that at least four components are required in the distribution of malicious mining scripts.

1. **Attacker**: These are the malicious actors who utilised client systems as mining infrastructure to generate profit. The scripts running on the client systems are configured with their unique wallet identification numbers so that they are rewarded for the mining activity.

2. **Miner Deployer**: Are the servers or domains that host the actual scripts that perform the cryptocurrency mining. The scripts are either custom written or commonly available from public sources.

3. **Mining Pool**: Are servers or domains that participate in distributed mining tasks. These include tasks such as verifying hashes.

4. **Distributor**: The intermediate domains that perform redirection to the destination containing the actual mining scripts are the distributors. These domains are typically changed frequently as to avoid downtime due to blacklisting. A proxy server is a valid example of a distributor.

A working example of these parties can be portrayed in the scenario whereby a victim browsing a page containing a malicious mining script would as follows:

1. The actual mining script containing the malicious code to implement mining would be obtained from the Miner Deployer configured by the Attacker.

2. Distributors could be utilised through domain redirection so that the URL of the Miner Deployed can be changed at any point. This aides in detection avoidance.

3. The actual mining tasks are assigned by the Mining Pools. These pools generate the actual revenue that is then paid to that attacker.

2.6 Browser Mining Prevalence

The research conducted by Rüth *et al.* (2018) utilised zone file databases in order to conduct analysis on the prevalence of sites hosting JavaScript that conduct cryptocurrency mining. The dataset included the Alexa top one million sites and the zone file databases for the `.com, .org and .net` top level domains. By downloading a pre-determined size of each homepage (256 kilobytes) of each site within each of the zone file databases, the retrieved output was then analysed for known cryptomining URLs using the adblock nocoin list as an input. The results provide detail as to which mining services have the largest share of sites using mining code as well as categorisation of all sites that include the mining code.

The aim of this research is to build on the work done by Rüth *et al.* (2018) in order to ascertain the current state of cryptomining within browsers. It is of particular relevance as Coinhive is no longer operational, as per section 2.2. As Coinhive was up until its closure, the largest Monero mining service (Varlioglu *et al.*, 2020). This research aims to uncover how the market has since changed and which if any services have filled the void left by Coinhive.

Coinhive ran the most widely utilised mining service. The service was provisioned via an optimised JavaScript miner for Monero. The monetisation of this service was in the form of a 30% fee of all the mined Monero. The Coinhive process for user registration and subsequent usage of the service was as follows:

1. A user who requires access to the Coinhive mining service registers on the Coinhive site.
2. Upon successfull registration, the user is provided with a token that is unique.
3. The unique token is used in configuring the necessary API calls used for the mined shares.
4. Upon a user visiting the website containing the Coinhive JavaScript miner, the miner is loaded when the page renders, connecting back to the Coinhive pool. This process authorises the Coinhive user‘s token for receiving input for hashing.
5. Once a hash has been found, it is then committed to Coinhive‘s pool.
6. Coinhive then pay their registered users 70% of the block reward. The remaining 30% is kept by Coinhive as payment for the service.

Figure 2.2 illustrates the dominance of Coinhive as the leading provider of mining services. The blue bar clearly indicating the large usage of the Coinhive service across the .com\.net\.org

and Alexa top one million datasets. The Authedmine[10] service was operated by Coinhive in addition to their primary mining service. The addition of the authedmine service provided a mining service for cryptomining with end user permission (Kharraz *et al.*, 2019).

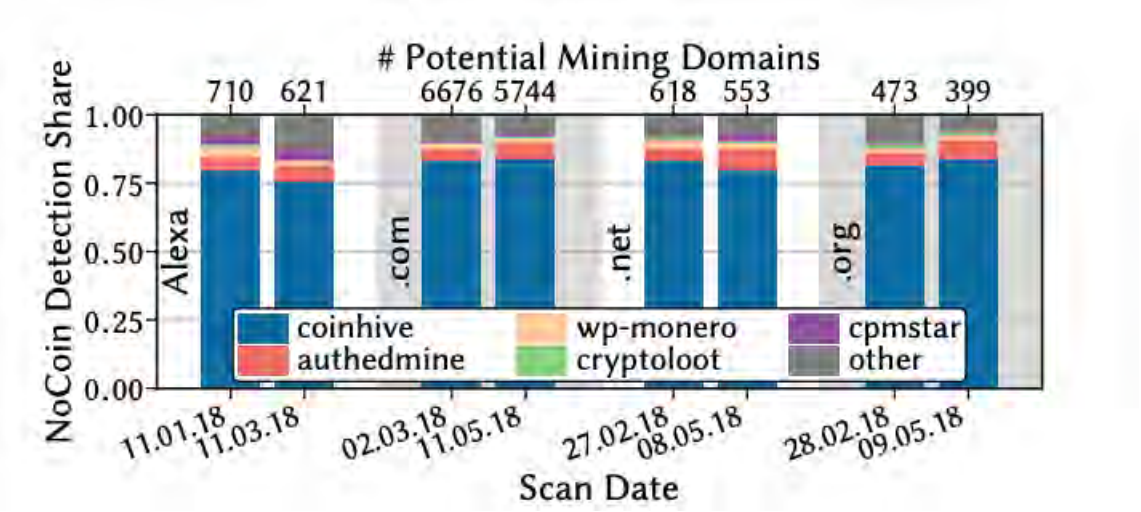

Figure 2.2: Coinhive prevalence in browser mining
(Rüth *et al.*, 2018)

Apart from the API services offered, Coinhive provided two additional services. Namely, a service for short link domain forwarding as well as a Captcha service.

2.6.1 The Coinhive Captcha service

The Coinhive Captcha service was shorted lived as it was blocked at both a DNS and ISP level soon after launching (Hohlfeld, 2018). The Captcha service behaved similarly to the Captcha service offered by Google, "reCaptcha" whereby the user is prompted to tick a box confirming that they are not a robot. Coinhive's Captcha required a few seconds of the user's computer resources to be utilised for Monero mining. The rewards for the mining would be paid to the Coinhive user who utilised the Captcha service on their web site.

2.6.2 The Coinhive short link forwarding service

The short link forwarding service was similar in functionality with other short link forwarding services such as as "bit.ly"[11]. Other short link services delay the redirection in order to display advertisements as the monetisation strategy. Coinhive's short link service operated in a similar fashion, but instead of the delay to display advertisements, the delay was used for the local computer to compute a number of hashes before the link was ultimately resolved (Varlioglu

[10] Authedmine - Consensual Coinhive mining script `https://authedmine.com/`
[11] Bitly - link shortering and sharing service. `https://bitly.com/`

et al., 2020). The number of required hashes for completion was entirely configurable and determined by the creator of the short link. From a user's perspective a progress bar would be visible indicating the progression of hashes that have been successfully computed. Once all the hashes had been successfully computed, they would be sent to the upstream service as the progress bar visible to the user illustrates that it is complete. The user would then be redirected to the intended link (Sato *et al.*, 2019). The creator of this short link would then receive a share of the mined block reward from all users who visit the site and utilise the short link forwarding service.

The Coinhive short links followed a distinctive alphanumeric pattern of "[a-z0-9]" directly after the URL of "https://cnhv.co". It was further noted by Rüth *et al.* (2018) that the links were increasing incrementally as new short links were created. This created a scenario whereby all created short links could be discovered and subsequently downloaded as HTML for further analysis. Analysis of the short link service was conducted in February 2018. Four characters were used in fuzzing the distinctive alphanumeric pattern, resulting in 1 709 203 valid short links. By collecting the HTML code from all the enumerated Coinhive short links, analysis of the code was used to extract the following:

- The Coinhive tokens from all the individual link creators.
- The configured number of required hash computations set by the link creator that are required by the user to solve in order to access the link redirection.

The resulting analysis uncovered that a few users had created a large numbers of links. One third of all the created links had been created by a single user. Furthermore, 85% of all created links had been created by ten users. A graphical representation of the "heavy users" can be seen in Figure 2.3.

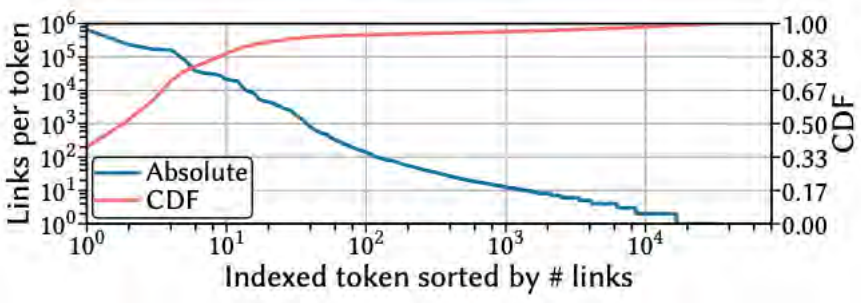

Figure 2.3: Tokens by number of links
(Rüth *et al.*, 2018)

In order to resolve the short link and be redirected the destination URL, the user is required to compute the pre-configured number of hashes that were set by the creator of the link (Allix *et al.*, 2016). Most of the links (85%) could be resolved in under 51 seconds, that being 1024 hashes. The heaviest user bias is 512 hashes, even when the user bias is removed, over two thirds of the hashes could be resolved in less than one minute.

A large number of the configured links required large amounts of time to resolve. Various users had configured the hash links to the maximum of 10^{19}. The network hash rate for the Monero network can be inferred from the difficulty combined with the rate at which blocks are found. Such a configuration would require billions of years to resolve.

Resolving all links was required to ascertain what kind of links are most commonly used as destination links for the Coinhive short link service. For this to be feasible, all links that required less and 10K hashes were computed. A random sampling of a thousand links for each of the top 10 users was extracted. This sample represents 80% of all the link destinations. As can be seen in table 2.1 (Rüth *et al.*, 2018), most of the links point to file sharing and streaming links.

Table 2.1: Destination URL classifications

Domain	Category	Frequency
youtube.be	Entertainment	20%
zippyshare.com	Filesharing	10%
icerbox.com	Filesharing	10%
hq-mirror.de	Ent and Music	10%
andyspeedracing.com	Automotive	10%
ftbucket.info	Msg Board	9.9%
getcoinfree	Finance	9.2%
ul.to	Filesharing	4.2%
share-online.biz	Filesharing	2.9%
oboom.com	Filesharing	2.8%

Detailed analysis of the Monero blockchain revealed which blocks were mined through the Coinhive service. Analysis over three months in 2018 (May to July) revealed that Coinhive mined 8.5 blocks a day in the median. This equates to 1.18% of the 720 blocks a day. Based on

these calculations, Coinhive earned 1,271XMR (Monero) per month. At the current exchange rate[12] of 96 US dollars for 1 XMR of Monero, Coinhive would have mined approximately 122 000 US Dollars worth a month (Nikiforakis *et al.*, 2014). Seventy percent of that revenue paid would have been paid to their users.

2.7 Detecting and Countering Web Based Cryptomining

Static and dynamic analysis can both be used in attempting to detect web based cryptomining. Both these techniques are detailed further in terms of application in the following subsections.

2.7.1 Static Analysis

The prevalence of Coinhive being the most widely used in-browser mining endpoint is corroborated by the search results for `coinhive.min.js` on PublicWWW[22], a search engine that indexes source code of publicly available websites. The existence of the term `coinhive.min.js` within the source code of a web site is indicative of it mining Monero cryptocurrency via the computational resources of visitors to the site, either consensually or not. Upon the initial release of the Coinhive JavaScript script, occurrences of it spiked but soon decreased as malware and organisations implemented measure for blocking access to the Coinhive website (Mursch, 2017). As a response to the blocking of its domain by enterprise organisations, Coinhive introduced an additional service domain name called Authedmine[23]. Authedmine requires users to consent before the in-browser mining is allowed. Although not receiving the same attention as the Coinhive service, Authedmine was seen to be gradually increasing its prevalence within the top one millions sites on Alexa. Other mining services such as Minr have employed randomised URL's in conjunction with obfuscated JavaScript to evade detection.

2.7.2 Dynamic Analysis

The majority of the discovered cryptojacking scripts have been noted by Sompolinsky and Zohar (2015) as using about 25% of the user's CPU resources. Such usage will be under the threshold of degrading performance and therefore unlikely to be discovered by the unsuspecting user. During the first few days after the Coinhive launch there were numerous reports of users CPU's reaching 100 percent utilisation (McCarthy, 2017). A factor in these high utilisation reports

[12] August 2019: 1XMR = 96USD

[22] PublicWWW - Online source code search engine. `https://publicwww.com/`

[23] Authedmine - Consentual Coinhive mining script `https://authedmine.com/`

is due to the Coinhive JavaScript library utilising all available CPU resources by default. The script throttling needs to be configured before being deployed to the web site in order to prevent occurrences of overly high user CPU usage.

The estimated revenue from in-browser cryptomining according to Coinhive developers is that of a monthly revenue of around 0.3 XMR (Monero), 46 US Dollars at the time of this writing, for a site with about 10 to 20 active miners. One of the biggest Coinhive script operators includes a domain parking with over eleven thousand parked domains (Bickford *et al.*, 2010). Visits to parked domains are generally short, resulting in approximately one hundred five thousand five hundred and eighty sessions at an average of 24 seconds a session over a period of three months. The in-browser mining for this period accumulated Monero to the amount of 0.02417 XMR in total (1.67 US Dollars as of March 2020) (Eskandari *et al.*, 2018).

Various cryptojacking tools have attempted to legitimise the action by obtaining user consent before commencing with mining. The Authedmine service from Coinhive is a legitimate example of such. There have been reports of abuse during the consenting process. As the consent is usually given via confirmation with a mouse click, reports have emerged of this process being vulnerable to clickjacking (Rydstedt *et al.*, 2010) attacks whereby users unknowingly consent.

Various discussions amongst browser developers have produced a variety of possible mitigations for in-browser cryptojacking, these include but are not limited to:

- Alerting users when client side scripts consume resources extensively.
- Throttling system resources used by client side scripts.
- Blocking known host names associated with cryptojacking via blacklisting.

The determination of the thresholds that classify processor usage as sufficiently high for the operation of legitimate applications is a problem open for research as is the threshold for processor usage low enough to dissuade in browser cryptojacking. The Opera browser has taken technical steps against the use of in-browser cryptojacking scripts by blocking them with via a blacklist they named NoCoin (Kolondra, 2018). In contrast, some browsers may promote the use of consensual in-browser mining, such as CryptoTab[24]. Its potential to monetise web sites independent of tracking cookies and advertisements make it a potential avenue of interest for browser development.

[24]CryptoTab - Web browser with builtin mining functionality. `https://cryptotab.net/en/`

The extensive research conducted by Saad *et al.* (2018) not only provides detailed analysis of cryptojacking code, both from a static and dynamic view point, but also an analysis of cryptojacking from an economic perspective in terms of viability when compared with advertising. This same economic analysis can be utilised as an indication as to the financial viability of cryptojacking from a criminal perspective. The analysis conducted by Saad *et al.* (2018) provides insight into:

- Content - In terms of the content of sites that were found to be hosting cryptomining software. This was noted as being widespread across various differing website genres and as such indicates a vast threat landscape.
- Currency - The currency analysis revealed various affinities between the cryptographic currencies and mining platforms. Coinhive was found to be the most widely used mining platform while Monero was the most widely mined cryptographic currency.
- Code - The code analysis revealed unique complexity features within the code of the cryptographic scripts. These same complexities were then used to identify cryptojacking code embedded within malicious and benign code samples.

The dynamic analysis highlighted the impact on system resources caused by cryptojacking. This includes CPU usage as well as battery usage on devices with a battery (Desnitsky and Kotenko, 2018). As a legitimate means for generating income instead instead of utilising online advertising, "in-browser" cryptojacking was not deemed feasible from a financial perspective (Saad *et al.*, 2018). The findings into both long and short term countermeasures for cryptojacking provided a platform for further reading and analysis of research in the field of cryptojacking detection. Although a vast amount of the research is countermeasure centric, the required detailed analysis of signatures, indicators of compromise and de-obfuscation techniques required for countermeasure research provides a great deal of data regarding the inner workings and behaviour of various cryptojacking instances.

The research conducted by Cova *et al.* (2010) presents a unique and novel method of detecting and subsequently analysing malicious JavaScript code. Their defined approach uses machine learning to provide a method to automatically identify the malicious JavaScript code via a systematic comparison of non-malicious JavaScript. The identified non-malicious characteristics are used as comparison benchmark allowing for the rapid detection of malicious JavaScript (Cova *et al.*, 2010). The research provides further insight into the capability of analysing obfuscated code as well as signature generation and detection for malicious code samples.

"*MineSweeper*: An In-depth Look into Drive-by Cryptocurrency Mining and Its Defense" by Konoth *et al.* (2018) is the research output of a comprehensive comparison of Alexa's top one million websites. The comparison details the profitability and prevalence of cryptojacking payloads delivered through "drive-by" mining. The "drive-by" mining attack utilise malicious JavaScript or web assembly modules to secretly induce a users browser to mine cryptocurrency unbeknown to the user visiting the site. The results of the comparison reveal multiple online services that cryptomining campaigns can utilise to generate funds from the nefarious cryptojacking payloads. The research further explores the techniques deployed to evade detection as well as how the latest web technologies are being utilised to effectively mine cryptocurrencies. These findings are subsequently utilised to discuss an array of ineffective countermeasures that include blacklisting and CPU usage heuristics. Their research introduces a novel technique used for the detection of silent browser based cryptomining by examining the intrinsic behaviour of cryptomining code.

Mineguard is the output of the research by Tahir *et al.* (2017). Its goal is the cessation of covert cryptomining software that causes large financial losses to organisations via substantial increases in power and cooling bills due to the constant GPU and CPU usage caused by consistent cryptomining. The findings from the research culminated in the release of the software. The output is the capability to create discernible signatures for the variety of crypto mining signatures. These signatures indicate the uniqueness of each algorithm and provide the software the ability to detect them effectively.

2.7.3 Deeper Analysis with CMTracker

Previous work by Cova *et al.* (2010) using static lists and the premise that 100% CPU utilisation is caused by cryptomining as shown by Carlin *et al.* (2018) does not uncover all instance of mining scripts. Such detective measures can be evaded by:

1. **Code obfuscation** - This will prevent keyword searches via static lists
2. **Throttling** - Miners configured to stay under or at a certain resource usage limitation could invariably evade detection that pursues resource usage as an indicator.

The research by Hong *et al.* (2018) resulted in the release of `CMTracker`. Its purpose is to automatically detect sites hosting cryptojacking mining scripts via a combination of a "hash based profiler" as well as a "Stack structure based profiler". These design inclusions were intended to provide a deeper level of robustness in detection that allow for more cryptojacking

scripts to be detected when compared with static based keyword searches. The premise of these two included profilers is to identify two common characteristics in cryptojacking scripts. That being that their workload includes are repeated and regular and include the computation of a hash based function.

2.7.3.1 The Hash based profiler

The "Hash Based Profiler" focuses entirely on monitoring for the core functionality of cryptocurrency miners, the proof of work system. By focusing on low level hash functions, various interfaces for hashing libraries were identified. In total nine of these libraries were identified from fixed set signatures from various commercial and open sourced cryptomining services. By then calculating the time that websites spend cumulatively hashing, the out-liars can be identified. An average site within the Alexa top 100 sites utilises less than 0.47% of its entire execution for hashing. Whereby cryptocurrency mining scripts utilise most of their time on hashing. As such, should a website spend more than 10% of its execution time on hash computations, `CMTracker` designates it as a cryptocurrency miner.

2.7.3.2 The Stack based profiler

The "Stack Structure Based Profiler" was implemented as an additional detection mechanism for the identification of cryptocurrency miners that employ a high level of code obfuscation techniques. As cryptocurrency miners execute heavy workloads repetitively, such a behavioural pattern if constantly repeated could provide for as an identifying factor within their execution stack. An average web site page is noted as rarely repeating the same call stack beyond 5.60% of its total execution time. As cryptocurrency mining is resource intensive, mining tasks are not usually placed at the main thread for execution upon the initial loading of a page, this is to avoid notice by the end-user. It is more likely that a single or even various dedicated threads will be created. `CMTracker` will therefore flag as a cryptocurrency any thread that periodically repeats its call chain and utilises more than 30% of the entire execution period in this particular thread.

This functionality provides a means for automatic identification of which scripts contain cryptocurrency mining payloads. Such a discovery does not automatically qualify the pages as malicious as certain websites commence mining after user consent. In order to differentiate between malicious and consensual pages hosting mining scripts, pre-defined keywords were searched for from extracted web page text. This included search for phrases such as "mining agreement". The analysis revealed that only 35 of the web pages discovered as hosting malicious

scripts were identified as consensual whereas the rest of the 85 3936 web pages were noted as containing non consensual mining scripts.

2.7.4 Sample Collection - for CMTracker

In order for the analysis to cover a large scale real world dataset, the Alexa top 100 000 websites and their sub domains were crawled by Hong *et al.* (2018)) in May 2018. As clandestine mining scripts require as much user browsing time as possible to be profitable, the assumption was that malicious payload injection tends to target the pages that would be most frequently accessed by visitors to the site in question. Additionally, external links from the homepages were also crawled for mining scripts as mining perpetrators tend to utilise the highly ranked pages to advertise the malicious pages containing the mining scripts thereby negating the need to embed the malicious scripts on the top ranked pages themselves. Visitors to the higher ranked sites are then lured to access the malicious pages by clicking links to the pages hosting the mining code. Time and resource limitations allowed for 20% of both the internal and external links to be followed. The sample collection process acquired 85 3936 unique web pages for the dataset used to locate pages containing cryptojacking code.

The `CMTracker` tool identified 2770 domains that contained cryptocurrency mining scripts from the aforementioned dataset. Of the identified domains, 868 of them were within the Alexa top one hundred thousand websites. The remaining 1902 domains were external links to sites other than the Alexa top one hundred thousand websites. This identification incorporated the traversal of 44 8660 unique domains linked to the top 100 hundred thousand. The highest ranked page within the Alex top one hundred thousand noted as hosting cryptocurrency mining code was `thepiratebay.org` that ranked at number 125.

Categories and Categorisation of the 2770 domains hosting cryptocurrency mining scripts by Hong *et al.* (2018) revealed that the "Art and Entertainment" and "Adult" categorised sites are the most prevalent, as can be seen in table 2.5. This supports the premise of extended user browsing time as a requirement for profitability. As these sites are primarily resources for downloaded pirated material as well as those hosting adult material, user engagement is generally longer due to the user searching for resources to download or view. This is in contrast to sites hosting the mining scripts on a landing page where a user is far less likely to be engaged for periods of time as long as when viewing "Art and Entertainment" as well as "Adult" categorised sites. In total, these two site categories combined accounted for 49% of all the discovered sites

(Hong *et al.*, 2018)

Table 2.2: Cryptojacking domains based on website categorisations

Website Category	# Websites with Scripts	Percentage(%)
Art and Entertainment	752	27.1
Adult	360	13.0
Internet and Telecom	323	11.7
Business	182	6.6
Game	180	6.5
Others	973	35.1
Total	2770	100

hosting cryptocurrency mining scripts.

2.7.5 Evasion Techniques

Three evasion techniques were noted by Hong *et al.* (2018) as being commonly applied. They are the following:

1. **Limiting CPU usage**: Although maximising CPU usage would increase profitability for the actor who has implemented the mining script, exhausting CPU resources would increase the likelihood of detection via automated means. Many cryptomining scripts contain functionality for throttling CPU usage. As varying CPU's run on various systems run at differing levels of CPU utilisation, identifying cryptomining scripts via CPU utilisation is difficult. Nonetheless, throttling is commonly configured to ensure that both automated detection mechanisms and overly poor user performance are avoided.

2. **Obfuscated code within mining scripts**: Various mining scripts use code obfuscation to hide their malicious intent. Code obfuscation techniques can aide in hindering both manual and static analysis. Varying degrees of obfuscation can be applied. The entire mining payload could be obfuscated or only the logic that performs the the mining.

3. **Payload Hiding**: Attackers have been noted hiding their malicious payloads within the libraries of third parties as opposed to injecting the payloads directly into web page source code. An example of such is when attackers use their own version of a widely used well

known JavaScript library, such as `JQuery.js`[26]. By appending the malicious code to the original code, the mining commences when `JQuery.js` is loaded.

2.8 Summary

Although various tools have been released to detect and prevent web based cryptomining, a mature infrastructure coupled with an ASIC resistant cryptographic currency has resulted in the web spread embedding of cryptomining JavaScript code within web sites. This has been implemented as both a method for supplementing income from web traffic, as well as a post exploitation means for generating income from compromised systems. The actual extent of web based cryptojacking in terms of number of systems and variants used remains an under researched area. This serves as the primary motivation for the research conducted.

[26] JQuery - JavaScript library designed to simplify HTML DOM tree traversal and manipulation. `https://jquery.com//`

Chapter 3

Data Collection and Enrichment

This chapter details the steps and techniques used to acquire and process the dataset used for analysis. This includes the commands used to access the PublicWWW[27] API endpoint for the initial dataset acquisition. The resulting dataset includes the counts of all cryptomining variants within the source code indexed on PublicWWW.

3.1 Research Method

The proposed data gathering process involves a two phased approach. The first phase being data accumulation via an online source code repository. The primary purpose of the accumulation is to search through the source code repository for all sites containing known cryptomining scripts. The accumulated list of sites containing cryptomining scripts will be the initial dataset.

The second phase of the data collection process involves data enrichment whereby the accumulated list of sites containing cryptomining scripts will be enriched with data such as:

- Geographic location
- Internet Service Provider details
- Website classification categorisations
- Historical data retrieved from VirusTotal

[27] PublicWWW - Online source code search engine. `https://publicwww.com/`

3.2 Initial Dataset Aquisition

PublicWWW was searched between the 7th and 11th of August 2019 for the existence of domains containing cryptomining scripts by utilising the Adblock nocoin block list[28] list of known cryptomining domains. The raw data was obtained from PublicWWW between the 7th and 11th of August 2019. The collected data was acquired to provide the initial dataset for analysis of empirical real world data.

The Adblock nocoin list was downloaded along with the purchase of a months‘ worth of API access from PublicWWW in order to access the data required for the data acquisition. Each entry of the Adblock nocoin 587 entries was systematically searched in PublicWWW via the API. The API endpoint was queried using a for loop that called the curl command to access and extract the raw data. The command and endpoint can be found in listing 3.1 with the API key redacted.

Listing 3.1: PublicWWW API curl loop

```
1 for url in $(cat noicoin.txt);
2 do curl -OJ https://publicwww.com/websites/%22/$url/%22/?export=csvu
3 &key=<redacted>
4 ; done
```

The PublicWWW database contained the source code for 556 116 024 web pages at the time of research during August 2019. The output of the initial research revealed 27 981 instances of cryptomining strings from the nocoin block list discovered within the PublicWWW database of web site source code. This does not indicate unique website occurrence as numerous websites were discovered to contain more than one instance of a known cryptomining script. In total 25 204 unique sites were flagged as containing a malicious cryptomining script listed in the nocoin block list. The collection process took around 20 hours over the course of 5 days. This collection process resulted in the accumulation of approximately 10MB of raw data for analysis.

The output resulted in a file in csv format for each searched nocoin entry. The file contained all websites that included the searched nocoin entry within its source code, as well as the sites rank according to PublicWWW. A raw output example can be seen in listing 3.2 containing the resultant output for the searched nocoin entry of `analytics.blue`. The pages containing the

[28] Adblock-nocoin-list - Block lists to prevent JavaScript miners. `https://github.com/hoshsadiq/adblock-nocoin-list`

string in its source code and the rank can be seen separated by a comma.

Listing 3.2: PublicWWW API output sample: `analytics.blue`

```
https://www.logo.com.tr/,127895
https://www.allaboutthejersey.com/,220859
https://www.gosquared.com/,228612
http://analytics.blue/,290742
https://www.blueport.com/,299475
https://kameleon.pics/,741306
https://clubdemo.com/,952365
https://www.nowherelan.com/,2102620
https://damgoodadmin.com/,2525094
http://vvolve.com/,4691722
http://expertnaya-ocenka.ru/,5002875
http://xn----7sbbdaxmh6bxb8ei.xn--p1ai/,5410329
http://www.hamradiooutlet.it/,5524475
http://nash-izberbash.ru/,5832607
http://seriesmegahd720p.blogspot.com/,6176947
http://hairstylemagz.net/,7105930
https://globas-i.ru/ru-RU/Home/Auth,7223001
http://www.happy-women.ru/,7698452
```

Of the 587 entries within the utilised nocoin block list from August 7th-11th 2019, 305 unique entries were noted in the collected dataset. The top 10 most prevalent entries are listed in table 3.1. These entries totalling 19 223 account for 76.81% of the total dataset of 25 204 sites containing cryptomining scripts. The entries are those of the domains hosting the cryptomining scripts that were discovered within the source code of the sites contained in the PublicWWW database.

Of the top 10 most prevalent cryptomining scripts, `coinhive.com` and `authedmine.com` make up a combined percentage total of 38,53% of the top 10.

Table 3.1: Top 10 encountered mining scripts

Rank	Mining Script	Instance Count	Percentage
1	coinhive.com	6164	32.06
2	feesocrald.com	3578	18.61
3	jsecoin.com	1708	8.89
4	exdynsrv.com	1655	8.61
5	hostingcloud.racing	1287	6.70
6	coin-hive.com	1232	6.41
7	authedmine.com	1223	6.36
8	monero-miner	832	4.33
9	coinpot.co	815	4.24
10	freecontent.date	729	3.79
	Total	19223	76.88
		N=25 204	

This is of particular relevance as the Coinhive mining service (which was also responsible for the running `authedmine.com`) is no longer operational as of March 2019. This is however a noted decrease in Coinhive's prevalence from January 2018. The sites containing scripts that point to `coinhive.com` and `authedmine.com` will function without hindrance, regardless of the embedded scripts failing to load. Although it is still the most noted entry, its total percentage of 38.53% is down from previous years when it was operational. The combination of `coinhive.com` and `authedmine.com` recorded a combined prevalence of between 75%-80% of all discovered cryptomining scripts from January 2018 to May 2018 (Rüth *et al.*, 2018). Their research was also conducted using the nocoin list.

My analysis of the discovered sites containing cryptomining scripts revealed that 390 of them were listed as being included in the Alexa top 1 million sites from August 7th-11th 2019. The Alexa rankings of the sites included ranged from 465 to 62 9452. The top 10 ranked URL's and their respective Alexa rank are shown in table 3.2.

The full list of 25 204 unique URLs with cryptomining scripts obtained from PublicWWW was processed on the 12th of August 2019 using Massdns[29] for DNS resolution from host names

[29]Massdns - High-performance DNS stub resolver for bulk lookups. `https://github.com/blechschmidt/massdns`

Table 3.2: Top 10 Alexa rank sites with cryptomining scripts

Rank	URL	Alexa Rank
1	thewhizproducts.com	465
2	thewhizmarketing.com	929
3	katfile.com	2536
4	ed-protect.org	5168
5	tudohd.com	6932
6	gamatotv.co	10796
7	pharmeasy.in	25194
8	songspk.mobi	25495
9	highestpayingfaucets.com	28794
10	megafilmeshdplus.org	30075

to IP addresses. The IP address resolution process took approximately 25 minutes to complete. This process resulted in the accumulation of 21 984 IP addresses. 3220 of the host names were unable to be resolved due to two noted errors, this accounted for 15% of the hostnames.

- 2064, 10% failed with NXDOMAIN, indicating the domain no longer exists
- 1156, 5% failed with SERVFAIL, indicating errors connecting to the relevant DNS server. Retries at resolving were attempted, resulting in the same number of failed lookup attempts.

The resolved IP addresses were then mapped to geographic locations using the MaxMind Geolite2 database[30]. The geographic location mapping of the 21 984 IP addresses revealed that servers hosting cryptomining scripts were physically located in 91 different countries. South Africa was noted as having 106 IP addresses hosting cryptomining scripts, ranking 25th of the countries hosting cryptomining scripts by IP address count.

The IP address count for each of the identified countries was tabulated to determine total numbers of IP address per country identified. Twelve countries were identified as having just one IP address hosting cryptomining scripts. The bulk of the IP addresses hosting cryptomining scripts were noted as being from a relatively small pool of countries. The top 10 countries with IP addresses hosting cryptomining scripts can be seen in the following figure. The United States of America is top of the list with more than the sum of the remaining top 10 countries and

[30]MaxMind Geolite2 - Free IP geolocation databases. `https://dev.maxmind.com/geoip/geoip2/geolite2/`

four times more IP addresses than the next country, Iran. The geographic determination of hosted IP addresses includes only the resolvable domain names. The percentage allocations of IP addresses to host country can be seen in table 3.3. These percentages are calculated from the entire dataset of resolved IP addresses, and not a percentage of the top 10.

Table 3.3: Top 10 counties hosting cryptomining scripts by IP address count

Rank	Country	Count	Percentage
1	United States	9256	42.10
2	Iran	2832	12.88
3	Germany	1583	7.20
4	Russia	1364	6.20
5	France	920	4.18
6	Netherlands	745	3.38
7	United Kingdom	524	2.38
8	Singapore	287	1.30
9	Sweden	285	1.29
10	China	271	1.23
	Sum	18076	82
		N=21 984	

The top 10 countries for hosting IP addresses with cryptomining scripts account for 82% of the dataset. In total, the top 10 countries account for 18 076 IP addresses in total. The remaining countries account for 1 337 IP addresses accumulatively whereas 2 580 IP addresses where not able to be mapped to a geographic location using the MaxMind Geolite2 database. This can be seen in table 3.4.

Table 3.4: IP address geolocation data compilation

Category	IP address Count	Percentage of Dataset
Top 10 Countries	18076	82
No Geolocation	2580	12
Remaining Countries	1337	6

3.3 Data Enrichment

The accumulated dataset of domains hosting cryptomining scripts from PublicWWW was further enriched via the use of the API service offered by VirusTotal for academic research. Each of the domains found to be hosting cryptomining scripts was queried against the VirusTotal API from the 16th of August 2019 until the 19th of August 2019. The results from the API provided an array of varying information per domain. This included but was not limited to:

- Domain categorisations
- Confirmation whether or not the domain exists in the VirusTotal dataset
- Known malware samples associated with the domain
- DNS and whois information

The API endpoint was queried using a for loop that called the curl command to access and extract the raw data. The command and endpoint can be found in listing 3.3 with the API key redacted.

Listing 3.3: VirusTotal API curl loop

```
1 for url in $(cat domains.txt);
2 do curl -OJ https://www.virustotal.com/vtapi/v2/domain/report?
3 apikey=<redacted>domain=$url; done
```

Whereas the output from the PublicWWW API provided a csv file containing all entries containing the searched phrase. The output from the VirusTotal API results in a single json file per searched domain. An example of such can be seen in listing 3.4 and 3.5.

Listing 3.4: VirusTotal sample JSON output - Example 1

```
"resolutions": [
  {
    "last_resolved": "2019-08-12 19:40:42",
    "ip_address": "104.31.86.134"
  },
  {
    "last_resolved": "2019-08-12 19:40:42",
    "ip_address": "104.31.87.134"
  }
 ],
 "subdomains": [
  "webdisk.films-list.com",
  "www.films-list.com",
  "mail.films-list.com"
```

The resulting accumulation of over 20 000 nested json files provided for both a rich and complex dataset. In order to effectively extract targeted data from each json file, the script in listing 3.6 was run against the entire raw dataset. The script parses all the json files in the current directory searching for the “id” and “ThreatSeeker” tags. A newline is then inserted between each discovered entry. All tabs are removed and only the lines containing “ThreatSeeker” are kept.

Listing 3.5: VirusTotal sample JSON output - Example 2

```
  "whois_timestamp": 1551499708,
  "response_code": 1,
  "verbose_msg": "Domain found in dataset",
  "Forcepoint ThreatSeeker category":
  "potentially unwanted software. compromised websites",
  "resolutions": [
    {
      "last_resolved": "2019-03-02 04:08:28",
      "ip_address": "154.220.38.203"
    }
  ],
  "detected_urls": [],
  "categories": [
    "parked",
    "potentially unwanted software. compromised websites"
  ]
}
```

Listing 3.6: JSON data enrichment script

```
for f in *.json;do grep -e ThreatSeeker -e id $f |tail -2 | tr -d
'\n'| awk '{print $0,"\n"}' >> results.csv ; done
cat results.csv | tr -d " \t" > results2.csv
sed -n '/^"Forcepoint/p' results2.csv > results3.csv
cat results3.csv |cut -f1,2,3 -d":" > results4.csv
```

This results in a workable data set matching the queried domain to the ForcePoint ThreatSeeker[31] domain categorisation group. The script was then rerun with "ThreatSeeker" amended to "BitDefender" in order to accumulate domain categorisations from both vendors for comparison. After the same sanitisation process, an additional dataset matching the queried domain to the BitDefender[32] domain categorisation group was created. These resulting datasets were merged with the PublicWWW dataset to create a dataframe containing data points listed in table 3.5.

[31]ForcePoint ThreatSeeker - Domain categorisation service. `https://www.forcepoint.com/`

[32]BitDefender - Domain categorisation service. `https://www.bitdefender.com/`

Table 3.5: Data points for enrichment

Data Points
Domain
URL
IP Address
ForcePoint Threatseeker categorisation
BitDefender categorisation
Miner in use
Geolocation

This data will be used to map domain categorisations to mining variants and geolocations in Chapter 4. A segment of the newly created enriched dataset, using Python Pandas and Jupyter Notebook can be seen in figure 3.1.

bitdefender_cat	domain	forcepoint_cat	url	rank	miner	ip	country
NaN	www.smoc-industries.fr	business and economy.compromised websites	http://www.smoc-industries.fr/	10000390	coinhive.com	213.186.33.2	France
NaN	www.aromatherapy.sk	shopping	https://www.aromatherapy.sk/	10007129	minera.js	37.9.175.8	Slovakia
NaN	suz.ir	uncategorized	http://suz.ir/	10011893	feesocrald.com	213.232.126.134	Iran
blogs	watchfreehd.net	sex	http://watchfreehd.net/	100142	exdynsrv.com	62.4.16.108	France
NaN	trzn.co	business and economy	http://trzn.co/	10017245	jsecoin.com	134.209.48.214	United States
NaN	cigoronline.com.ar	uncategorized	http://cigoronline.com.ar/	10020776	coinhive.com	186.138.139.19	Argentina
NaN	hoshinkidohapkido.ir	NaN	http://hoshinkidohapkido.ir/	10021731	feesocrald.com	213.232.126.134	Iran
parked	jeewandevelopers.com	uncategorized	http://jeewandevelopers.com/	10032105	coinhive.com	95.216.124.146	Finland

Figure 3.1: Example of content from enriched dataset

3.4 Summary

The data acquired in this chapter provided the platform for the analysis and interpretation performed in chapter 4. This includes the geographic data with the most IP addresses hosting cryptomining scripts and domain categorisations of sites hosting cryptomining scripts. The enriched dataset aligns the domain categorisation and geographic location data with the most prevalent cryptomining scripts. The finalised dataset, compromised of merging numerous individual datasets totalled 2.8MB. The dataset is available for download as a Python Pandas Dataframe from `https://www.dropbox.com/sh/ggbo2x1wc60l3ah/AAD6zMjktnTdDPNN2yM_feJMa?dl=0`. It is comprised of 24 528 rows and 10 columns.

Chapter 4

Data Analysis

This chapter discusses the outcomes of the data analysis and details the analytical steps performed on the data acquired and enriched in the previous chapter. The analysis included a geographic, domain categorisation as well as an ISP and subnet analysis of the top 10 mining variants. The chapter further includes graph network analysis and finally, a financial analysis of a mining variant.

4.1 Geographic analysis by mining variant

Analysis of the primary geographic locations of the top 10 discovered mining scripts revealed a similar pattern to the combined total. That being the United States of America being the country with the most IP addresses hosting the individual cryptomining scripts. Outliers to this pattern are:

4.1.1 `feesocrald.com`

2715 IP addresses are geolocated to Iran. Only 9 IP addresses were geolocated to the United States of America. This can be seen in table 4.1.

4.1.2 `jsecoin.com`

Sweden is second to the United States of America in terms of addresses that are able to geolocated. This can be seen in table 4.2.

4.1.3 `freecontent.date`

Both Turkey and Germany are second to the United States of America in terms of addresses that are able to geolocated. This can be seen in table 4.3.

Table 4.1: Top 10 countries hosting `feesocrald.com` cryptomining scripts

Rank	Country	IP Address Count	Percentage
1	Iran	2715	99.05
2	United States	9	0.33
3	Germany	5	0.18
4	France	4	0.14
5	Indonesia	2	0.06
6	United Kingdom	1	0.04
7	Poland	1	0.04
8	South Africa	1	0.04
9	Ukraine	1	0.04
10	Netherlands	1	0.04
	Sum	2740	99.96
		N=2741	

Table 4.2: Top 10 countries hosting `jsecoin.com` cryptomining scripts

Rank	Country	IP Address Count	Percentage
1	United States	716	49.96
2	Sweden	241	16.81
3	Germany	75	10.47
4	Russia	54	3.76
5	France	48	3.34
6	Hungary	43	3.00
7	Netherlands	40	2.79
8	United Kingdom	27	1.88
9	Canada	19	1.32
10	Singapore	16	1.11
	Sum	1279	89.25
		N=1433	

Table 4.3: Top 10 countries hosting `freecontent.date` cryptomining scripts

Rank	Country	IP Address Count	Percentage
1	United States	211	37.61
2	Turkey	53	9.44
3	Germany	52	9.26
4	Finland	50	8.91
5	Slovakia	40	7.13
6	Italy	26	4.63
7	Russia	24	4.27
8	United Kingdom	12	2.13
9	Netherlands	12	2.13
10	France	11	1.96
	Sum	491	87.47
		N=561	

4.2 Domain Categorisation Analysis

The dataset enriched with domain categorisation data was analysed to determine the prevalent domain categorisations across the domains identified as hosting cryptomining scripts. This was achieved by counting the occurrence of each categorisation type for each entry of the enriched dataset. Both the "ForcePoint" and "BitDefender" categorisations were analysed by counting each domain categorisation occurrence in the dataset. The "BitDefender" categorisation column was considerably less populated than the "ForcePoint" column with 21 514 compared to 4724 null entries. Although the "ForcePoint" column had far fewer null entries, its "uncategorised" group was by far the largest of all the listed categories, as can be seen in table 4.4.

The combined null entries and "uncategorised" domains totalled 17 038 "ForcePoint" domains without categorisation, 4476 less uncategorised domains than the considerably smaller "BitDefender" dataset. The "ForcePoint" categorisations counts with the uncategorised group removed as well as the "BitDefender" categorisation counts can be seen in figures 4.1 and 4.2.

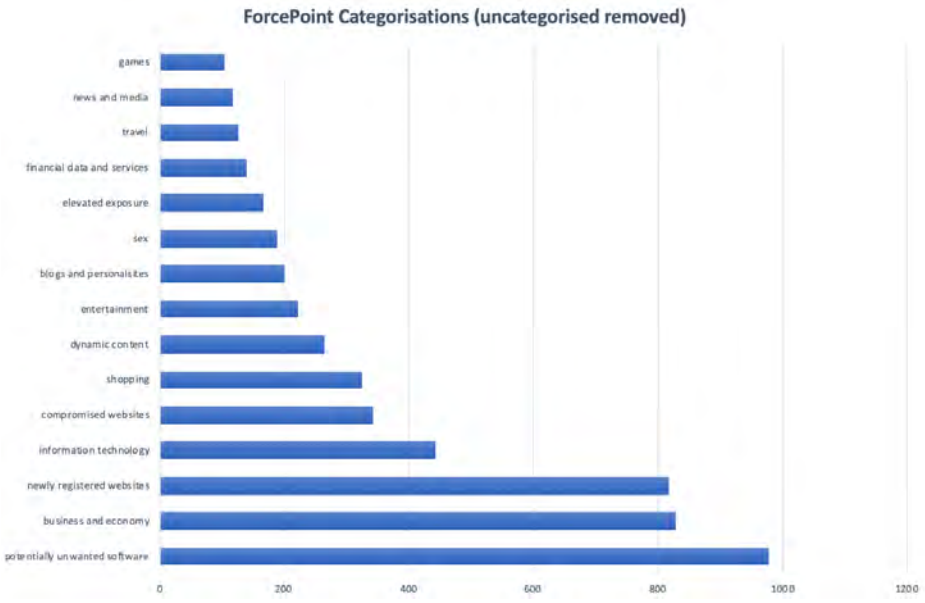

Figure 4.1: ForcePoint domain categorisations with uncategorised removed

As the large disparity in categorisation dataset sizes between "ForcePoint" and "BitDefender" is apparent, comparison based on categorisation counts alone is not sufficient. Various similarities between the categotisation datasets were noted in table 4.5 where categorisation percentages are compared.

Although the "ForcePoint" and "BitDefender" datasets are disparate. The categorisation

Table 4.4: Domain categorisations - ForcePoint

Rank	Category	Occurrence Count	Percentage
1	Uncategorised	12314	62.17
2	Potentially Unwanted Software	978	4.93
3	Business and Economy	829	4.18
4	Newly Registered Websites	817	4.12
5	Information Technology	442	2.23
6	Compromised Websites	342	1.72
7	Shopping	325	1.64
8	Dynamic Content	263	1.32
9	Entertainment	221	1.11
10	Blogs and Personal Sites	200	1.00
11	Sex	188	0.94
12	Elevated Exposure	166	0.83
13	Financial Data and Services	139	0.70
14	Travel	125	0.63
15	News and Media	118	0.59
	Sum	17 467	88.11
		N=19 804	

Table 4.5: ForcePoint and BitDefender correlations

BitDefender	BitDefender %	ForcePoint	ForcePoint %
Parked	18	Newly Registered Domains	16
Computers and Software	6	Information Technology	9
Porn	3	Sex	4
Travel	2	Travel	2
News	2	News and Media	2

classifications between datasets show similarities in the percentage classifications of each category. This serves to validate the domain classifications of the sites identified to be hosting cryptomining scripts.

- The “parked” and “newly registered websites” are a similar categorisation comparison

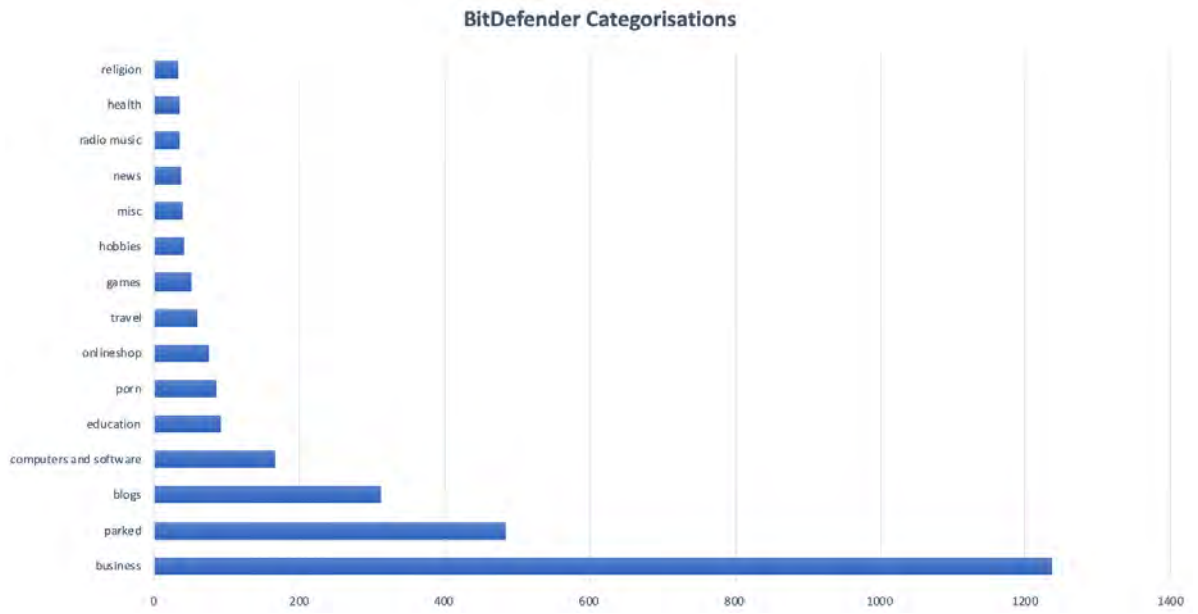

Figure 4.2: BitDefender domain categorisations

with similar percentages across both datasets at 18% and 16% respectively

- The "computers and software" and "information technology" are similar categorisation comparisons and also share similar percentages across the datasets at 6% and 9% respectively.
- Both the "sex" and "porn" categories are similar categorisation comparisons and have similar percentages across both datasets at 3% and 4% respectively.
- The "travel" category is at 2% across both datasets.
- The "news" and "news and media" are at a very similar 1% and 2% in both respective datasets.

4.3 ISP and subnet Analysis

Analysis of the dataset revealed the top 10 endpoints that the collected domains hosting cryptomining scripts are resolved to. Table 4.6 shows the most resolved endpoints based on a count of the resolved IP addresses (the list includes two google hostnames due to CNAME records pointing to other domains). CNAME records are DNS records that resolve a domain to another domain name, as opposed to resolving the domain name to an IP address. The CNAME records in positions 4 and 7 in table 4.6 are hostname entries for "Content Delivery Networks", or CDN's. These networks consist of geographically distributed servers that deliver web pages

and content to users. CDN's tend to service users from servers that are geographically located close to the user in order to facilitate faster delivery speeds. As these two hosts are part of the Google CDN network, the IP address endpoints cannot be determined by reverse lookups on the DNS name.

The counts were determined by counting the number of each IP occurrence in the dataset. The IP addresses were accumulated via the process of data enrichment using Massdns for DNS resolution of the collected URLs with known cryptomining scripts.

Table 4.6: Top 10 endpoints of cryptomining domains

Rank	Host	Count	Percentage
1	213.232.126.134	2709	11.04
2	66.96.161.196	238	0.97
3	81.231.232.61	228	0.92
4	ghs.google.com	216	0.88
5	109.108.145.100	116	0.47
6	62.210.16.62	111	0.45
7	blogspot.l.googleusercontent.com	108	0.44
8	159.65.245.16	92	0.37
9	178.128.243.171	92	0.37
10	206.189.153.135	92	0.37
	Sum	4002	16.28
		N=24 522	

Utilising the accumulated VirusTotal data, whois detail for each domain was collected and analysed. Table 4.7 illustrates the top 10 ISP's hosting IP addresses with cryptomining scripts. The table includes both the ISP name and the number of IP addresses as well as the percentage make up of the top 10.

Table 4.7: Top 10 IPS's hosting cryptomining domains

Rank	Name	Count	Percentage
1	Clouflare Inc	2565	21.03
2	GoDaddy LLC	713	5.84
3	Websitewelcome.com	399	3.27
4	Hetzner Online Gmbh	302	2.47
5	Namecheap Inc	222	1.82
6	Unified Layer	220	1.80
7	Africa Network Info Center	215	1.76
8	OVH SAS	192	1.57
9	Amazon Technologie	185	1.51
10	Asia Pacific Network Info Center	181	1.48
	Sum	5194	42.55
		N=12 192	

The results show Cloudflare[33] as the ISP with the most IP addresses containing cryptomining scripts with 21.03%. Cloudflare is a content delivery network designed to provide its clients application and network level security as well as the functionality to hide the IP address of the server hosting the client website. The remaining nine ISP's share the remaining 21.52%. Table 4.8 details the subnet count as well as the ISP that owns the subnet. ISP analysis allows for the identification of a potentially favoured ISP for sites with cryptomining scripts. This could be due to various factors, such as weak controls that allow mining code to be embedded without the ISP identifying it. It could also point to the potential breach of an ISP, whereby mining code could have been inserted into multiple sites at the ISP.

The intention of the subnet analysis is to determine whether any of the subnets contain a disproportionately large amount of endpoints containing cryptomining scripts. Such a scenario could indicate a breach at an ISP whereby servers on a subnet have been compromised and subsequently used to host cryptomining code.

Besides for OVH SAS, none of the top 10 ISP's hosting cryptomining domains contain a subnet within the top 10 most prevalent subnets in table 4.8. As discussed in section 4.3, the

[33]Cloudflare - Infrastructure and web site security provider `https://www.cloudflare.com/`

Table 4.8: Top 10 most prevalent subnets (/24)

Rank	Subnet (/24)	Number of IPs on subnet	ISP
1	217.160.0.0	102	1&1 Internet AG
2	46.30.215.0	69	One.com A/S
3	74.208.236.0	64	1&1 Ionos Inc.
4	87.236.16.0	56	Beget Ltd
5	92.53.96.0	45	Timeweb
6	81.169.145.0	45	Strato AG
7	90.156.201.0	44	MasterHost
8	192.0.78.0	42	Automattic Inc
9	31.31.196.0	41	Reg.Ru
10	213.186.33.0	39	OVH SAS

identified pattern between tables 4.7 and 4.8 appear to illustrate that the IP addresses hosted at the top 10 ISP's are somewhat sporadic and not within subnets populated with other systems hosting cryptomining scripts. This is in contrast to the smaller ISP's where various subnets that are largely populated with other systems hosting cryptomining scripts were noted. Two possible reasons for this discovery are proposed, namely:

1. Smaller ISP's were chosen by webmasters to host multiple sites all containing cryptomining scripts on the same segment network. This is due to the relative ease in obtaining an entire subnet from smaller ISP's.

2. The compromise of individual systems at the smaller ISP's led to further compromise of additional systems hosted on the same subnet. This allowed for malicious cryptomining content to be embedded across systems on the same subnet.

4.4 Graph Network Analysis

In order to further analyse the top 10 IP addresses and their relationships with cryptomining domains and the kind of miners being implemented, the dataset was imported into a Neo4j[34] instance. An example of a configured Neo4j node can be seen in listing 4.1. These properties are required as they will be used to match nodes to each other based on these included properties, thereby creating the graph database output that illustrates the relationships between nodes.

[34]Neo4j - Graph database management system. `https://neo4j.com/`

The blue nodes represent the IP address of the host being analysed. The green nodes represent the mining script variants whilst the red nodes represent the domains hosted on the IP address.

Listing 4.1: Neoj4 node example

```
{"country": "United States",
  "bitdefender_cat": "parked",
  "forcepoint_cat": "newly registered websites",
  "ip": "159.65.245.16",
  "domain": "comparegoodshoes.com",
  "url": "http://comparegoodshoes.com/",
  "miner": "hashing.win"}
```

Utilising the Cypher[35] query language, queries were written that detailed the association between country of origin and type of miner implemented for each of the top 10 endpoints listed in table 2.8. The reason for graphically analysing these cases is to determine and visually illustrate patterns in the strategies utilised to host either single or multiple variants of cryptomining scripts on the assessed infrastructure.

4.4.1 Case 1 - 213.232.126.134

Figure 4.3 illustrates the highest ranking IP address in terms of domain resolutions with 2 709 domains resolving to 213.232.126.134. All of the domains associated to the IP address implemented the `feesocrald.com` mining script. This is indicative of a targeted campaign whereby multiple sites are configured on a server implementing the same cryptomining script, potentially a web server that was compromised.

Figures 4.4 and 4.5 show a zoomed in view of both the blue and green nodes. This indicates a single IP address and single mining variant across all 2709 domains.

[35]Cypher - Neo4j's graph query language that allows users to store and retrieve data from the graph database. `https://neo4j.com/developer/cypher-query-language/`

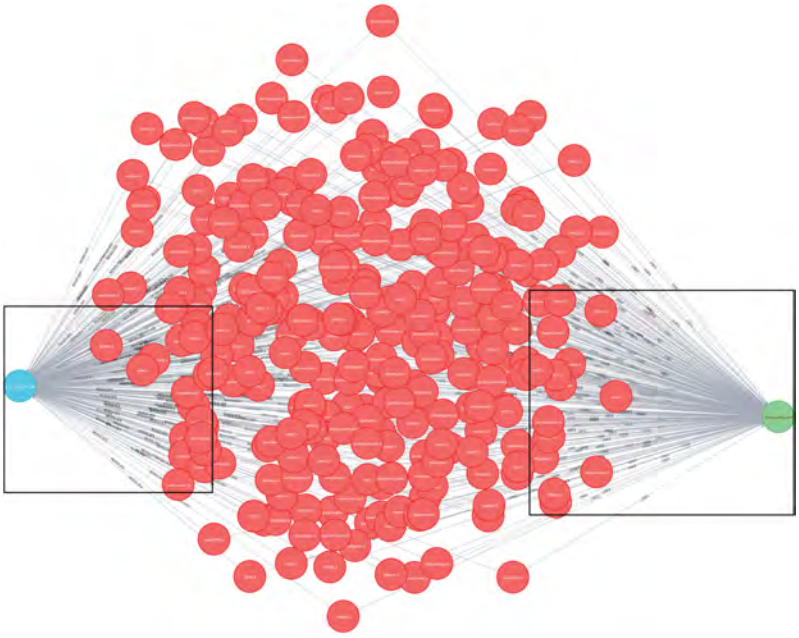

Figure 4.3: 213.232.126.134 endpoint analysis

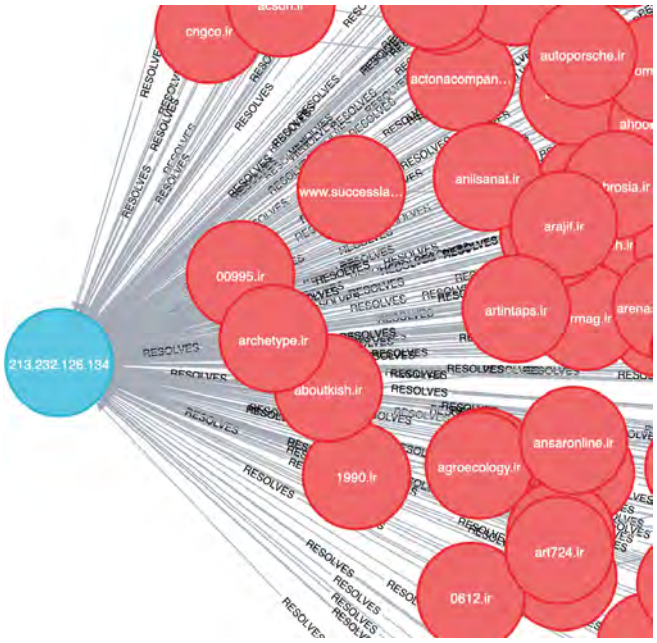

Figure 4.4: IP node analysis

4.4.2 Case 2 - 66.96.161.196

Figure 4.6 illustrates the graphic analysis of endpoint with the second most domain resolutions attributed to it, IP address 66.96.161.196. The IP address is geolocated to the United States of America and has a considerably less amount of domains that resolve to it compared to 213.232.126.134 with 238. Unlike 213.232.126.134 that has one specific cryptomining variant across all domains, 66.96.161.196 includes two different cryptomining variants, namely:

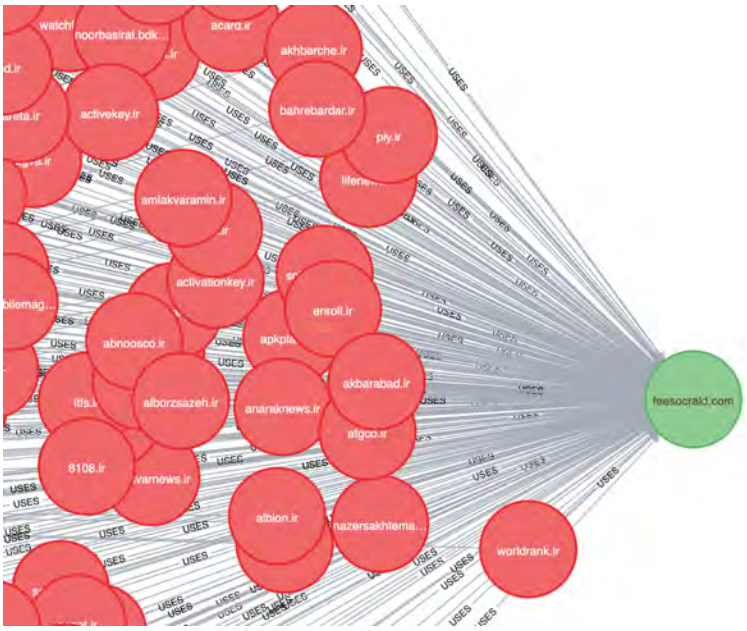

Figure 4.5: Mining variant node analysis

- `exdynsrv.com`
- `traffic.tc-clicks.com`

The green nodes represent the mining variants while the blue node represents the IP address of the endpoint. Each red node represents an individual domain hosted on the IP address.

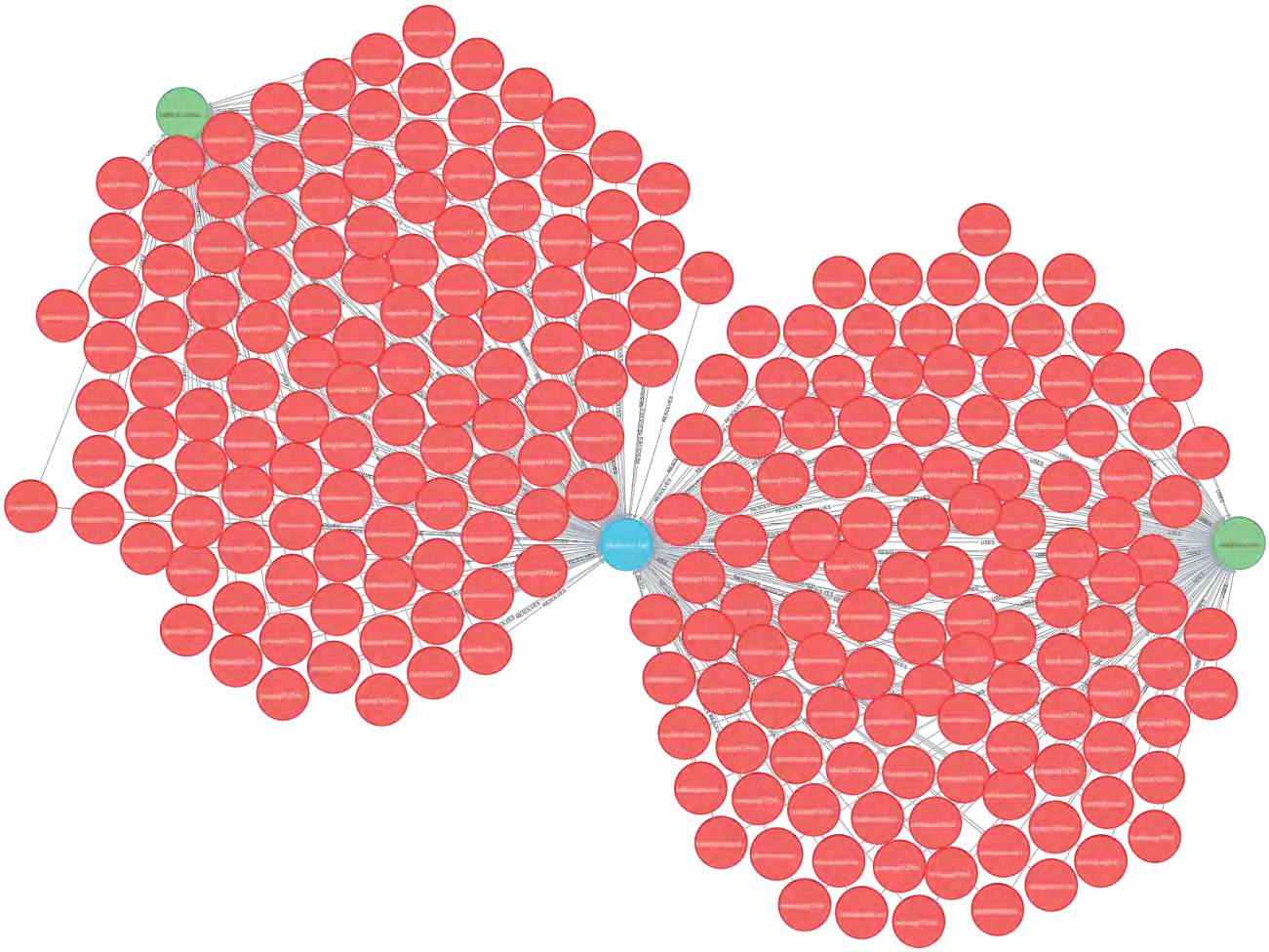

Figure 4.6: 66.96.161.196 endpoint analysis

4.4.3 Case 3 - 81.231.232.61

Analysis of 81.231.232.61, the third highest resolution endpoint with 216 domains also revealed a singular cryptomining script implementation, that of `jsecoin.com` hosted on a server in Sweden. This can be seen in figure 4.7. The single miner (blue node) used across multiple domains resolving to one IP address is indicative of a web server compromise. Everyone single domain node (red) is associated and thereby visually connected with both the blue and green nodes, representing the IP address and mining variant respectively.

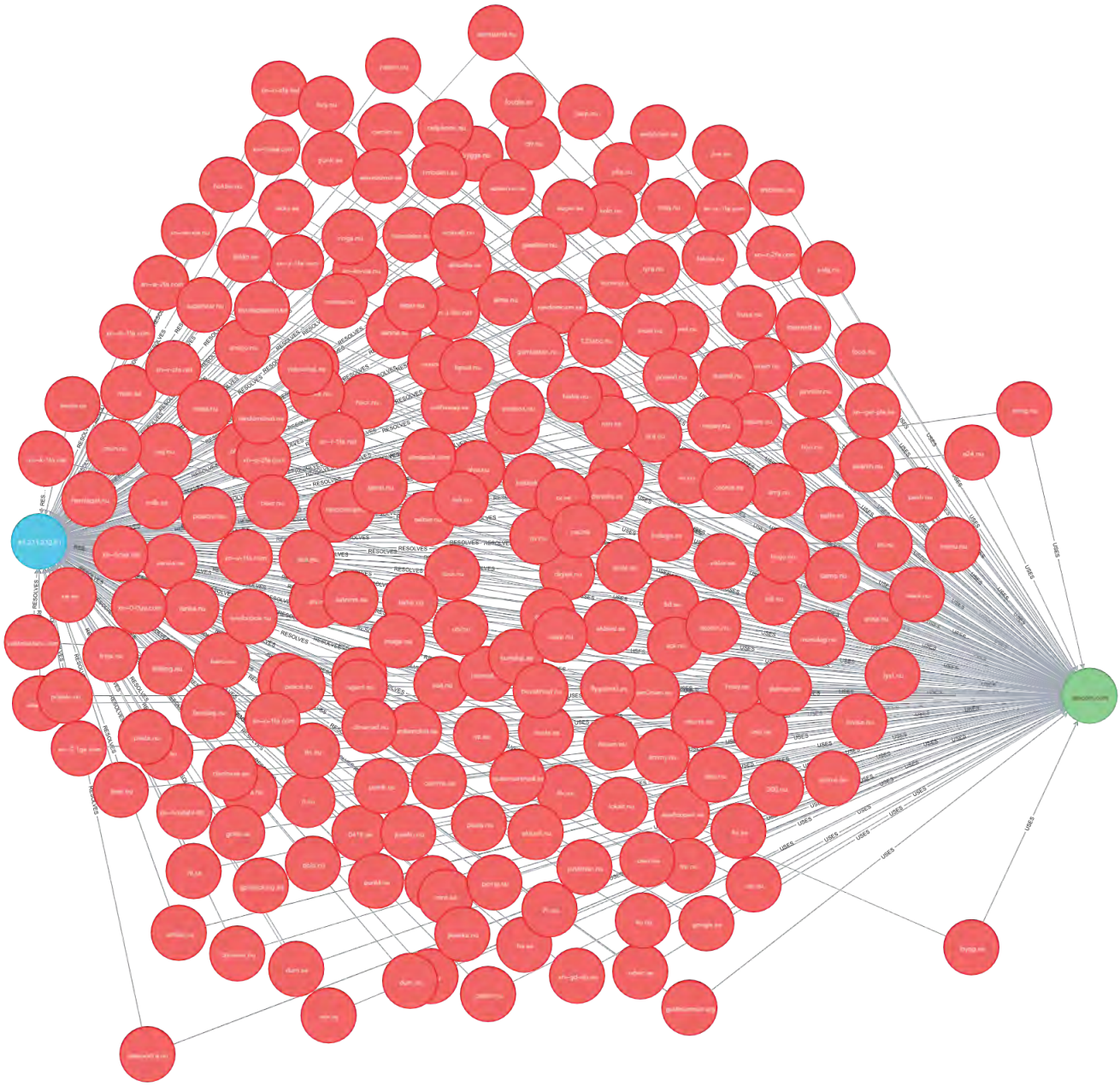

Figure 4.7: 81.231.232.61 endpoint analysis

4.4.4 Case 4 - `ghs.google.com`

The Google hosting services domain are hosted as opposed to one or two. This is expected as the `ghs.google.com` has 216 domains using it as a CNAME resolver, as discussed in section 4.3. As a CDN[36], the service masks the endpoint domain by using are hosted as opposed to one or two. This is expected as the `ghs.google.com` as the resolved address with the target site being included in the HTTP host header. As a result of this masquerading functionality that a CDN

[36]Content delivery network -geographically distributed network of proxy servers and their data centres. The goal is to provide high availability and high performance by distributing the service spatially relative to end-users. `https://en.wikipedia.org/wiki/Content_delivery_network/`

service offers, a vast array of different cryptomining scripts were discovered on a variety of different domains across the Google hosting services. This can be seen figure 4.8, a vastly different graph database compared to the first three analysed endpoints whereby multiple cryptomining scripts are hosted as opposed to one or two. This is expected as the `ghs.google.com` endpoint represents more than one single endpoint.

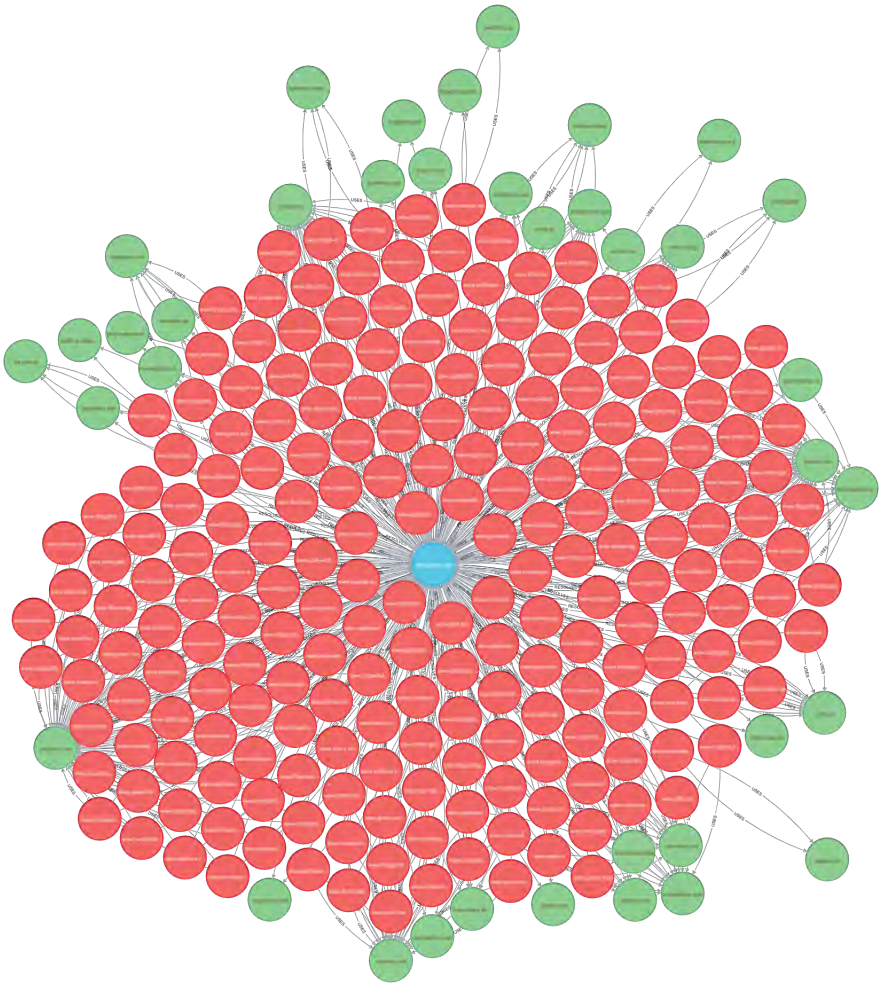

Figure 4.8: `ghs.google.com endpoint Analysis`

4.4.5 Case 5 - `blogspot.l.googleusercontent.com`

The Neo4j database graph for the `blogspot.l.googleusercontent.com` endpoint shares a common pattern to that of the `ghs.google.com` endpoint. It is the seventh most resolved endpoint with 108 domains that resolve to it via a CNAME resolver. It is also a Google owned domain that offers CDN services. An array of varying cryptomining scripts across the domains can be seen in figure 4.9. The large number of green nodes representing multiple mining code variants is expected due to the endpoint (blue) being a CDN hostname. The hostname endpoint is in fact masquerading for multiple web sites across varying geographic locations that are hosting differing mining scripts.

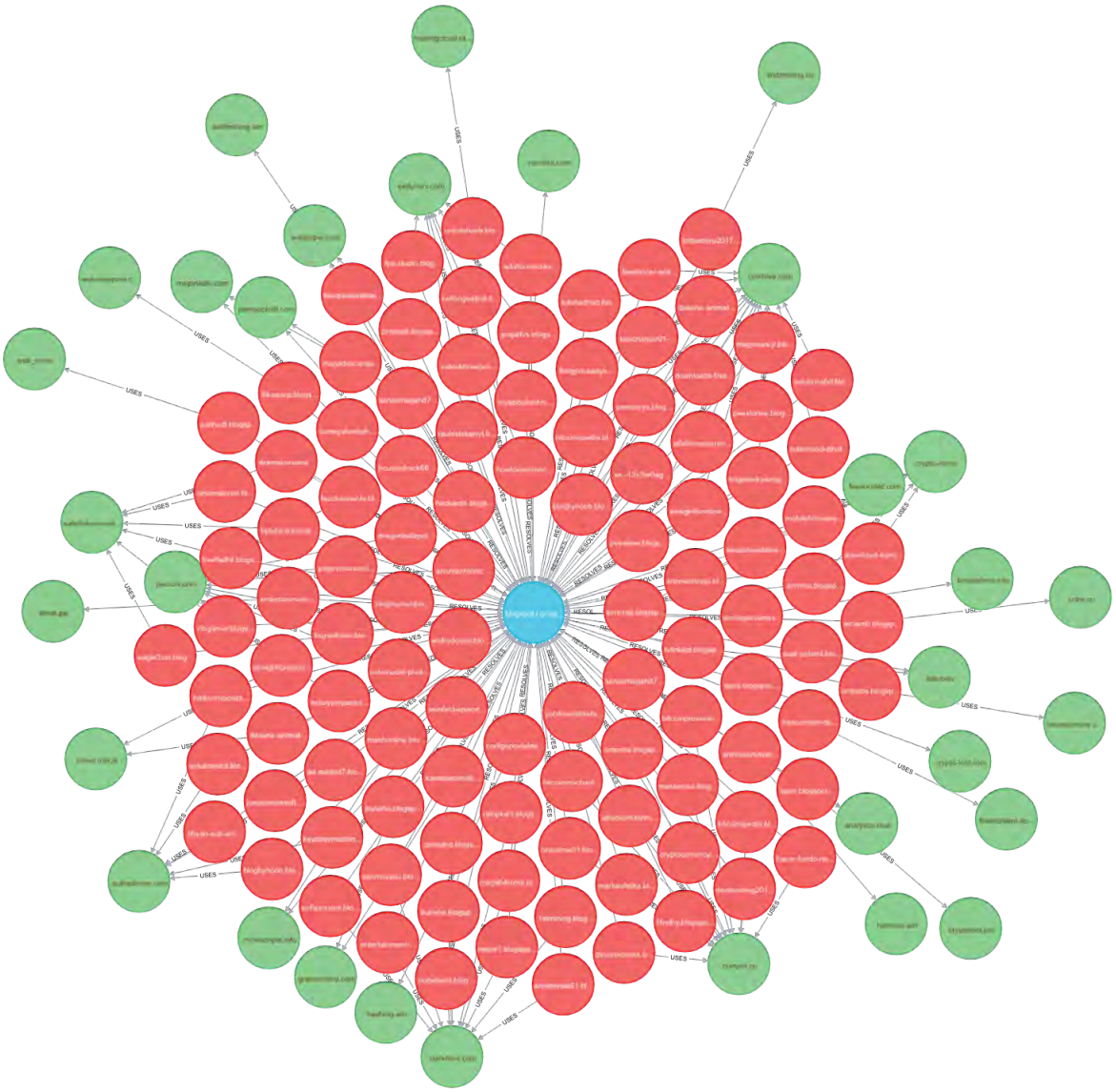

Figure 4.9: `blogspot.l.googleusercontent.com endpoint analysis`

4.4.6 Case 6 - 62.210.16.62

The remaining five endpoints in the top 10 list all follow the same pattern of one cryptomining variant for all domains that resolve to the IP, with the exception of 62.210.16.62, an IP address located in France with 111 domains that resolve to it. The majority of the domains utilise the `datasecu.download` cryptomining variant where a small number of other domains that resolved to the same IP address implemented the `coinhive.com`,`coin-hive.com` and `hostingcloud.racing` mining scripts, as can be seen in figure 4.10.

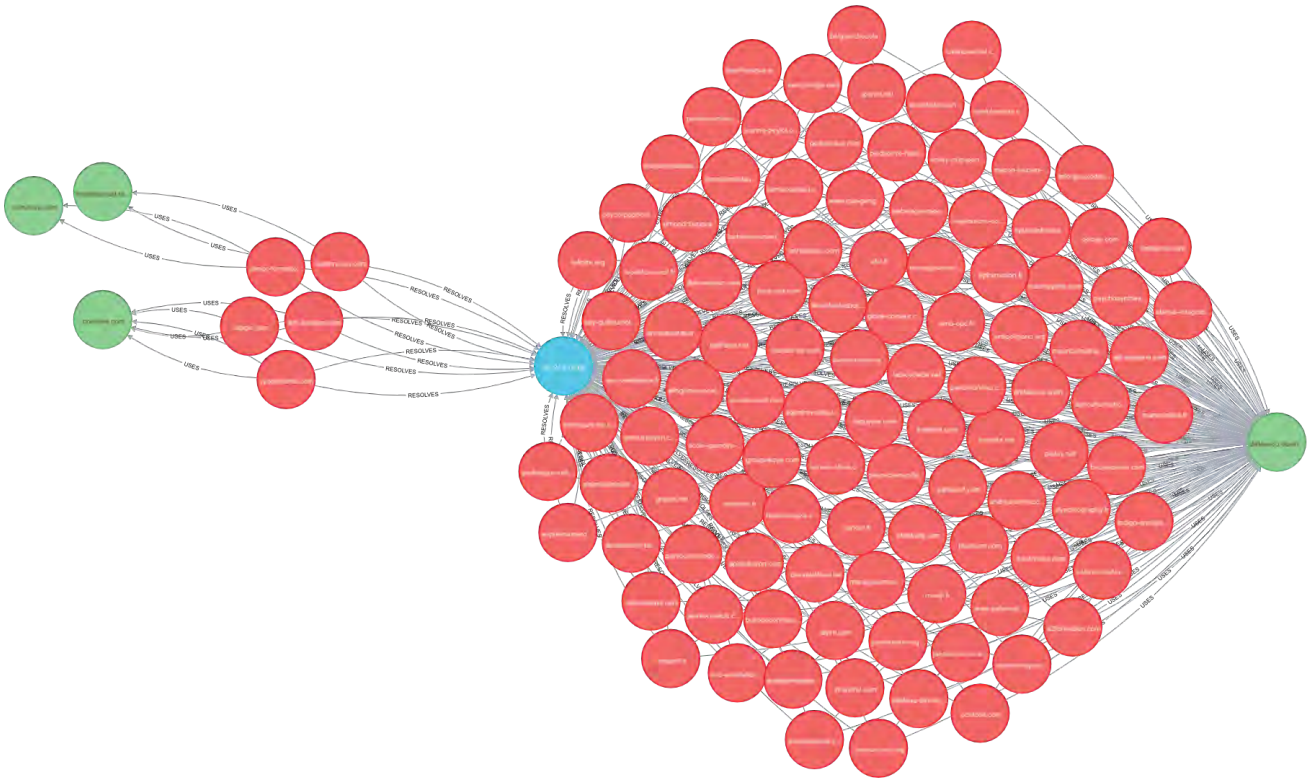

Figure 4.10: 62.210.16.62 endpoint analysis

4.5 Domain Categorisation Graph Analysis

The domain categorisation analysis serves to visually illustrate the domain categorisations of various sites hosting cryptomining scripts on two selected endpoints. Analysis on the IP address endpoints that contained a single cryptomining variant revealed a similar pattern across five of the six instances, that being that the majority of domains were uncategorised by ForcePoint except for a few outliers. This can be seen in figure 4.11 where a segment of the ForcePoint categorisations for the domains hosted on 206.189.153.135 are illustrated. This outcome is not unexpected due to the transient nature of domains hosting malicious scripts. The domains are often temporary and not in existence long enough to be categorised.

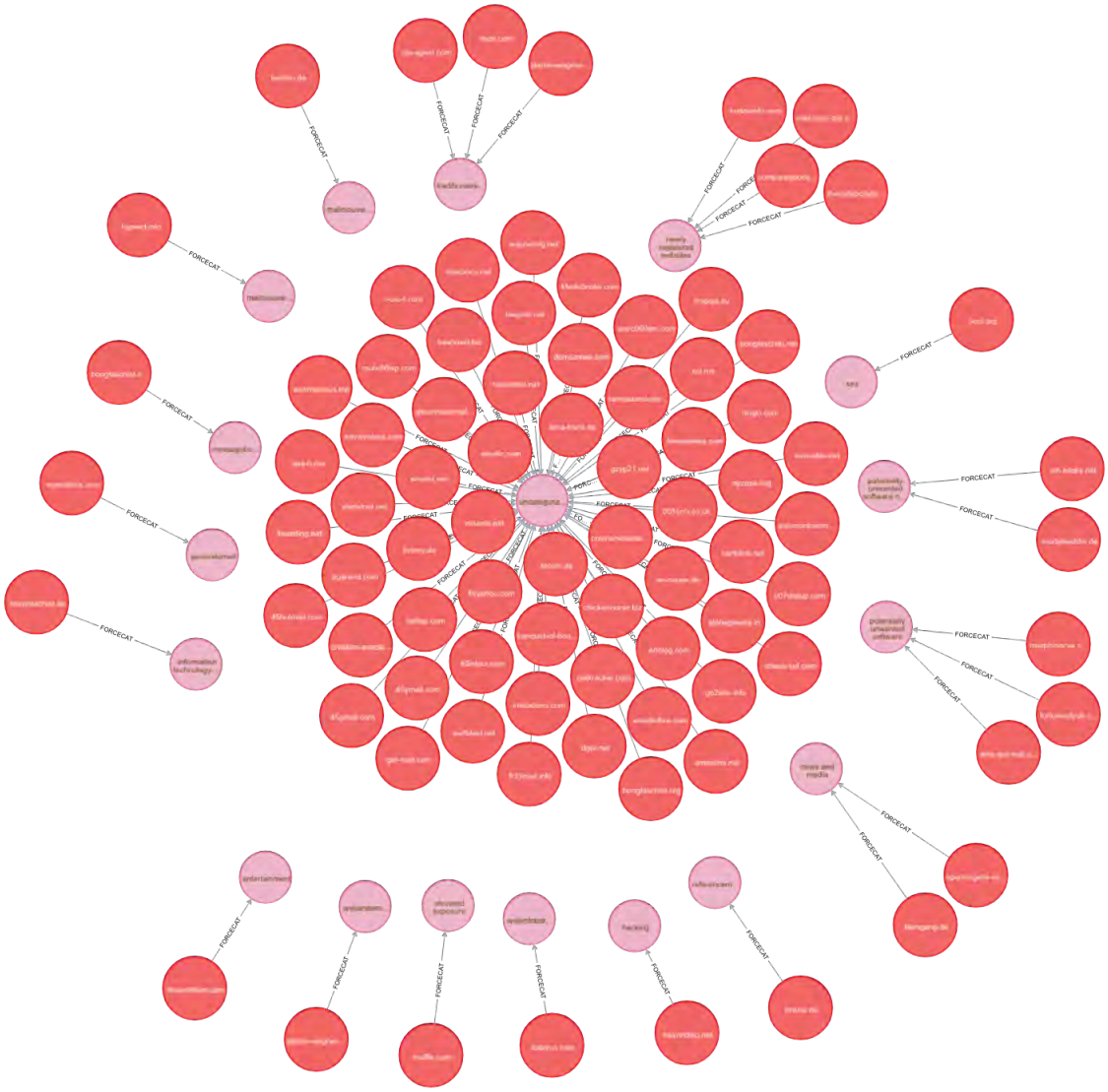

Figure 4.11: 206.189.153.135 domain categorisation analysis

This is in contrast to the domains hosted on 109.108.145.100 where the majority of domains are categorised as "shopping" by ForcePoint. This is notable as the cryptomining variant on 109.108.145.100 was noted as being `authedmine.com`, a consensual based cryptominer. The categorisations for the domains on 109.108.145.100 can be seen in figure 4.12. These domains mine cryptocurrency in the visitors browser only with explicit permission, therefore the large number of categorised domains in not unexpected as the sites are most likely established and not serving malicious content.

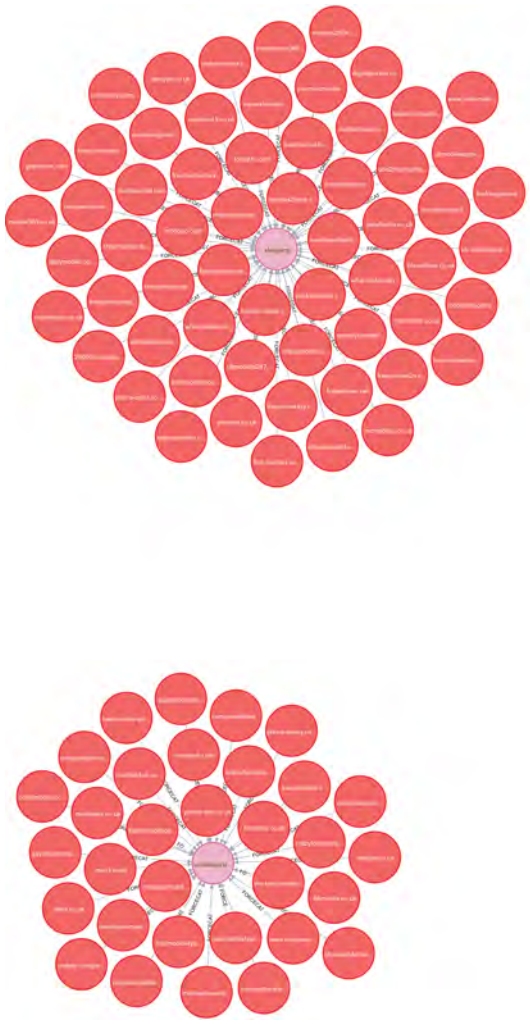

Figure 4.12: 109.108.145.100 domain categorisation analysis

This domain categorisation analysis shows a visual representation of the difference in domain categorisations of sites containing a non-consensual cryptomining variant as opposed to the sites using a mining script requiring user consent. The pattern in figure 4.11 illustrates the non-consensual mining variant including various sites with varying domain categorisations. This is in contrast to figure 4.12 where the consensual mining script is hosted on sites with a far smaller range of varying domain categorisations. This suggests an intentional process was followed to

establish legitimate reputationally sound shopping sites supplemented with consensual mining scripts as opposed to haphazard domain categorisations across an array of sites.

4.6 VirusTotal Graph Analysis

In order to analyse the top 10 IP addresses used for hosting mining scripts in the VirusTotal database, Maltego[37] was used to both incorporate the VirusTotal database and visualise the mined data effectively. The IP addresses where discussed in section 4.3. The previously acquired VirusTotal API key was integrated with Maltego in order to access the VirusTotal API. VirusTotal categorises detected samples under various conditions, two of these categories were noted amongst the top 10 IP addresses. They are the following:

- **Detected communicating samples** - these are detected files that have communicated with the specific domain or IP address.
- **Detected downloaded samples** - these are detected files that were downloaded from the domain or IP address, not from VirusTotal.

Figure 4.13 illustrates the associated detected communicating samples with each of the top 10 resolved hosts, examples are illustrated by the black arrows. Although some of the top 10 hosts utilise the same mining scripts as other hosts, none of the detected communicating sample hashes associated with any of the top 10 hosts are common to any of the other hosts. As per figure 4.10, the host with IP address 62.210.16.62 was noted as hosting four different mining scripts across 111 domains. During the VirusTotal analysis of the top 10 hosts in September 2019, the host with IP address 62.210.16.62 had 12 detected communicating samples associated with it, the same number of detected communicating samples as both the CDN hosts in the top 10 resolved hosts.

An IP address with the equivalent number of detected communicating samples as Google CDN endpoints is deemed to be high. As the functionality of Content Delivery Networks allows for multiple systems to be accessible via a single CNAME record, it is most likely that the 12 detected communicating samples access varying systems via the CDN's CNAME record as opposed to the 12 detected communicating samples that directly access the host with IP address 62.210.16.62.

[37] Maltego `https://paterva.com/`

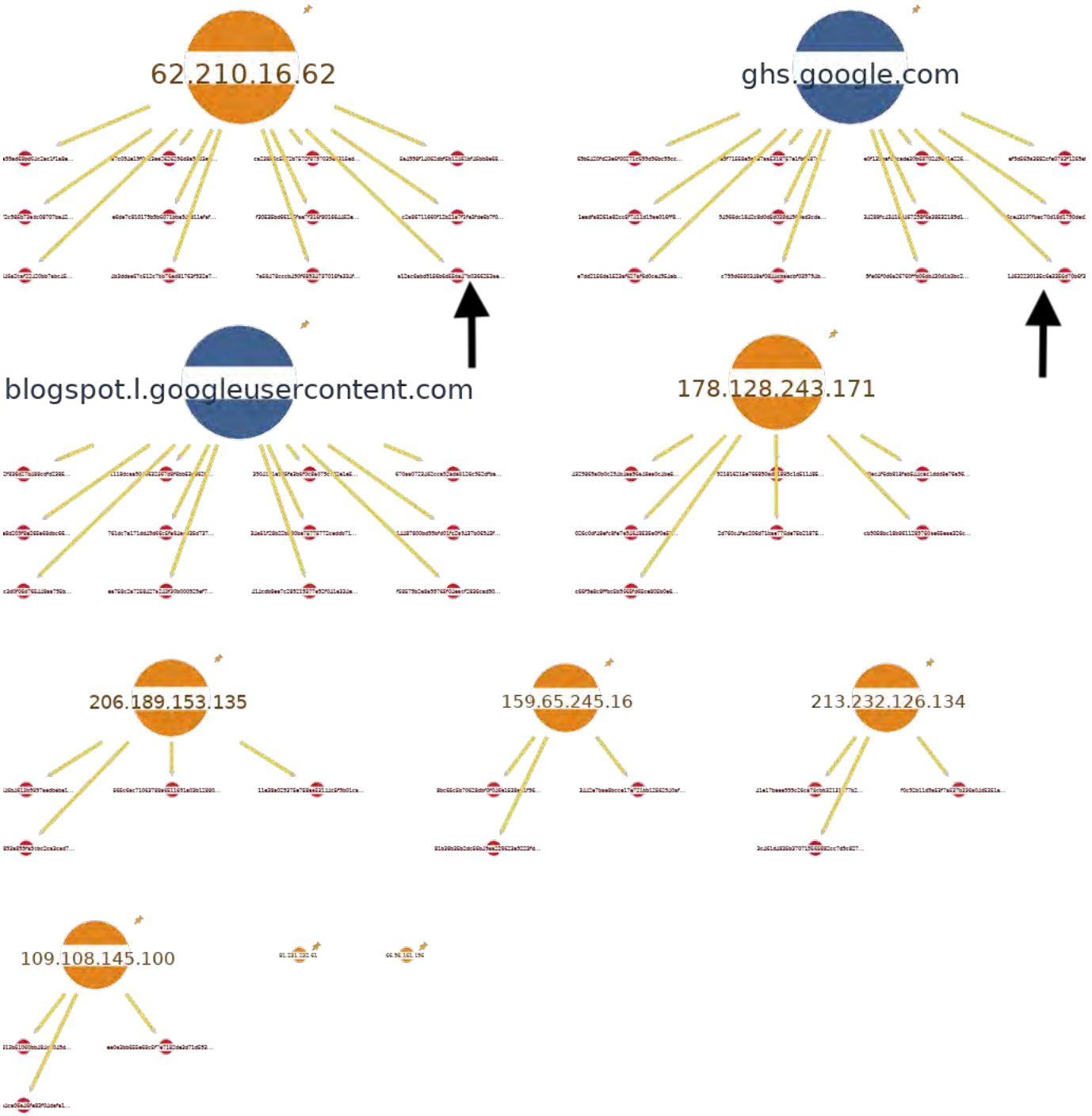

Figure 4.13: Detected VirusTotal communicating samples from the top 10 IP addresses

Figures 4.14 and 4.15 illustrate the associated detected downloaded samples with each of the top 10 resolved hosts. This analysis was conducted in September 2019 using Maltego and an associated VirusTotal API key. As was the case with the detected communicating samples, no overlap of the detected samples was noted across the hosts. The host with IP address 62.210.16.62 had 12 hashes of detected downloaded samples associated with it. This was more

than any other of the top 10 hosts. Neither of the CDN hosts had any hashes associated with it. This was expected as any associated detected downloaded samples would be associated with the root domain and not that of the CDN's CNAME record. The two other hosts from the top 10, each with two associated detected download samples are:

1. 66.96.161.196 - Both the `exdynsrv.com` and `traffic.tc-clicks.com` mining scripts were associated with this host.

2. 109.108.145.100 - The consent requiring `authedmine.com` mining script was associated with this host.

Table 4.9 details specific data points around each of the 3 discussed IP addresses.

Table 4.9: Assessed VirusTotal IP Addresses

IP Address	Reverse DNS	ASN	Whois Detail	AbuseDB Entries
62.210.16.62	pointer pf-lb-2.online.net.	AS12876	ONLINE SAS	1
66.96.161.196	196.161.96.66.static.eigbox.net.	AS29873	Endurance Group	0
109.108.145.100	109.108.145.100.srvlist.ukfast.net.	AS34934	Ukfast.net Limited	5

The host with IP address 62.210.16.62 is the outlier in terms of suspected malicious activity. Four different cryptomining scripts on the host is higher than any other individual IP address within the top 10 hosts. The same can be said for detected samples with 12 detected hashes for both the downloaded and the communicating samples. This is the highest number amongst the top 10 hosts for both the VirusTotal downloaded and communicating samples.

Figure 4.14: Detected VirusTotal downloaded samples from the top 10 IP addresses - 62.210.16.62

Each of the individual hashes illustrated in figure 4.14 represents code that VirusTotal has identified as being malicious. They are samples that were identified as being downloaded from the 62.210.10.62 IP address as opposed to being identified as samples that communicated with 62.210.10.62. The likelihood is that these hashes were not cryptomining scripts, but rather other kinds of malicious code hosted and potentially distributed by 62.210.10.62.

Figure 4.15: Detected VirusTotal downloaded samples from the top 10 IP addresses - Remaining top 9 hosts

4.7 JSECoin Analysis

The JSECoin[38] cryptocurrency was selected for dedicated analysis. This case study of a specific cryptocurrency was conducted due to the noted prevalence of the JSECoin cryptocurrency in the observed results of the data analysis.

JSECoin was noted in section 3.1 in table 3.1 as being the third most prevalent cryptomining script in use. As Coinhive is no longer operational and the `feesocrald.com` campaign no longer active, JSEcoin was selected for further detailed analysis as the most prevalent and currently active cryptomining script. The enriched dataset from figure 3.1 was queried using the Pandas Python library to create a new dataframe containing only sites embedded with the JSECoin mining script, this command can be found in listing 4.2.

[38] JSECoin - JavaScript embedded cryptocurrency `https://jsecoin.com/`

Listing 4.2: JSECoin dataframe creation

```
1 df_jsecoin=df[df["miner"] == "jsecoin.com"]
```

The newly created dataframe was exported as a csv file. All the domains were extracted from the file, which in turn was fed through a curl loop that downloaded the HTML source code for each page within the JSECoin dataframe. The HTML source code was downloaded during October 2019. The code for the download loop is included in listing 4.3.

Listing 4.3: JSECoin HTML download loop

```
1 for url in $(cat jsecoin_urls.txt);
2 do curl -OJ $url; done
```

Listing 4.4 details the initialisation configuration of the JSECoin mining script. The parameters are used as follows:

1. **AccountNo** - This is the publishers account number. It is the account that will be paid the reward for the cryptomining.
2. **PublisherSite** - This is the domain where the mining script is located.
3. **optionalSubID** - Is used for reference purposes. It will default to 'optionalSubID' if left unconfigured.

Listing 4.4: JSECoin mining script parameters

```
1 https://load.jsecoin.com/load/{:AccountNo}/
2 {:PublisherSite}/{:optionalSubID}/0/
```

The resulting HTML output from the curl loop in listing 4.3 was searched for all instances of "load.jsecoin.com". A sample of the output is included in listing 4.5. The numerical values following "load" are those of user accounts.

Listing 4.5: JSECoin initialisation code sample

```
1  https://load.jsecoin.com/load/108434/gold-indian.pro/gold/0/
2  https://load.jsecoin.com/load/108441/thejapanesesex.com/0/0/
3  https://load.jsecoin.com/load/109532/equipejob.com/optionalSubID/0/
4  https://load.jsecoin.com/load/110426/tourtiranabytaxi.com/0/0/
5  https://load.jsecoin.com/load/110849/itintroducer.fi/0/0/
6  https://load.jsecoin.com/load/110859/pornobox.cz/0/0/
7  https://load.jsecoin.com/load/112173/freevideobacks.com/0/0/
8  https://load.jsecoin.com/load/112173/livewallpaper.net/0/0/
9  https://load.jsecoin.com/load/112173/mywebprice.net/0/0/
10 https://load.jsecoin.com/load/112173/redboxjobs.com/0/0/
```

Analysis of the collected HTML source code revealed 376 unique account numbers. The top 5 user account numbers and their noted number of occurences in the HTML source code is displayed in table 4.10.

Table 4.10: Top 5 JSECoin user account prevalence

Rank	Account Number	Account Number Occurrences
1	15838	228
2	112173	8
3	274	6
4	12971	6
5	57168	6

Analysis of the user account ranked first with a count of 228 occurences revealed that all sites were hosted on the same server. That being the server with IP address 81.231.232.16, as illustrated in figure 4.7. Each site hosted on 81.231.232.16 was configured with an identical landing page offering the domain name for sale. An example of a site hosted on 81.231.232.16 is displayed in figure 4.16. The landing page includes the embedded JSECoin mining initialisation script, an advertising banner and an affiliate link to a Bitcoin exchange. It is apparent that the webmaster is attempting to diversify the monetisation strategy of each site on the server via a combination of three differing mechanisms.

Figure 4.16: 81.231.232.16 landing page example

4.7.1 JSECoin financial analysis

In order to determine to ascertain how much those implementing the `jsecoin.com` script were earning and whether the income correlated to the sites Alexa ranking, access to the ledger and account balances for JSECoin users was required. A JSECoin API key was obtained via registration on `https://platform.jsecoin.com/`. The JSECoin developer API provides access to user account balances via the endpoint in the listing 4.6.

Listing 4.6: JSECoin developer API user balance endpoint

```
1 https://api.jsecoin.com/v1.7/checkuserid/123/auth/ \
2 -H "Content-Type: application/json" \
3 -H "Authorization: <redacted>"
```

The list of unique JSECoin account ids obtained from the HTML code collection loop in listing 4.3 was combined with a curl loop to access the API endpoint in listing 4.6. The resulting code obtained the JSECoin balance for all collected unique JSECoin user ids. The curl loop used to obtain the balances during October 2019 can be found in listing 4.7.

Listing 4.7: JSECoin developer API user balance curl loop

```
1 for id in $(cat jsecoin_ids.txt);
2 do curl -v -X GET https://api.jsecoin.com/v1.7/checkuserid/$id/auth/
3 -H "Content-Type: application/json" \
4 -H "Authorization: <redacted>"
5 > jsecoin_balances ; done
```

A sample of the output file containing account balances can be seen in listing 4.8. The

JSECoin balance values equate to 0,09 and 32,79 US Dollars respectively.

Listing 4.8: JSECoin account balance output file


```
{
  "success": 1,
  "uid": 101497,
  "publicKey": "0459c3035ad3eb7698c2e8cb3a5292e97bb72648a3a64393e747
  ee7962a928e1404931103f6ba2aa7921fc7e939c9c2dad0b1a7e682b1cde921506
  c1c6e009e293",
  "balance": 555.17
}
{
  "success": 1,
  "uid": 10210,
  "publicKey": "04f7aa0e57f3e1427181af55c3c03af54dfb7cd1cd72a667d1d9f
  cecdd9135d1427302046b8d481c708c066c48e9e76b355cc89bf7a2c0dc1b0d1996
  324aed2763",
  "balance": 191763.94541743
}
```


The associated uid and balance fields were extracted for each response and combined into a single csv file for analysis. The resulting csv file was imported as a dataframe using the Python Pandas library.

As of the 1st of November 2019, 1 JSECoin was valued at 0,00028006 USD. Analysis of the 376 unique JSECoin user account balances revealed the following details:

- The sum of the combined user balances amounts to 22 647 154 JSECoin, 5795,40 US Dollars at the quoted exchange rate.
- The average user balance across the 376 unique users ids is 60 232 JSECoin. The average balance in US dollars equates to 15,40 US Dollars.
- The lowest balance was noted as -42 894 JSECoin. Three user accounts with negative balances exist within the dataset.
- The highest balance was noted as 93 813 15 JSECoin, that equates to 2475,82 US Dollars.

Table 4.11 details the top five JSECoin user accounts in terms of their counted occurrences across different websites, with the inclusion of their JSECoin and USD account balances. The

top five was chosen as opposed to the top 10 due to the low occurrence count of users outside the top 5.

Table 4.11: Top 5 JSECoin user account prevalence with balances

Rank	User account	Count	JSECoin balance	USD balance
1	15838	228	2292	0.55
2	112173	8	-271	-0.06
3	274	6	50485	12.22
4	12971	6	4493	1.89
5	57168	6	100300	26.55

Both tables 4.11 and table 4.12 indicate that user account prevalence does not appear to have an impact on increasing the user's JSECoin account balance. In contrast to table 4.11, table 4.12 shows all but one account of the top 10 accounts was noted as having more than one occurrence, account 7076. All of the remaining top nine accounts in terms of highest balances were counted only once within the sites listed as including the JSECoin mining script. The following however must be noted when analysing user account balances associated with JSECoin mining:

- The age of the user account cannot be determined and therefore there could be vast disparities of mining time between accounts.
- Other sources of JSECoin income may be included within the balances, such as purchasing of JSECoin.
- JSECoin may have been accumulated and subsequently sold or transferred out from an account in question.

Analysis of the sites associated with the top 10 user accounts in table 4.13 includes the PublicWWW rank for each of the eleven sites. The PublicWWW ranking of each site was included in the original during the initial dataset acquisition from PublicWWW. The PublicWWW ranking does not appear to have a notable influence of the underlying user accounts balance. Of the eleven sites associated with the top user account balances:

- Four of the sites do not have a ranking at all.
- Four of the sites have a ranking higher than 30 million
- Of the three sites with a ranking, the lower the sites ranking (the closer to number 1) the higher the user account balance.

Table 4.12: Top 10 JSECoin user account balances

Rank	User account	Count	JSECoin balance	USD balance
1	9250	1	9381315	2627.42
2	20490	1	1037387	290.80
3	97582	1	1028446	288.30
4	142434	1	704357	197.44
5	40622	1	680666	195.60
6	54635	1	498712	143.34
7	7076	2	406953	116.97
8	13778	1	358501	103.04
9	83980	1	332111	95.45
10	3431	1	328489	94.41

Table 4.13: Ranking of sites associated with top 10 user accounts by value

Rank	User account	Domain	PublicWWW Rank	USD balance
1	9250	thebrandmoneycantbuy.com	>30M	2627.42
2	20490	bagsracing.com	N\A	290.80
3	97582	agrande.pl	>30M	288.30
4	142434	sellingvps.site	>30M	197.44
5	40622	csubakka.hu	72725	195.60
6	54635	digitaldredger.com	>30M	143.34
7	7076	pokemongotoolkit.com	N\A	116.97
		langolonerd.it	N\A	
8	13778	webmailad.com	1784119	103.04
9	83980	tarot-cartas.com	2253724	95.45
10	3431	delphisources.ru	N\A	94.41

4.8 Discussion of Results

The data analysis revealed that although Coinhive is no longer operational, it (including Authedmine) continues to maintain the largest percentage of discovered cryptomining scripts amongst searched source code. It's total percentage has however declined since Rüth *et al.* (2018) mea-

sured the prevalence of cryptomining variants in 2018 with the Adblock nocoin list, while Coinhive was still operational. The fact that the most prevalent cryptomining script is that of the now defunct Coinhive is indicative of the decline in popularity of cryptojacking. The fact that no competitive services have emerged to fill the vacuum left by Coinhive is a clear indicator of the decline in demand for browser based cryptomining services.

In terms of geographic analysis, the vast majority of systems hosting cryptominig scripts were geolocated to the United States of America. There were however some outliers whereby some cryptomining variants had the majority of systems geolocated to Iran and Sweden, these were `feesocrald.com` and `jsecoin.com` respectively, as discussed in Section 4.2. The geographic analysis revealed that the United States of America remains very much the global centre of hosting cryptomining scripts. The geographic footprint of hosted cryptomining scripts is therefore relatively small in terms of coverage with no widespread international coverage except for the outliers previously mentioned.

Although most domains were unsurprisingly not categorised according to various domain categorisation services (due to the transient nature of hosting malicious content), many domains were identified as having domain categorisations, as per Section 4.3. The majority of these categorised domains were highlighted as "newly registered" or "parked", however, porn,sex and IT related categories accounted for the second largest segment of sites containing cryptomining scripts. The high volume of sites containing cryptoming scripts categorised as pornographic indicates a somewhat traditional approach for the hosting of malware\malicious content. The hosting of porn has the potential to acquire large amounts of browser traffic, coupled with hosted video or streaming content designed to keep sites users on the site for long periods of time, thus generating more income via the embedded cryptomining scripts.

While Cloudflare was noted as the Internet Service Provider with the most systems hosting cryptomining scripts, the IP addresses of the identified systems at Cloudflare were noted as being on different subnets. Section 4.4 explains that the Internet Service Providers with smaller numbers of identified systems were more likely to have these systems on related subnets. One such Internet Service Provider had 102 IP addresses found with cryptomining scripts, all of these were discovered on the same subnet. The prevalence of Cloudflare as the most widely used ISP for hosting cryptomining scripts is an indicator that the individuals hosting the cryptomining scripts are seeking the security and privacy services offered by Cloudflare. The use of Cloudflare offers a protective layer between the server containing the malicious code and the end user. This

hides the real IP address of the server hosting the cryptomining code.

Integration with VirusTotal in Section 4.7 showed that the top 10 servers containing the most cryptomining scripts were present within the VirusTotal dataset. These IP addresses had previously been recorded as having either hosted or been contacted by previously identified malware. In some cases up to 12 different malware hashes were associated with a server. These included both strains that had communicated with the server as well as samples downloaded from it. This indicates that the most prevalent servers in terms of hosting cryptomining scripts were previously involved in other nefarious cyber related activities. This correlation between hosted cryptomining scripts and other kinds of malicious code gives insight into the use of cryptojacking as an individual element in an ecosystem. It does not appear to be used in isolation by malicious participants but rather as part of a variety of malicious vectors.

As the largest active cryptomining variant, a detailed analysis on `jsecoin.com` was performed in Section 4.8. The analysis revealed no direct correlation between the number of sites a mining account holder runs and the their JSEcoin account balance. An individual user with 228 sites running cryptomining code had a noted balance equating to 0.55 USD. This is in contrast to the user with the highest JSEcoin balance of 2 627 USD, only having one site linked to their mining account. This aligns with high volume traffic being a key requirement for accumulating cryptocurreny through based cryptomining. Numerous low volume traffic sites will not be as profitable as a single site hosting cryptomining code with high volume traffic.

Chapter 5

Conclusion

5.1 Introduction

This chapter concludes the research by recapping the work done in the previous chapters. It further discusses the research objectives and whether or not they were achieved. Finally, the chapter concludes with the closing statement and discussion around future work in the field that should be researched.

5.2 Summary of Previous Chapters

- **Chapter 2** Provided the relevant background information regarding the basics of cryptocurrency mining as well as the introduction into how the cryptojacking infrastructure operates. It includes reviews of previous work around cryptojacking prevalence as well as technologies around cryptojacking detection and prevention.
- **Chapter 3** Discusses the acquisition of the initial dataset as well as the steps and requirements required to enrich it with the necessary data. Theses steps are documented and illustrated resulting in the enriched dataset required for the analysis.
- **Chapter 4** Details the analysis process whereby the enriched dataset is processed using various data analysis tools. The output of the analysis is discussed and interpreted accordingly.

5.3 Review of Research Objectives

The research objectives for this paper were declared in section 1.6. These objectives are further discussed in this section to determine whether each of the individual objectives were achieved

or not.

- How prevalent is illicit cryptomining throughout the Internet ? This question was sufficiently answered by the research. Section 4.1 details the prevalence of all encountered mining variants.
- Which cryptocurrencies are preferred for cryptojacking ? This was question was sufficiently answered. Sections 4.1 and 4.7 illustrate the top cryptocurrencies in use as well as the financial viability of the most widely used current cryptocurrency.
- Which Internet Service Providers and countries contain the most servers hosting cryptomining websites ? This question was sufficiently answered. Sections 4.1 and 4.3 provide the context into which ISP's and countries are most prevalent for hosting cryptomining scripts.

5.4 Closing Statement

The research conducted revealed that cryptojacking remains highly prevalent on the Internet today. Thousands of websites, hosted around the world are hosting websites with various cryptomining scripts hidden within their source code. These scripts illicitly mine cryptocurrencies for account holders by utilising the web user's computer resources. In most cases the web user is totally unaware of this process taking place. The research methodology should be continually repeated to continuously monitor the prevalence and trends of illicit cryptojacking on the Internet.

Despite the closure of the largest player in the browser based Monero mining infrastructure, cryptojacking remains widespread. The closure of Coinhive has not brought an end to crpytojacking but instead opened the market for new cryptographic currencies offering the means to mine in-browser. While block lists are continually updated to aide in the detection of site hosting mining scripts, their prevalence remains widespread. Operationally, both end users and organisations need to be cognisant of the cryptomining threat and look to utilise technology built to detect and counter cryptomining, as previously discussed in Section 2.8.4.

5.5 Future Work

There is much scope for future work and development in the area of cryptojacking analysis:

The methods in this research should be expanded upon to include the monitoring and discovery of Web Assembly cryptojacking being used on the Internet, previously discussed in Section 2.8.2.

Further analysis in the comparison of consensual cryptomining versus banner advertisements on websites should be conducted. This serves as a means to determine the validity of consensual cryptomining as a legitimate form of web based revenue, introduced in Section 2.4. Whether current detection and mitigation solutions are effective or not is a suggested area for further research as well as the categorisations of domains primarily used by websites hosting embedded cryptojacking code.

References

Allix, K., Bissyandé, T. F., Klein, J., and Traon, Y. L. Androzoo: Collecting millions of android apps for the research community. In *2016 IEEE/ACM 13th Working Conference on Mining Software Repositories (MSR)*, pages 468–471. May 2016. doi:10.1109/MSR.2016.056.

Anjum, A., Sporny, M., and Sill, A. Blockchain standards for compliance and trust. *IEEE Cloud Computing*, 4(4):84–90, July 2017. doi:10.1109/mcc.2017.3791019.

Bentov, I., Lee, C., Mizrahi, A., and Rosenfeld, M. Proof of activity. *ACM SIGMETRICS Performance Evaluation Review*, 42(3):34–37, December 2014. doi:10.1145/2695533.2695545. URL `https://doi.org/10.1145/2695533.2695545`

Berecz, G. and Czibula, I.-G. Hunting traits for cryptojackers. In *Proceedings of the 16th International Joint Conference on e-Business and Telecommunications*. SCITEPRESS - Science and Technology Publications, 2019. doi:10.5220/0007837403860393.

Bickford, J., O'Hare, R., Baliga, A., Ganapathy, V., and Iftode, L. Rootkits on smart phones. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications - HotMobile '10*. ACM Press, 2010. doi:10.1145/1734583.1734596. URL `https://doi.org/10.1145/1734583.1734596`

Bijmans, H. L. J., Booij, T. M., and Doerr, C. Inadvertently making cyber criminals rich: A comprehensive study of cryptojacking campaigns at internet scale. In *Proceedings of the 28th USENIX Conference on Security Symposium*, SEC'19, pages 1627–1644. USENIX Association, Berkeley, CA, USA, 2019. ISBN 978-1-939133-06-9. URL `http://dl.acm.org/citation.cfm?id=3361338.3361451`

BitcoinWiki. Important milestones of the Bitcoin project. 2009. Accessed February 2019. URL `https://en.bitcoin.it/wiki/Category:History`

Carlin, D., OrKane, P., Sezer, S., and Burgess, J. Detecting cryptomining using dynamic

analysis. In *2018 16th Annual Conference on Privacy, Security and Trust (PST)*. IEEE, August 2018. doi:10.1109/pst.2018.8514167.

Caviglione, L., Gaggero, M., Lalande, J.-F., Mazurczyk, W., and Urbanski, M. Seeing the unseen: Revealing mobile malware hidden communications via energy consumption and artificial intelligence. *IEEE Transactions on Information Forensics and Security*, 11(4):799–810, April 2016. doi:10.1109/tifs.2015.2510825.

Cimpanu, C. The internet is rife with in-browser miners and it's getting worse each day. 2017. Accessed February 2019.
URL `https://www.bleepingcomputer.com/news/security/the-internet-is-rife-with-in-browser-miners-and-its-getting-worse-each-day/`

Cova, M., Kruegel, C., and Vigna, G. Detection and analysis of drive-by-download attacks and malicious javascript code. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 281–290. ACM, New York, NY, USA, 2010. ISBN 978-1-60558-799-8. doi:10.1145/1772690.1772720.

Darabian, H., Homayounoot, S., Dehghantanha, A., Hashemi, S., Karimipour, H., Parizi, R. M., and Choo, K.-K. R. Detecting cryptomining malware: a deep learning approach for static and dynamic analysis. *Journal of Grid Computing*, January 2020. doi:10.1007/s10723-020-09510-6.

Desnitsky, V. and Kotenko, I. Analysis of energy resource depletion attacks on wireless devices. *Izvestiâ vysših učebnyh zavedenij. Priborostroenie*, 61(4):291–297, April 2018. doi:10.17586/0021-3454-2018-61-4-291-297.

Dev, J. A. Bitcoin mining acceleration and performance quantification. In *2014 IEEE 27th Canadian Conference on Electrical and Computer Engineering (CCECE)*, pages 1–6. IEEE, 2014.

Draghicescu, D., Caranica, A., Vulpe, A., and Fratu, O. Crypto-mining application fingerprinting method. In *2018 International Conference on Communications (COMM)*, pages 543–546. IEEE, 2018.

Dunn, J. E. Unsecured AWS led to cryptojacking attack on la times. 2018. Accessed February 2019.
URL `https://nakedsecurity.sophos.com/2018/02/27/unsecured-aws-led-to-cryptojacking-attack-on-la-times/`

Edelman, B., Ostrovsky, M., and Schwarz, M. Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. Working Paper 11765, National Bureau of Economic Research, November 2005. doi:10.3386/w11765.

Eskandari, S., Andrea, Leoutsarakos, Mursch, T., and Clark, J. A first look at browser-based cryptojacking. *Concordia University, Bad Packets Report*, 2018.
URL `https://arxiv.org/pdf/1803.02887.pdf`

European Commission. Information provider's guide:cookies. 2011. Accessed February 2019.
URL `http://ec.europa.eu/ipg/basics/legal/cookies/index_en.htm`

Hackett, R. Popular Google Chrome extension caught mining cryptocurrency on thousands of computers. 2018. Accessed February 2019.
URL `http://fortune.com/2018/01/02/google-chrome-extension-cryptocurrency-mining-monero/`

Hajri, H. H. A., Mughairi, B. M. A., Hossain, M. I., and Karim, A. M. Crypto jacking a technique to leverage technology to mine crypto currency. *International Journal of Academic Research in Business and Social Sciences*, 9(3), March 2019. doi:10.6007/ijarbss/v9-i3/5791.
URL `https://doi.org/10.6007/ijarbss/v9-i3/5791`

Hoffman, J. J. New Jersey division of consumer affairs obtains settlement with developer of bitcoin-mining software found to have accessed New Jersey computers without users' knowledge or consent. 2015. Accessed February 2019.
URL `https://nj.gov/oag/newsreleases15/pr20150526b.html`

Hohlfeld, O. Operating a DNS-based active internet observatory. In *Proceedings of the ACM SIGCOMM 2018*. ACM Press, 2018. doi:10.1145/3234200.3234239.

Hollister, S. Hotel caught injecting advertising into webpages on 'complimentary' wi-fi network. 2012. Accessed February 2019.
URL `https://www.theverge.com/2012/4/7/2931600/hotel-caught-injecting-advertising-into-web-pages-on-complimentary-wi`

Hong, G., Yang, Z., Yang, S., Zhang, L., Nan, Y., Zhang, Z., Yang, M., Zhang, Y., Qian, Z., and Duan, H. How you get shot in the back: A systematical study about cryptojacking in the real world. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 1701–1713. ACM, New York, NY, USA, 2018. ISBN 978-1-4503-5693-0. doi:10.1145/3243734.3243840.

Hruska, J. Browser-Based Mining Malware Found on Pirate Bay, Other Sites. 2017. Accessed February 2019.
URL https://www.extremetech.com/internet/255971-browser-based-cryptocurrency-malware-appears-online-pirate-bay

Huang, D. Y., McCoy, D., Dharmdasani, H., Meiklejohn, S., Dave, V., Grier, C., Savage, S., Weaver, N., Snoeren, A. C., and Levchenko, K. Botcoin: Monetizing stolen cycles. In *Proceedings 2014 Network and Distributed System Security Symposium*. Internet Society, 2014. doi:10.14722/ndss.2014.23044.

Kauthamy, K., Ashrafi, N., and Kuilboer, J.-P. Mobile devices and cyber security - an exploratory study on user's response to cyber security challenges. In *Proceedings of the 13th International Conference on Web Information Systems and Technologies*. SCITEPRESS - Science and Technology Publications, 2017. doi:10.5220/0006298803060311.
URL https://doi.org/10.5220/0006298803060311

Kharraz, A., Ma, Z., Murley, P., Lever, C., Mason, J., Miller, A., Borisov, N., Antonakakis, M., and Bailey, M. Outguard: Detecting in-browser covert cryptocurrency mining in the wild. In *The World Wide Web Conference on - WWW '19*. ACM Press, 2019. doi:10.1145/3308558.3313665.

Kolondra, K. New year, new browser. Opera 50 introduces anti-bitcoin mining tool. 2018. Accessed February 2019.
URL https://blogs.opera.com/desktop/2018/01/opera-50-introduces-anti-bitcoin-mining-tool/

Konoth, R. K., Vineti, E., Moonsamy, V., Lindorfer, M., Kruegel, C., Bos, H., and Vigna, G. Minesweeper: An in-depth look into drive-by cryptocurrency mining and its defense. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, CCS '18, pages 1714–1730. ACM, New York, NY, USA, 2018. ISBN 978-1-4503-5693-0. doi:10.1145/3243734.3243858.

Krishnan, H., Saketh, S., and Tej, V. Cryptocurrency mining – transition to cloud. *International Journal of Advanced Computer Science and Applications*, 6(9), 2015. doi:10.14569/ijacsa.2015.060915.

Kroll, J. A., Davey, I. C., and Felten, E. W. The economics of bitcoin mining, or bitcoin in the presence of adversaries. In *Proceedings of The Workshop on the Economics of Information Security*, volume 2013, page 11. 2013.

Kshetri, N. and Voas, J. Do crypto-currencies fuel ransomware? *IT Professional*, 19(5):11–15, 2017. doi:10.1109/mitp.2017.3680961.

Liao, S. Showtime websites secretly mined user CPU for cryptocurrency. 2017. Accessed February 2019.
URL https://www.theverge.com/2017/9/26/16367620/showtime-cpu-cryptocurrency-monero-coinhive

Liu, C. and Chen, J. C. Malvertising campaign abuses google's doubleclick to deliver cryptocurrency miners. 2018. Accessed February 2019.
URL https://blog.trendmicro.com/trendlabs-security-intelligence/malvertising-campaign-abuses-googles-doubleclick-to-deliver-cryptocurrency-miners/

McCarthy, K. CBS's showtime caught mining crypto-coins in viewers' web browsers. 2017. Accessed February 2019.
URL https://www.theregister.co.uk/2017/09/25/showtime_hit_with_coinmining_script/

Meland, P. H., Johansen, B. H., and Sindre, G. An experimental analysis of cryptojacking attacks. In *Secure IT Systems*, pages 155–170. Springer International Publishing, 2019. doi:10.1007/978-3-030-35055-0_10.

Mursch, T. Cryptojacking malware coinhive found on 30,000+ websites. 2017. Accessed February 2019.
URL https://badpackets.net/cryptojacking-malware-coinhive-found-on-30000-websites/

Musch, M., Wressnegger, C., and And Konrad Rieck, M. J. Web-based cryptojacking in the wild. *Computing Research Repository*, abs/1808.09474, 2018.

Nakamoto, S. Bitcoin: A peer-to-peer electronic cash system. 2008. Accessed February 2019.
URL https://bitcoin.org/bitcoin.pdf

Narayanan, A., Bonneau, J., Felten, E. W., Miller, A., and Goldfeder, S. Bitcoin and Cryptocurrency Technologies - A Comprehensive Introduction. Princeton University Press, 2016. ISBN 978-0-691-17169-2.

Nikiforakis, N., Maggi, F., Stringhini, G., Rafique, M. Z., Joosen, W., Kruegel, C.,

Piessens, F., Vigna, G., and Zanero, S. Stranger danger: Exploring the ecosystem of ad-based url shortening services. pages 51–62. 04 2014. doi:10.1145/2566486.2567983.

O'Gorman, B. Cryptojacking: A modern cash cow. Technical report, Trend Micro, September 2018. Accessed November 2018.
URL `https://www.symantec.com/content/dam/symantec/docs/security-center/white-papers/istr-cryptojacking-modern-cash-cow-en.pdf`

Pastrana, S. and Suarez-Tangil, G. A first look at the crypto-mining malware ecosystem: A decade of unrestricted wealth. *arXiv preprint arXiv:1901.00846*, 2019.

Raikos, B. G. PayTech and blockchain: Adjusting for security and risk. November 2019. doi:10.1002/9781119551973.ch42.

Rauchberger, J., Schrittwieser, S., Dam, T., Luh, R., Buhov, D., Pötzelsberger, G., and Kim, H. The other side of the coin. In *Proceedings of the 13th International Conference on Availability, Reliability and Security - ARES 2018*. ACM Press, 2018. doi:10.1145/3230833.3230869.

Razali, M. A. and Shariff, S. M. CMBlock: In-browser detection and prevention cryptojacking tool using blacklist and behavior-based detection method. In *Advances in Visual Informatics*, pages 404–414. Springer International Publishing, 2019. doi:10.1007/978-3-030-34032-2_36.

Reddy, N. Bitcoin forensics. In *Practical Cyber Forensics*, pages 401–432. Apress, 2019. doi:10.1007/978-1-4842-4460-9_13.
URL `https://doi.org/10.1007/978-1-4842-4460-9_13`

Romano, D. and Schmid, G. Beyond bitcoin: A critical look at blockchain-based systems. *Cryptography*, 1(2):15, September 2017. doi:10.3390/cryptography1020015.

Rosenfeld, M. Analysis of bitcoin pooled mining reward systems. *Computing Research Repository*, abs/1112.4980, 2011.
URL `https://arxiv.org/pdf/1112.4980.pdf`

Rüth, J., Zimmermann, T., Wolsing, K., and Hohlfeld, O. Digging into browser-based crypto mining. In *Proceedings of the Internet Measurement Conference 2018 on - IMC '18*. ACM Press, 2018. doi:10.1145/3278532.3278539.

Rydstedt, G., Bursztein, E., Boneh, D., and Jackson, C. Busting frame busting: a study of clickjacking vulnerabilities at popular sites. In *In IEEE Oakland Web 2.0 Security and Privacy Workshop*, page 6. 2010.

Saad, M., Khormali, A., and Mohaisen, A. End-to-end analysis of in-browser cryptojacking. *arXiv:1809.02152*, 2018.

Sato, M., Imamura, Y., Orito, R., and Yamauchi, T. (short paper) method for preventing suspicious web access in android WebView. In *Advances in Information and Computer Security*, pages 241–250. Springer International Publishing, 2019. doi:10.1007/978-3-030-26834-3_14.
URL `https://doi.org/10.1007/978-3-030-26834-3_14`

Sompolinsky, Y. and Zohar, A. Secure high-rate transaction processing in bitcoin. In *International Conference on Financial Cryptography and Data Security*, pages 507–527. Springer, 2015.

Tahir, R., Huzaifa, M., Das, A., Ahmad, M., Gunter, C., Zaffar, F., Caesar, M., and Borisov, N. Mining on someone else's dime: Mitigating covert mining operations in clouds and enterprises. In *Research in Attacks, Intrusions, and Defenses*, pages 287–310. Springer International Publishing, 2017. doi:10.1007/978-3-319-66332-6_13.

Varlioglu, S., Gonen, B., Ozer, M., and Bastug, M. F. Is cryptojacking dead after coinhive shutdown? *arXiv:2001.02975*, 2020.

Zhao, T., Zhang, G., and Zhang, L. An overview of mobile devices security issues and countermeasures. In *2014 International Conference on Wireless Communication and Sensor Network*. IEEE, December 2014. doi:10.1109/wcsn.2014.95.
URL `https://doi.org/10.1109/wcsn.2014.95`