Inadvertently Making Cyber Criminals Rich: A Comprehensive Study of Cryptojacking Campaigns at Internet Scale

Hugo L.J. Bijmans
Delft University of Technology

Tim M. Booij
Delft University of Technology

Christian Doerr
Delft University of Technology

Abstract

Since the release of a browser-based cryptominer by Coinhive in 2017, the easy use of these miners has skyrocketed illicit cryptomining in 2017 and continued in 2018. This method of monetizing websites attracted website owners, as well as criminals seeking new ways to earn a profit. In this paper, we perform two large studies into the world of cryptojacking, focused on organized cryptomining and the spread of cryptojacking on the Internet. We have identified 204 cryptojacking campaigns, an order of magnitude more than previous work, which indicates that these campaigns are heavily underestimated by previous studies. We discovered that criminals have chosen third-party software – such as WordPress – as their new method for spreading cryptojacking infections efficiently. With a novel method of using NetFlow data we estimated the popularity of mining applications, which showed that while Coinhive has a larger installation base, CoinImp WebSocket proxies were digesting significantly more traffic in the second half of 2018. After crawling a random sample of 49M domains, ~20% of the Internet, we conclude that cryptojacking is present on 0.011% of all domains and that adult content is the most prevalent category of websites affected.

1 Introduction

Unlike traditional currencies, such as the Euro or Dollar, cryptocurrencies are digital assets created as a medium of exchange based on cryptography and a blockchain, which are used to secure both the creation and transactions of units. In 2009, Satoshi Nakamoto released the Bitcoin [33], the first ever decentralized cryptocurrency, which made it possible to transfer monetary value to another person by creating a transaction and committing this to the blockchain, a list of blocks secured by cryptographic operations maintained by a peer-to-peer network of miners. These miners secure the blockchain by constantly collecting transaction data from the network and validating it by solving cryptographic challenges based on the previous block, the transaction and the receiver of the transaction. After validation, the confirmed transaction is inserted into the blockchain again in the form of a validated block. As a reward, the miner gets a (part of a) cryptocurrency. This network guarantees that only the rightful owner of a Bitcoin wallet can make transactions and prevents malicious actors from inserting false information into the blockchain.

Solving these cryptographic challenges as a miner has however become so difficult that Bitcoin cannot efficiently be mined anymore on regular PCs. Over the past years, over 4,000 other cryptocurrencies have been created, so-called alt-coins. One of them is Monero (XMR), launched in 2014 and nowadays the most popular cryptocurrency in browser-based mining [34]. In contrast to Bitcoin, Monero uses a private blockchain, meaning that while anybody can use it to make transactions, nobody is allowed to view them [47]. It also builds upon a different proof-of-work algorithm to validate its transactions, called CryptoNight, a fork of CryptoNote [43]. This algorithm is designed to be memory-hard and therefore requires a large set of bytes in memory to perform frequent read and write operations on. Simple consumer-grade CPUs have exactly that memory available at their processor caches, making this kind of mining the most efficient on regular consumer-grade hardware. To speed up the mining process, mining jobs can be distributed among individual miners in a mining pool. In such a pool, miners work together to mine new blocks and share the rewards. Work is distributed among miners in the pool based on the difficulty of the cryptographic challenge. As a consequence, powerful machines solve the more difficult puzzles, while low-end machines receive the easier ones. Rewards are shared according to the same principle. Mining pools closely monitor the submissions from their miners and state that they will block any wallet address after receiving evidence that a wallet is used for malware or botnet activities [28].

The introduction of alt-coins that by design can be effectively mined on regular PCs also made them an attractive target for cybercriminals. Both the private blockchain and the ASIC-resistant mining algorithm of Monero quickly made Monero one of the preferred choices. In addition to being

included in malware [37], there also exist implementations to perform *drive-by mining* or *cryptojacking*, where cryptocurrency is mined in the user's web browser while visiting a web site. While originally developed as an alternative mechanism to donate to the upkeep of a website in presence of now ubiquitous ad-blockers, many methods exist to maliciously apply browser-based mining: for example, criminals hack vulnerable websites to install mining scripts [3] or create malicious advertisements with cryptojacking code that are displayed on benign websites [30], but actors have also compromised routers [35] or setup malicious Wi-Fi networks [38] to inject cryptominers into their users' traffic.

Previous studies have performed surveys on the use of cryptominers across the most commonly visited websites and have identified groups of criminals installing cryptominers on a large number of domains for their own profit [22, 39]. It makes sense for a cyber criminal to lure as many users as possible into such mining, which could be accomplished not only by deploying the cryptojacking code into popular websites, but hacking a large number of websites or injecting a resource such as a common library that is used by a large number of unsuspecting websites. These individual installations are working together in a coordinated campaign, thus significantly increasing the profits of the criminal, but at the same time also indicating an elevated level of knowledge and sophistication of the adversary. The presence and extent of such coordination is however largely unknown.

In this paper, we address this gap and systematically investigate the coordination and collaboration of cryptojackers on the Internet and make the following four contributions:

- We are the first to systematically analyze the relationships between websites that perform cryptomining and the actors behind them. By this campaign analysis, we find the existence of massive installations. In fact, we have identified 3 times as much cryptojacking activity as [39] and the five largest campaigns we detected exceed the *total* size of cryptomining reported in [22].
- We show that the bulk of organized mining activity is the result of compromised (parts of) third-party software and that comparatively little organized activity is the result of hacked websites or an explicit choice to mine by the website owner.
- Through a survey of 1,136 top level domains and by comparing the installation base with actual mining traffic on the Internet using NetFlow data, we find that the most prominently installed miner is actually not the one that generates the most mining activity in practice. We also see that applications and attack vectors come and go, and that different TLD zones exhibit clear differences in mining application popularity.
- Estimating cryptojacking by solely crawling the Alexa Top 1M results is an overestimation of the size, as we see that cryptojacking activity is almost 6 times higher in that subset compared to the rest of the Internet.

To enable follow up research, we make our data and software publicly available at `https://www.cyber-threat-intelligence.com/cryptojacking-campaigns`.

2 Background

WebAssembly & asm.js To enable faster execution of code inside the browser, Mozilla developed *asm.js*, a technique for translating high-level languages, such as C and C++, into JavaScript to be used by the browser [29]. Multiple validation methods enable the JavaScript engine to compile this code ahead-of-time and improve execution speed. This technique made it possible to execute code faster inside the browser after its release in 2013.

WebAssembly (*Wasm*) is a more recently released scripting language developed by the World Wide Web consortium in 2017 and is able to compile high-level languages like C, C++ and Rust inside the browser to be used in web applications [50]. It runs in a sandbox within the browser and it aims to execute as fast as native machine code. Wasm is complementary to JavaScript, as it is being controlled by JavaScript code after its compilation.

The difference between asm.js and Wasm is the fact that the latter is compiled only once and is started directly at native speed, whereas code in asm.js is compiled and optimized at run time, therefore decreasing execution speed. Both techniques are supported by all four major browsers (Chrome, Firefox, Edge and Safari) and have drastically improved the execution speed of applications inside the browser, which made them very attractive for browser-based mining.

WebSockets & Stratum WebSockets is a HTML5 protocol providing two-way communication between the client and a server over a single TCP connection [52]. The protocol enables easy real-time data transfer without refreshing (a part of) the web page. Communication is done over the same TCP ports as the web browser, making it robust to strict firewall rules or other blocking.

Developers are free to define the format of messages sent over WebSocket connections. However, there is a protocol specifically designed for cryptomining communications: the Stratum Mining Protocol, a line-based protocol with messages encoded in plain-text JSON-RPC format [46]. Servers communicate with their clients using Stratum to authorize new miners in the pool, distribute jobs based on difficulty and retrieve found hashes from the miners. An example of a WebSocket connection using the Stratum protocol is given in Table 1.

Browser-based mining Triggered by the rise of CPU-mineable cryptocurrencies (such as Monero) and the rapid

WebSocket traffic frames
⇑ {"type":"**auth**", "params":{"site_key":"<site_key_of_website>", "type":"anonymous","user":null,"goal":0, "version":3000,"coin":"xmr"}}
⇓ {"type":"**authed**", "params":{"token":"<random_36_characters>", "hashes":0}}
⇓ {"type":"**job**", "params":{"blob":"<random_152_characters>", "job_id":"<random_28_characters>", "target":"ffffff01", "id":"<random_36_characters>", "algo":"cn","variant":"4","height":1808537}}
⇑ {"type":"**submit**", "params":{"job_id":"<random_28_characters>", "nonce":"377c32b8", "result":"<found_64_characters_hash>"}}
⇓ {"type":"**hash_accepted**", "params":{"hashes":128}}
⇓ {"type":"**job**", "params":{"blob":"<random_152_characters>", "job_id":"<random_28_characters>", "target":"ffffff01", "id":"<random_36_characters>", "algo":"cn","variant":"4","height":1808537}}

Table 1: Example of a WebSocket connection using the Stratum Mining Protocol to communicate with a mining pool

development of useful web standards (e.g.WebAssembly and the Stratum protocol), browser-based cryptomining gained an enormous momentum in the autumn of 2017. Coinhive, a German company, created an easy to use browser-based mining application as an alternative to advertisements [9, 23]. They provide a JavaScript library, an API and a WebSocket proxy infrastructure to developers to easily integrate a browser-based miner into their website and let their visitors mine for Monero. 70% of the mined Monero is transferred to the owner of the account, the remaining 30% is kept by Coinhive [10]. Soon after Coinhive released their miner application, similar ones appeared, such as Cryptoloot [11] and Coin-Have [6]. Nowadays, miner applications come and go, with various capabilities and usage fees, but Coinhive still has a prominent place in the cryptojacking landscape.

Overview of a cryptojacking attack Although different mining applications exist, all browser-based miners show great similarities. As depicted in Figure 1, the user visits the cryptomining website (1) and receives a valid HTTP response (2). The cryptomining website requests a JavaScript file (3), which controls the mining operation. This script first explores the host system, searches for the number of CPU threads available, downloads the WebAssembly mining script

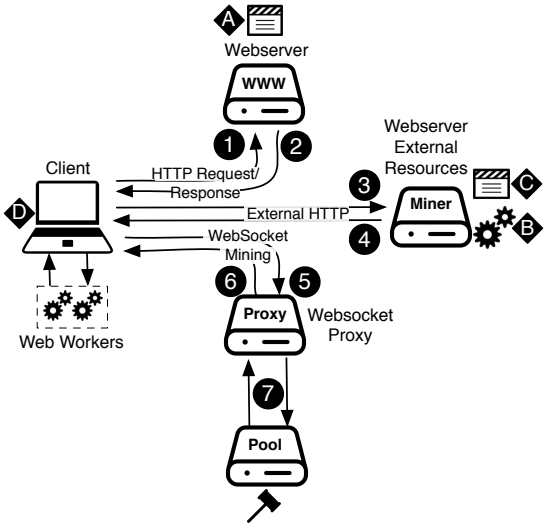

Figure 1: Browser-based cryptomining attack

for the actual mining operation (4) and distributes it over a number of WebWorkers (a JavaScript instance running in the background, without affecting the page performance). It also sets up a WebSocket connection with the mining pool through a proxy (5). The script authenticates itself to the mining pool server (in Stratum format) and, if successful, receives the first job to work on (6). The WebWorkers start working on that job and found hashes are submitted to the mining pool by the controller script (7).

Campaign analysis Campaign analysis is the field of research focused on discovering clusters of malicious online entities. The term originates from analysis of large volumes of SPAM or phishing emails, but can also be used in other areas, such as browser-based cryptomining. In this particular case, campaign analysis is focused on finding clusters of the same cryptominers on different domains. Since those miners always include a form of identification to which funds need to be transferred, clustering cryptojacking websites can be done relatively easily. Most mining applications define a *siteKey*, a unique (random) string used to identify the user to which earnings have to be transferred, which can be found in either the source code or the WebSocket traffic. A similar *siteKey* guarantees that the same account is rewarded for the mining that takes place. Identifying campaigns can also be done by searching for similar WebSocket proxy servers, if the website is not using a popular one, but instead hosting its own server. We have used these, and other techniques to discover campaigns as discussed in Section 6.1. We have chosen to define a cluster of websites as a campaign once they share identical features more than 5 times. E.g. a cluster of 6 websites with the same *siteKey* or private WebSocket proxy server is considered a campaign.

3 Attack vectors

Mining cryptocurrencies with the computing power of website visitors is not illegal, as long as users are asked permission to mine. When a user cannot consent to the mining activities their computer is involved in, it is called *cryptojacking*. Although browser-based cryptomining is a recent phenomenon, jurisdiction on cryptomining without consent already exists. In 2015, a US court settled a case with a developer of Bitcoin-mining software, in which the Attorney General stated that no website should tap into a person's computer processing power and that the user has to be informed about the cryptomining activities which take place on the visited website [18]. However, this is often not the case. In this section, we summarize the attack surface for cryptojacking on the Internet. All attack vectors are marked in Figure 1 by their corresponding characters.

Website owner (A) The owner of a website can add a cryptomining script to his web page without informing its users. This can be done as a replacement for advertisements, which was the case for The Pirate Bay, one of the most popular torrent websites [48]. Only a few days after the Coinhive service was launched, they added a miner to their website which started mining without user consent, as a replacement for the intrusive advertisements they would normally show. Nowadays, the website shows a disclaimer on the bottom of the homepage, notifying their visitors that their CPU will be used for cryptomining. Another major source of website owner initiated cryptojacking is parked domains [13].

Compromised websites (A) A cryptomining script can also be present on a web page without knowledge of the website owner. When a website gets hacked, an attacker is able to inject cryptomining scripts. Now, the attacker receives the rewards for the visitors mining on that website. There are numerous examples of this kind of attack. There have been cryptojacking scripts found on web pages of the Indian government [3], CBS Showtime [26] and many others.

Third-party software (B) Gaining unsolicited access to large number of domains is a time-consuming operation. As a consequence, attackers have tried different tactics to infect multiple websites at once by infecting third-party software. In the last year, we have seen attacks in which cryptojacking code is injected into popular third-party software, such as JQuery or Google Tag Manager [5]. Drupal, a widely used open-source CMS, was the victim of a large attack involving more than 100,000 websites [32] and WordPress, a similar CMS, suffered from a weather plugin [53] secretly injecting a cryptojacking script into the website it was installed on.

Malicious advertisements (C) Advertisement-supported websites let their advertisements be sold by advertisement networks, such as Google. The downside of this system is that attackers can attach cryptomining scripts to advertisements and distribute them through an advertisement network over a large number of websites. In January 2018, Youtube was a victim of this kind of attack, in which cryptomining scripts were injected in the ads shown on the website [30].

Man in the middle (D) The most effective method of gaining large groups of miners for an attacker is by being the man-in-the-middle. In August 2018, 200,000 MikroTik routers were infected by malware, which inserted a Coinhive script into every website the user visits [35]. The bug was patched within a day, but many MikroTik routers are not, leaving them still vulnerable. In our research, we are not able to detect these attacks, since they are not originating from a website.

4 Related Work

Academic research on browser-based cryptomining has only started in 2017 and is, due to the recent developments of the used web standards, very topically. The first explorations into this research field have been performed by Eskandari et al. [13]. In their analysis, the authors queried two large source code datasets for strings known to be part of cryptomining scripts (such as `coinhive.min.js` or `load.jsecoin.com`) and found a large number of domains. This method is only able to detect known mining applications, not the obfuscated or new ones. While calculating the profitability, the authors stumbled upon a Coinhive campaign which ran a miner on over 11,000 parked websites. This study kicked-off a number of subsequent investigations, which were all aimed at detecting browser-based cryptomining. Rauchberger et al. [39] created their *MiningHunter*, a crawler able to detect mining scripts even when their malicious activities are obfuscated. The detection method relied on analyzing executed JavaScript code and WebSocket traffic frames. After a successful crawl of the Alexa Top 1M in the beginning of December 2017, they were able to detect 3,178 websites running a cryptominer. 1,210 unique keys were retrieved and one large campaign involving 1,116 websites infected by a malicious advertisement network was identified. At the same time Parra Rodriguez et al. [40] worked on *RAPID*, a resource and API-based detection method, which is able to detect browser-based cryptomining and is resistant to JavaScript obfuscation. Their classification was able to classify mining samples with a precision of 96%. Eventually 656 actively mining websites were found in the Alexa Top 330,550. A similar classification study was performed by Carlin et al. [2], in which they demonstrated that dynamic opcode tracing is extremely effective at detecting cryptomining behavior. Liu et al. [24] proposed a novel approach for detecting browser-based mining applications by

creating *BMDetector*, a detection system based on a modified Chrome kernel. Using this modified kernel, the authors were able to perform JavaScript code block analysis on the compiled JavaScript code, which allowed them to detect heavily obfuscated miner applications as well. Hong et al. [19] built *CMTracker*, a behavior-based detector with two runtime profilers for tracking browser-based cryptomining. The first profiler monitors incoming JavaScript files for known fingerprints, the second profiler observes the call stack and searches for periodic executions. Their approach was able to detect 868 actively mining websites among the Alexa Top 100K in April 2018. More than half of the found keys were used only once and they noticed that domains hosting mining scripts were migrating faster than the mining pool domains. The authors also mentioned evasion techniques, such as code obfuscation and payload hiding inside third-party libraries. Periodic execution in mining scripts was also noticed by Wang et al. [49], who created *SEISMIC*, a monitoring service to interrupt browser-based mining scripts based on this finding.

A different view on the subject was given by Papadopoulos et al. [36], who tried to answer the question whether browser-based cryptomining could be a suitable alternative to advertisements. After crawling a dataset of 200K websites running advertisements or cryptominers, they concluded that advertisements are still more than 5 times more profitable than cryptominers. This will only change once a visitor stays on the same website for more than 5.3 minutes or when Monero becomes more valuable [36]. A broader view of the browser-based cryptomining ecosystem is given by Saad et al. [42], who researched both cryptomining code and user impact. Besides various JavaScript static code analysis clustering methods and battery drainage studies when cryptomining, they did not perform any crawling of the web. This is in great contrast to the work of Rüth et al. [41], who digged deep into browser-based cryptomining by conducting two large web crawls. A first crawl using `zgrab`, which downloaded the first 256 kB of 137M *.com*, *.net*, and *.org* domains, as well as from the Alexa Top 1M websites. Consequently, the resulting HTML file was checked against the NoCoin [14] block list. A second crawl was performed on a subset of 10M websites, with a customized Chrome browser, instructed to dump WebAssembly modules for further inspection. They conclude their work by stating that 0.08% of the probed websites is actively mining [41].

Another large web crawl study is conducted by Konoth et al. [22] as a study for the creation of *MineSweeper*. Again, the Alexa Top 1M (including three internal pages) was crawled, with a crawler extracting information from all loaded JavaScript and HTML files, WebSocket traffic, and requests made while visiting the website. A total of 1,735 websites was found to be actively mining, the majority of them using Coinhive. 20 mining campaigns were discovered in their analysis, of which the largest involved 139 websites. Based on these findings, a novel detection technique was developed, which focused on the aspects all mining scripts have in common: high CPU cache usage and WebAssembly. They developed *MineSweeper*, able to successfully identify mining scripts based on the CPU's L1 and L3 cache usage and cryptomining characteristics in WebAssembly, thus hardening it against miner obfuscation.

As shown by this summary of related work, most attention of academic investigation has been on detecting these browser-based cryptominers. Multiple studies have shown to be able to detect them with high precision [19, 22, 24, 39–41]. Academic research is less focused on finding campaigns of cryptomining websites, while the online research community (such as Badpackets [31] or Krebs on Security [23]) is particularly interested in finding those relations. The first explorations into this area have been taken by [22], [13] and [39], but campaigns have not been systematically explored in their research. This paper aims to resolve this gap, by focusing on identifying campaigns, methods used in these campaigns and their evolution. We are also interested in the spread of cryptojacking on the Web, but as previous work is mostly crawling (subsets of) the Alexa Top 1M, we will analyze a broader set of websites online. In this paper we will not try to create a new detection method, but we build upon the work of [22] to perform our crawls.

5 Methodology

In a measurement study like this, suitable datasets and methods are essential for conducting proper research. In this section we first discuss the datasets used or created, followed by a summary of our crawler implementation.

5.1 Dataset creation

In our first crawl, we focus on finding campaigns of cryptojacking websites. Previous work of [19, 22, 39, 41] mainly investigated the popular parts of the Internet by crawling the Alexa Top 1M, or subsets of it. But, as pointed out by Scheitle et al., the Alexa Top 1M is not the only list measuring the popular Internet and the method Alexa uses to create this list raises questions whether it is the most reliable list to use for research on cryptojacking [44]. To overcome this issue, we have decided to use the union of three top lists on the Internet; the Alexa Top 1M [1], the Cisco Umbrella 1M [4] and the Majestic 1M [25], all using different measurement strategies, to include the popular part of the Internet in our dataset. These last two also include subdomains and domains not serving a web page. Therefore, we have only added the domains to the list of URLs to be crawled and omitted the subdomains from the latter two. Since we are interested in finding as many cryptojacking domains as possible for our campaign analysis, we have decided to extend our list even further with a list of websites gathered from querying PublicWWW – a source code search engine – with the keywords listed in Appendix A.

Table 2: Dataset creation for the campaign focused crawl

List	No. of websites	Date (2018)
Alexa Top 1M	1,000,000	Dec 24
Cisco Umbrella 1M	233,145	Dec 24
Majestic 1M	897,767	Dec 24
Custom PublicWWW set	87,051	Nov 23 – Dec 24
Total	**1,896,503**	

The union of these sets formed the dataset to be crawled and consisted out of 1,896,503 websites (unique effective TLDs + 1), as listed in Table 2. To estimate the prevalence of cryptojacking on the Internet in general, we will not use a top list as the Alexa Top 1M, because it is not a random sample of the Internet. We therefor also download a random sample of ~20% of the websites in 1,136 TLDs. We discuss this crawl in more detail in Section 7.

Operator NetFlows While the aforementioned datasets provide insights into the landscape of cryptomining installations at a given moment, these data sources do not reveal much about the actual usage of such services. In order to bridge this gap, we analyzed NetFlow traces from the network of a Tier 1 operator from September 2017 until December 2018, which were collected at a 1:8192 sampling ratio. For our analysis, we obtained NetFlow records for all traffic from and to the various WebSocket proxy servers belonging the mining services. Although NetFlows do not reveal the actual contents of a connection, the used ports and packet sizes can indicate connection types. The identity of the source connecting to the WebSocket proxy is however irrelevant, and was anonymized to a pseudo-random value by the operator using the CryptoPan algorithm [54].

5.2 Crawler implementation

As mentioned in Section 4, this research builds upon the work of Konoth et al. [22]. Therefore, we have used their crawler implementation as a starting point for our crawler. The following paragraphs will highlight the major changes and additions made to their work for our research.

Addition of new miner applications The publicly available *Minesweeper* crawler supports 22 different mining applications. Based on previous work and online research, we have added another 9 miner applications to the crawler, in order to also identify the newest miner applications. The added applications and their keywords are listed in Appendix B. For some of the already supported miner applications we have extended the fingerprints and improved the regular expressions to find *siteKeys*.

Active mining detection We have instructed the crawler to never explicitly consent to any mining operation. Therefore, we define that website to be actively mining without consent when: a mining code signature is found, together with a *siteKey*, more than two WebWorkers and a WebSocket connection, or, when the Stratum protocol communication or login credentials for a mining pool are found in WebSocket traffic. If one of these conditions holds, we mark the domain as actively cryptojacking.

WebSocket stack trace The miner application communicates with the mining pool using WebSocket connections. WebSocket traffic was already logged in the crawler, but the initiator of the WebSocket connection was not. By inspecting the stack trace of the WebSocket initiation, we can determine which script was responsible for opening the WebSocket connection and therefore the mining initiator. Using this method, we can easily distinguish between miners started from the main HTML page or the ones hidden inside other resources. Moreover, similar stack traces are a powerful indicator for campaign analysis, since it shows what component started the mining application. We have used this method successfully in our campaign analysis to identify attack vectors. Miners hidden inside third-party software such as WordPress are easily noticed in the stack trace, as we will show in Section 6.1.

Changed logic and exhaustive key finding Our crawler visits every website twice. First, by using a custom Chrome build, with the `-dump-wasm-module` flag enabled to dump any WebAssembly on the page. If present, these Wasm modules are analyzed for cryptojacking code by the *MineSweeper* application. Second, by using another Chrome build, which visits the website and saves every file it encounters. Instead of visiting 3 internal pages (as Konoth et al. did), we instructed the crawler to visit just one internal page. Besides that, we have implemented a more exhaustive *siteKey* search. The crawler first searches for fingerprints of known miner applications and afterwards for the *siteKey* in the following order: WebSocket traffic, the HTML page and finally in all other HTML and JavaScript resources. A minor addition has been made to automatically decode a base64 encoded *siteKey* of the Mineralt miner [27]. This addition allowed us to retrieve more *siteKeys*, which improves the campaign analysis afterwards.

5.3 Infrastructure

We deployed the crawler in Docker containers on 60 servers within the university network, each running 8 Docker instances in parallel. The crawl started on the December 24, 2018 and completed on January 9, 2019. In total, 1,769,183 websites have been successfully visited in this initial crawl. Afterwards, we have performed a second crawl using the same infrastructure, which we discuss in Section 7.

Table 3: Summary of the results of the first crawl

Crawling period	24/12/2018 – 9/1/2019
# websites crawled	1,769,183 (93%)
# potential cryptojacking websites	21,022
# active cryptojacking websites	10,100
# active miner applications	22
# websites with unknown miners	323
# cryptojacking campaigns identified	204
# websites in largest campaign	987
# websites in Alexa Top 1M	648 (0.065%)
# websites in Cisco Umbrella 1M	109 (0.047%)
# websites in Majestic 1M	506 (0.056%)

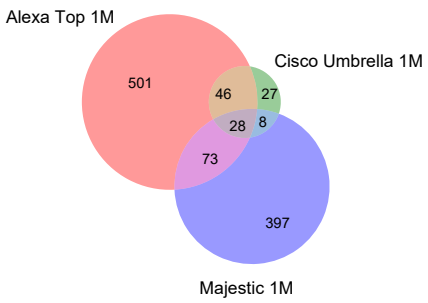

Figure 2: Venn-diagram showing the distribution of identified cryptojacking domains over the used top lists

6 Current state of cryptojacking campaigns

We have identified 21,022 websites with traces of cryptomining activities of which 10,100 websites are actively mining without the visitor's explicit consent. Only 648 of these websites are listed in the Alexa Top 1M. 22 different miner applications have been identified among the crawled websites, most of them running at least the Coinhive miner application (71%). Also, 509 websites are deploying multiple miners. For 323 websites, the used miner application could not be detected, which indicates heavily obfuscated or unknown miner applications. The results are summarized in Table 3.

Among the identified websites, 204 campaigns have been detected, of which the largest one covers 987 websites. This number of campaigns is a magnitude larger compared to previous work [22, 39]. We have identified the use of third-party software, such as Drupal and WordPress, to be the driving factor behind the largest cryptojacking campaigns.

Mining with consent There are two mining applications focused on mining solely with visitor consent. JSEcoin, a mining service presenting itself as *"The future blockchain & ecosystem for ecommerce and digital advertising"*, allows website owners to let their users mine JSE tokens, after explicit opt-in consent [20]. Another consent-focused mining application is AuthedMine, the opt-in version of Coinhive, introduced after adblockers started blocking Coinhive [8]. In our crawl, we have identified 2,477 websites using the JSEcoin miner and 227 websites using AuthedMine. None of the websites using AuthedMine opened a WebSocket connection, which indicates that no mining activity took place. 143 websites using JSEcoin did however open a WebSocket connection, but never actually started mining. By analyzing the WebSocket traffic, we observed that in most cases the WebSocket connection initiation was followed by two probes sent back and forth, waiting for the user to opt-in. Since these mining applications did not started mining without consent of the visitor, we have omitted them from our results.

Identified domains in top lists Of the 10,100 domains identified as actively cryptojacking, only 925 were found in one of the three top lists. The Alexa Top 1M contains the most cryptojacking domains (648), meaning that 0.065% of the websites in the Alexa Top 1M are cryptojacking, slightly less than previous work [22, 41]. For both other lists this number is lower. The addition of the Cisco Umbrella 1M resulted in only 27 additional findings, whereas the addition of the Majestic 1M led to the discovery of 397 new cryptojacking domains. In Figure 2, a Venn diagram depicts these differences in subsets. Only a small number of websites is shared among the Alexa Top 1M and the Majestic 1M. Also note that 9,175 (86%) of the identified websites are not listed in any of these top lists. This finding stresses the necessity of looking further than top lists while performing campaign analysis and to study the current state of cryptojacking on the Internet.

Categorization of websites We have discovered various sorts of cryptojacking websites on the Internet. By complementing the list of identified domains with website categorization data of Webshrinker [51], we categorized each cryptojacking website. We confirm previous work by identifying adult content (such as pornography) as the most prevailing category within our dataset, with over 2,000 websites in this category. Illegal content, a category known being home to abusive web resources, contains a lower number of cryptojacking websites compared to what we expected.

Installation base Coinhive is still the most popular cryptomining application installed on the identified cryptojacking websites (75%), followed by Cryptoloot (5.3%) and CoinImp (3.2%). But, there are noticeable differences between the complete crawl and the subset of domains in the Alexa Top 1M. Coinhive's share is halved, whereas CoinImp and Cryptoloot installations are doubled in size. Nerohut and Webminerpool miners are relatively more present in the Alexa Top 1M subset, while Mineralt has a similar share in that subset. The bottom

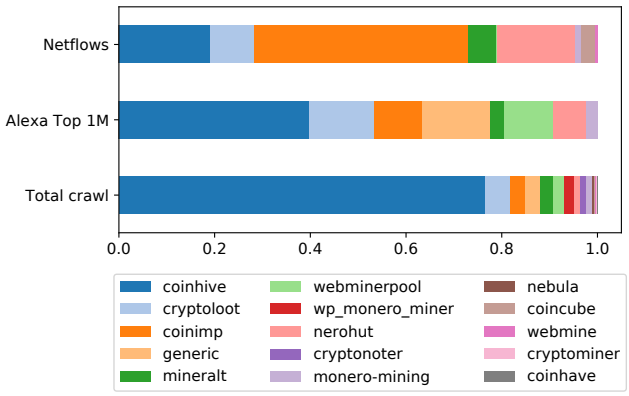

Figure 3: Distribution of cryptomining applications based on the total crawl, the Alexa Top 1M and NetFlows analysis.

two stacked bars in Figure 3 show the distribution of miners according to our analysis.

We have also discovered services which combine multiple cryptomining applications. The most popular mining combination is the set of Coinhive, Cryptoloot and Cryptonoter, which are bundled in the implementation of the WordPress Monero Miner plugin [21]. A combination of a Nerohut miner with a Cryptoloot or Webminerpool miner is also regularly encountered. Usually, only one miner starts (due to another script deciding which one to use), but we also encountered domains on which multiple miners were started concurrently.

Actual mining activity The distribution of mining applications installed on domains gives an insight into their popularity by actors pursuing cryptomining, but not into their actual usage. The amount of actual mining that takes place can however be estimated by tracing the connections website visitors make to the mining application's WebSocket proxy, as explained in Figure 1. We obtained a trace of connections transported by a Tier 1 network operator in 1:8192 sampling for a period of 14 months, and followed the WebSocket proxy server IPs from these mining applications to estimate the traffic to these servers. This gives an insight into how much traffic these WebSocket proxies digest, and is therefore a more reliable source for popularity measures. The upper stacked bar chart in Figure 3 shows the distribution of NetFlows to the WebSocket proxy servers of known mining applications for the month of December. The results show a drastic difference between installation base and mining traffic: while Coinhive is found on most websites, CoinImp proxy servers handle more than twice as much traffic than the dominant application. WebSocket traffic to servers of Cryptoloot is similar in size compared to its installation base.

Table 4: Mining pools the identified domains are mining in

Mining pool	Occurrence
`supportxmr.com`	93
`xmrpool.eu`	15
`greenpool.site`	13
`minexmr.com`	6
`xmr.omine.org`	4
`moneroocean.stream`	2
`seollar.me`	1
`xmr.nanopool.org`	1

Mining pool participation Most mining applications do not disclose the actual mining pool they are mining for in WebSocket traffic. However, on 135 identified domains, the WebSocket traffic did reveal that, as listed in Table 4. Most of these websites are participating in the `supportxmr.com` mining pool, which is commonly orchestrated by a Webminerpool or Nerohut mining script. Other pools are less commonly used or were not revealed in WebSocket traffic.

Throttling of applications Most cryptomining applications allow for a throttle value to be set, which limits the percentage of the CPU the miner can use. It is not necessary to set a throttle value, in this case the miner uses 100% of the available processing power. We have discovered that when a throttle value is set, this is often set to 0.3, meaning that 70% of the processing power can be used by the miner. Setting a throttle to use 70% of the resources seems to be balancing between gaining enough profit and not disturbing the browsing experience too much. In the identified campaigns, the throttle value is mostly set to the same value on all domains. An exception is listed in Table 5, in which a campaign involving 180 websites uses two different throttle values.

Attack vectors encountered We were able to retrieve the *siteKey* of actively cryptomining websites in 92% of the cases. Most of the gathered *siteKeys* are only used once (78%) and only a small portion (5%) is used on more than 5 different websites. However, the *siteKeys* in this last category are found on 4,663 different websites (46% of the total). The high number of *siteKeys* used only once suggests a large amount of website owner initiated cryptojacking, since every domain uses its own key. The fact that almost half of the websites is part of a campaign involving at least 5 websites also indicates different attack vectors. We have manually analyzed the used *siteKeys* in the latter category, and we can conclude that, besides website owner initiated cryptojacking, the use of third party software is a prevailing attack vector. Third-party applications like WordPress, Drupal or Magento are often abused to spread cryptojacking injections. These applications play a major part in campaign analysis, as discussed in Section 6.1.

Hiding techniques With the rise of cryptomining blocking applications such as NoCoin [17] or Minerblock [16], mining scripts are more often hidden to prevent detection. We have encountered a number of hiding techniques in our crawl and distinguish the following levels of obfuscation:

1. *No obfuscation.* The script is loaded in clear text, key and other options are visible to the user.

   ```
   var miner = new CoinHive.Anonymous('key');
   miner.start();
   ```

2. *Limiting CPU usage.* Script is loaded in clear text, key and other options are visible to the user, but CPU usage is throttled, so detection by the user is less likely.
3. *Renamed variables.* The script is loaded in clear text, but (some) variable names have been changed. These variable names are either replaced by random strings, or by completely different words, such as on `http://www.2001.com.ve/`:

   ```
   startHarryPotter("boddington", "2001");
   ```

4. *Renamed mining script.* The loaded script is still in clear text, but hosted on the web server itself instead of fetched from a mining service. The file name is changed to prevent blacklist blocking, frequently to general names, such as `jquery.js` or `stat.js`.
5. *Hidden inside other scripts*. The miner is appended or inserted into another script. The benign script still functions as normal, but also starts up the mining process.
6. *Obfuscated code.* The loaded scripts are masked by a code obfuscator and contain packed or CharCode code. All application-specific strings are encoded, stored in an array and variable names are replaced by random strings.

   ```
   var _0x5d02=["\x75\x73\x65\x20\x73\x74", ..]
   ```

7. *Obfuscated code and WebSocket traffic*. The loaded script is obfuscated by a code obfuscator and WebSocket traffic is sent encrypted to the proxy server.
8. *Obfuscated and hidden*. Scripts are hidden inside other files and/or via multiple redirects. Every script is randomly named and obfuscated, and so is the WebSocket traffic. WebAssembly is not retrieved from the server, but included inside the script.

In our crawl, most website owner initiated cryptojacking is not obfuscated, often not even throttling CPU usage. Attacks using third-party software are usually hiding cryptomining code inside other scripts and apply some obfuscation. We have encountered multiple WordPress themes and Drupal plugins with such a hidden miner. Only 391 websites with encrypted WebSocket have been identified, whereas most websites are using plain text Stratum communication. The highest level of obfuscation is rarely encountered.

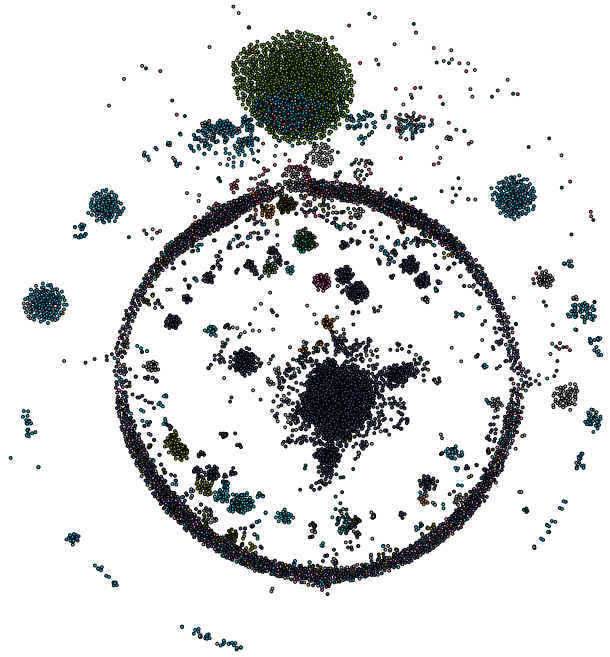

Figure 4: Relationships between the identified cryptojacking domains depicted in a force-directed graph

6.1 Cryptojacking campaigns

We have identified 204 cryptojacking campaigns, covering 5,733 websites, meaning that 57% of all cryptojacking websites encountered are part of a campaign. We define a cluster of more than 5 websites to be a campaign, as stated in Section 2. Figure 4 shows all the identified cryptojacking domains in a force-directed graph, where domains with similar features attract each other, colored according to the used application. Clear clusters can be distinguished, such as a Monero-Mining campaign shown in pink and a large Mineralt campaign shown in green right above it. Coinhive, the application used the most, is shown in dark blue with multiple large clusters all over the graph. The circle represents the cryptojacking domains not part of a campaign. In the following paragraphs, we highlight our findings based on different possibilities for identifying campaigns as introduced in Section 2.

Found on shared siteKey We were able to successfully retrieve the *siteKey* of 92% of the actively cryptojacking domains, which enabled us to cluster domains sharing the same *siteKey*. A shared *siteKey* guarantees that the rewards for mining will be transferred to the same account. We have identified 192 cryptojacking campaigns based on the same *siteKey* being installed on more than 5 different websites. As shown in Table 5, the largest campaign covers 987 websites, all using WordPress. A variety of plugins and themes include a malicious file named `jquory.js`, which is responsible for starting

a Coinhive miner. A similar attack vector is observed in a campaign involving 317 Drupal websites. This campaign is part of the Drupalgeddon 2 and 3 attacks, which took advantage of major remote code execution vulnerabilities in Drupal to inject their malicious scripts [45]. The only large campaign using the Mineralt miner, also focused on WordPress, has base64 encoded its *siteKey* inside the `script` tags. This makes them seem different, but match once decoded, since only the throttle value is changed. Not just vulnerabilities in CMS systems are used to spread cryptojacking code, also Magento, an e-commerce system, is involved in a Coinhive mining campaign targeting 175 websites in our crawl. The largest campaign using the compromised websites attack vector involved 376 Chinese websites, which share a miner script injected on the bottom of the page. A provider of The Pirate Bay proxies orchestrates the largest website owner initiated campaign on our list, with 70 proxy domains using the same Cryptoloot miner. These findings indicate that the most successful and largest cryptojacking campaigns are created by abusing third-party software.

Found on shared WebSocket proxy server Most cryptojacking campaigns are using the infrastructure of popular applications, such as Coinhive, to connect to a mining pool. Thus, clustering domains on these WebSocket proxy servers will not create meaningful clusters. However, when we discard these popular proxy servers, we are able to identify another 12 campaigns, which have not already been identified by shared *siteKeys*. Those are listed in Table 6. A Coincube miner campaign involving 27 websites uses `coin-services.info` as a WebSocket proxy server on a variety of ports. This campaign hosts its miner scripts on code repositories such as GitHub and BitBucket, where a number of accounts is created to host the miner files, which are all named `main.js`. On one of the GitHub accounts, even a picture of stacked Ukrainian money can be found [15]. 28 very similar websites, all offering illegal video streams, were found to be using a WebSocket proxy server on `wss://ws**.1q2w3.life/proxy` with, after manual inspection, `seriesf.lv` as the accompanied *siteKey*. This proxy server was also discovered by [22] on 5 websites in their crawl. They estimated that this campaign made a profit of $2,012.90 per month, which is likely to be a lot more, since we have found almost 6 times as many domains involved in this campaign. We have discovered that websites using a private WebSocket proxy are more likely to hide their activities by using higher levels of obfuscation.

Additionally, we have discovered 14 WebSocket proxy servers with very similar addresses on 75 domains (e.g. `nflying.bid, flightzy.bid` and `flightsy.bid`). These servers are contacted by the most obfuscated miner encountered in this crawl. The miner code is hidden inside a randomly named file, the miner code is heavily obfuscated and the WebSocket traffic is sometimes encrypted. Our efforts to reverse engineer the obfuscated miner code are so far unsuccessful. Therefore, we can not cluster them as being a campaign based on the shared proxy servers, but we have added the signature to our crawler as a separate mining application for the next crawls.

Found on shared initiator file In our crawling process, the stack trace of an initialized WebSocket connection is saved for every website. While examining these stack traces, some file names emerged and lead to the identification of another 4 cryptojacking campaigns. The oddly named file `gninimorenomv2.js`, responsible for opening WebSocket connections on 24 websites seemed to be part of a malicious advertisement campaign, which injects cryptojacking scripts into served advertisements. As shown in Table 6, this file opens a connection to `wss://heist.thefashiontip.com:8182/` to earn the profits from the displayed mining advertisements. Another campaign was identified by grouping the websites in which `adsmine.js` was responsible for opening a WebSocket connection. These websites turned out to be 17 very similar pornography websites, which indicates that this campaign is website owner initiated. The newly discovered mining application, as described in the previous section, served obfuscated mining scripts to its miners. Although obfuscated, inspection of the random file names revealed clusters of websites injected with the same randomly named miner, which lead to the discovery of another 3 campaigns, all targeting solely WordPress websites.

Found on shared mining pool login Most miner applications submit their solved hashes to a WebSocket proxy server, which combines the hashes of multiple miners before forwarding it to the actual mining pool. However, we have discovered 238 websites directly submitting their hashes to a mining pool. These websites are using only six unique cryptocurrency wallet addresses. The shared wallet addresses guarantee that profits made by cryptojacking are transferred to the exact same wallet. These findings did not lead to the discovery of any new campaigns, but did confirm previous findings. E.g., proxy `wss://delagrossemerde.com:8181/` (used by 15 sites) is solely receiving traffic from domains using the same wallet.
The different methods used in this section enabled us to find 204 cryptojacking campaigns. We can conclude that the largest campaigns are using third-party services like WordPress, Drupal or Magento as their method of spreading. Only one campaign using advertisements with injected cryptojacking scripts has been identified, this in contrast to previous work by [22, 39], who reported malicious advertisements as a significant attack vector. Compromised websites or website owner initiated campaigns are generally smaller in size. The obfuscation level used in most campaigns is rather low, heavily obfuscated code is encountered rarely and in more than half of the identified campaigns a miner added in plain text.

Table 5: Identified campaigns based on a shared *siteKey* (HT = hiding technique encountered)

SiteKey	#	Type	Attack vector	HT
I2OG8vG[..]coQL & hn6hNEm[..]w1hE	987	Coinhive	Third-party software (WordPress)	5
I8rYivhV3ph1iNrKfUjvdqNGfc7iXOEw	376	Coinhive	Compromised websites	2
oHaQn8u[..]EvOS, XoWXAWvi[..]JfGx, no2z8X4[..]w2yK	317	Coinhive	Third-party software (Drupal)	2
TnKJQivLdI92CHM5VDumySeVWinv2yfL	213	Coinhive	Third-party software (WordPress)	1
GcxML3FZ;60;1 & GcxML3FZ;-70;1	180	Mineralt	Third-party software (WordPress)	6
ZjAbjZv[..]9FiZ, PQbIwg9H[..]gfVW	175	Coinhive	Third-party software (Magento & WordPress)	4
w9WpfXZJ9POkztDmNpey3zA1eq3I3Y2p	103	Coinhive	Compromised websites	2
j7Bn4I56Mj7xPR2JrUNQ9Bjt6CeHS3X1	79	Coinhive	Third-party software (WordPress)	2
cb8605f33e66d9d[..]6af74f86e6882899a8	70	Cryptoloot	Website owner initiated (The Pirate Bay)	2
49dVbbCFDuhg9nX[..]K2fkq5Nd55mLNnB4WK	70	Coinhive	Compromised websites	1

Table 6: Identified campaigns based on shared WebSocket proxy servers (HT = hiding technique encountered)

WebSocket proxy server	#	Type	Attack vector	HT
wss://ws**.1q2w3.life/proxy	28	Nebula	Website owner initiated	6
wss://coin-services.info:****/proxy	27	Coincube	Compromised websites	6
wss://heist.thefashiontip.com:8182/	24	Webminerpool	Malicious advertisements	5
wss://delagrossemerde.com:8181//	15	Webminerpool	Website owner initiated	8
wss://wss.rand.com.ru:8843/	13	Coinhive	Third-party software (WordPress)	8
ws://185.165.169.108:8181/	8	Webminerpool	Website owner initiated	2
ws://68.183.47.98:8181/	7	Webminerpool	Website owner initiated	2
wss://gtg02.bestsecurepractice.com/proxy2/	6	Unknown	Third-party software (WordPress)	3

6.2 A in-depth campaign search

The sizes of the campaigns identified in Section 6.1 depend on the dataset we crawled, so they could have been incomplete. To find more websites belonging to the identified campaigns, we have taken the indicators of compromise for a large number of campaigns and queried PublicWWW for domains matching these IoCs. This resulted in a dataset of 7,892 websites. Combined with the 21,022 potentially cryptojacking websites from the initial crawl, a total of 25,121 URLs was crawled on February 12, 2019, more than a month after the initial crawl. We successfully obtained 24,187 (96%) of them.

Most of the campaigns remained of similar size in this crawl, except for a campaign involving three keys, ef937f99557277ff62a6fc0e5b3da90ea9550ebcdfac, 06d93b846706f4dca9996baa15d4d207e82d1e86676c and dd27d0676efdecb12703623d6864bbe9f4e7b3f69f2e. This advanced campaign is targeting domains using Bitrix24, a CRM platform used by a variety of organizations. The most remarkable website it has been found on is the website of the Ministry of Education of Belarus (https://edu.gov.by/). The malicious code is hidden as the core loader of Bitrix24 and uses both Nerohut and Cryptoloot to mine with. It has a built-in anti-detection method, since it stops mining once a developer tools window is opened. In our initial crawl, we have identified only 68 domains belonging to this campaign, which turned out to be 855 in our in-depth search, making this campaign the second largest campaign we have identified so far. Another campaign, involving key vPfPDHk89TxmH1arysiJDrutpYGntofP, is displaying fake loading screens on 86 websites, whereas only 47 of these have been identified in our initial crawl.

All other campaigns remained similar or slightly smaller in size. Except for the two aforementioned campaigns, we conclude that our initial crawl likely identified the correct size of campaigns, given the database of PublicWWW. Their database contains source code snapshots of over 544M websites, which should provide a proper approximation.

6.3 Evolution of cryptojacking

To study the evolution of cryptojacking on the Internet, data is needed from different moments in time. Fortunately, Konoth et al. [22] shared their crawling results and Hong et al. [19] shared their list of identified cryptojacking domains, which made it possible for us to crawl these exact same sets of URLs and to analyze whether these domains were still mining. Additionally, we have followed the domains identified in our crawls over a period of 3 months, and analyzed WebSocket proxy traffic over time using operator NetFlows.

Comparison with previous crawls Konoth et al. [22] crawled from March 12 until 19, 2018 and identified 1,735 potential cryptojacking domains. We crawled their list on January 21, 2019 and obtained 1,725 of them. 85% of the

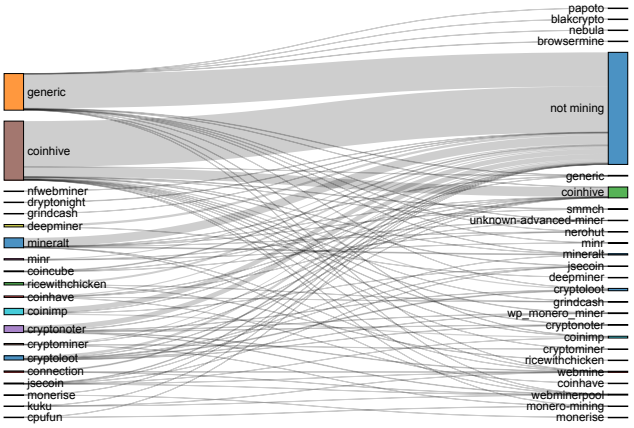

Figure 5: Usage evolution between March 2018 and January 2019 in the list of identified domains by [22]

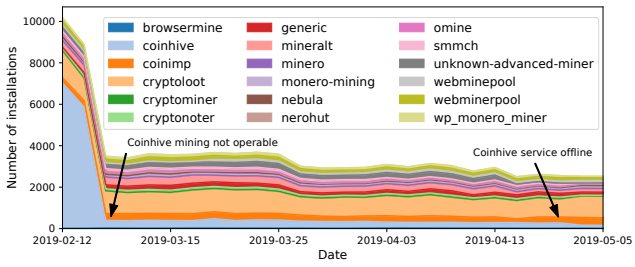

Figure 6: Evolution of the cryptojacking domains per type

websites are not cryptomining anymore, and only 10% is still using the same application. On 136 websites (7%), the same key was found in both crawls. As Figure 5 shows, a large number of websites using a Coinhive miner removed the miner application. Some continued using Coinhive, but also a small shift into less popular mining applications can be observed. Websites already using these miners tend to stick to their choice and are still using the same miner almost a year later. We have also seen a number of mining applications become extinct, such as Deepminer and NF Webminer. Hong et al. [19] also published the list of identified cryptojacking domains from their crawl in February 2018. A year later, on February 12, 2019, we have crawled this list of 2,770 domains. We obtained 2,435 (88%) of them and only 340 (14%) domains are still actively cryptojacking. Both crawls show that a large number of websites stopped cryptojacking themselves or removed the miner infection. After one year, approximately 85% of the domains are not actively cryptojacking anymore. We have also observed a small portion of domains switching to less popular applications. The low number of 7% of websites that are still mining with the same *siteKey* indicates the fast changes in the cryptojacking threat landscape.

Evolution of identified domains We have followed all previously identified cryptojacking domains for a period of 3 months (until May 5, 2019) and crawled them initially occasionally, but afterwards every other day. Within this time period, Coinhive announced to end its mining application, due to decreased Monero prices and hash rate [7]. The announcement was made on February 26, 2019 and stated that mining would not be operating anymore after March 8, 2019, and that the service would be discontinued by the end of April 2019. This lead to a drastic change in the cryptojacking landscape, as Coinhive's dominance in actively mining installations collapsed when their mining service was set non-operationally. Mining applications were however not massively replaced, which confirms our finding that a large portion of browser-based cryptomining is not initiated by the website owner. Only when the Coinhive mining service was actually discontinued and errors were shown while requesting the offline Coinhive mining resources, we observe a small increase in Cryptoloot and CoinImp installations.

WebSocket proxy traffic over time As discussed in Section 2, most miner applications use a WebSocket proxy server to forward traffic from their miners to the mining pool. Using NetFlow data mentioned earlier, we have analyzed traffic towards popular WebSocket proxies from September 2017 till December 2018, which gives an insight into the evolution of cryptomining applications usage, as shown in Figure 7. We have taken the set of WebSocket proxy IPs the miners connect to as a basis, which we extended by using passive DNS data to discover other WebSocket proxy server IPs used by these applications, but hosted on different servers, not encountered during our crawls. The same passive DNS data was used to verify whether these IP addresses were solely used as WebSocket proxy servers. To prevent other traffic to these servers from being in our dataset, we have both set the maximum packet size to 550 kB and verified that only WebSocket traffic was counted towards these servers. For most proxies, this is traffic towards port 80 or 443, and for a few servers using specific ports, this could be different. An example is the WebSocket proxy server of the WP-monero-miner which uses port 8020.

The blue line from September 2017 on shows how the web-mining ecosystem is monopolized by innovator Coinhive at the start, where after copycats like Cryptoloot and Webmine start to emerge in October. We see that CoinImp essentially starts to eclipse all other miner applications from mid April 2018 onwards in terms of mining traffic to the proxies, which is unexpected given the distribution of installations on websites and previous studies. Some mining proxies only have transient success: a remarkable example is the WP-monero-miner, released shortly after Coinhive in 2017. The applica-

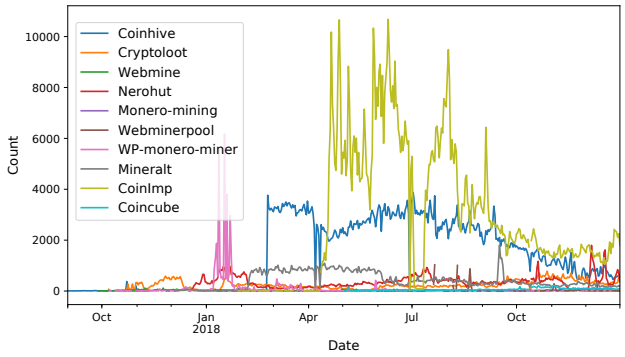

Figure 7: Number of NetFlows involving WebSocket proxy servers for popular miners between Sep 2017 and Dec 2018

tion hosts its own mining pool and digested a lot of traffic in January 2018, only to almost disappear again weeks later. Coinhive, the application used by most websites, is a constant factor in the miner landscape with over 4,000 NetFlows a day in mid 2018 (given our 1:8192 sampling, thus 32M connections per day), but not as large as one would expect from its installation base. Additionally, a clear declining trend can be observed in the NetFlow counts to all mining services after the summer of 2018. The last months of NetFlow data show a diverse set of mining applications actively used.

7 An Internet scale study on cryptojacking

In order to estimate the prevalence of browser-based cryptojacking on the Internet and to indicate any differences between Top Level Domains (TLDs), we have performed another crawl, in which we have crawled ~20% of the websites belonging to each of the 1,136 existing TLDs. We obtained a daily zone transfer for all generic top level domains (gTLDs) – such as *.top*, *.loan* – from the Internet Corporation for Assigned Names and Numbers (ICANN), as well as a feed of registered country code top-level domains (ccTLDs) – such as *.uk*, *.jp*, or *.ru* – from a security intelligence provider. From these lists, we randomly picked a sample of ~20% of the size of each TLD [12]. Based on the results of the previous crawl, we have added another 5 mining applications to the crawler implementation, as listed in Appendix C. From January 11 until April 3, 2019, we crawled the random sample including 48.9M domains. This yielded a total of 125 TB of network traffic.

7.1 General findings

After crawling a random sample of 48.9M websites in a large number of different top level domains, we are able to draw conclusions about the prevalence of browser-based cryptojacking on the Internet. We estimate that 0.011% of all domains are actively cryptomining without their visitors' explicit consent,

Table 7: Distribution of cryptomining applications installations in the Internet scale crawl (sum of percentages is >100%, because of websites using multiple applications)

Type	# of websites	Percentage
Coinhive	2,531	48.767%
Unknown	689	13.276%
CoinImp	513	9.884%
Cryptoloot	504	9.711%
Mineralt	276	5.318%
Nerohut	247	4.760%
Webminerpool	233	4.489%
Unknown-advanced-miner	92	1.773%
SMMCH	80	1.541%
Browsermine	73	1.407%
Webminepool	62	1.195%
WP-Monero-Miner	60	1.156%
Omine	56	1.079%
Monero-mining	55	1.060%
Cryptonoter	50	0.963%
Cryptominer	26	0.501%
Minero	24	0.462%
Nebula	23	0.443%
Webmine	19	0.366%
Coincube	19	0.366%
Project-poi	4	0.077%
Adless	1	0.019%

meaning that one in every 9,090 websites is cryptojacking. Comparing this number to the statistics of the top lists used in our initial crawl, we conclude that cryptojacking activity is mainly focused on the popular parts of the Internet. In the Alexa Top 1M, 0.065% of the websites was actively cryptojacking, in this random sample only 0.011% of the websites, which is almost 6 times lower. This can be explained by the lucrativeness of cryptojacking, in which a higher popularity means more visitors, yielding more potential miners and thus higher potential profits. Additionally, it shows that researching the prevalence of cryptojacking by crawling the Alexa Top 1M overestimates the problem size. However, the distribution of used applications in our random sample is fairly similar to the distribution in the Alexa Top 1M. The distribution of mining applications in this crawl is listed in Table 7.

The categories of domains identified in this crawl are very similar to the initial crawl. As depicted in Figure 8, *Adult content* remains the most prevailing category, while other large categories are *Technology* and *Under Construction*, the category involving parked, expired or yet-to-be developed domains. Based on these two very different crawls we can conclude that cryptojacking is indeed more prevailing on domains hosting adult content.

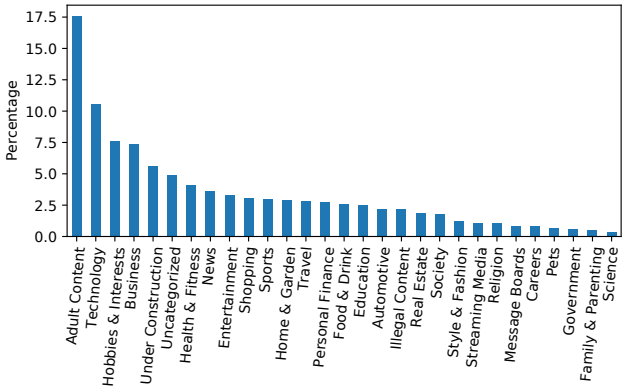

Figure 8: Categories of mining domains in the second crawl

7.2 Cryptojacking on different TLDs

We have crawled domains of roughly ~20% of 1,136 different TLDs in order to analyze the prevalence of cryptojacking. As Table 8 shows, cryptojacking activity varies enormously within different TLD zones. The four largest TLDs, *.com*, *.de*, *.net* and *.org* have a similar percentage of cryptojacking websites, but we have discovered almost 6 times as much cryptojacking activity in the Russian TLD. Also, domains in the Brazilian and Spanish zones are more susceptible to cryptojacking, having respectively 4 and 3 times more cryptojacking activity than average. On the contrary, the *.top*, *.us* and *.loan* zones host only a few cryptojacking websites.

Our website category analysis showed that adult content is the most prevailing category for cryptojacking activities. This triggered our attention for the *.xxx* domain, which is specially created for adult content, which we therefore crawled completely instead of ~20%. Surprisingly, the *.xxx* domain contains only one website actively cryptomining.

When comparing used mining applications on the different TLDs, large differences can be distinguished, as shown in Figure 9. Coinhive is the most popular miner in most zones, whereas Cryptoloot is preferred in the Russian zone, and French and Czech websites contain more Nerohut miners. The Russian zone is also the only TLD where browsermine is used regularly. The high number of generic miner applications in the Dutch and Belgian zone is remarkable. A large number of these domains in the *.nl* and *.be* zone are part of a campaign using expired domain names of a Dutch registrar (*Totaaldomein B.V.*) to host porn and unknown cryptominers.

Our results show a different popularity of used mining applications compared to previous work of [41]. They detected Coinhive on 85% to 90% of the *.com*, *.net* and *.org* TLDs, whereas we determine that this market share is significantly lower (~50%). This result proves that a simple solution like the NoCoin block list is unable to detect all miners and analyses with such techniques result in different outcomes.

Table 8: Results of the TLD crawl. Listed are the top 10 largest domains, followed by remarkable TLDs

TLD	Size	Crawled	Cryptojacking
.com	149,937,597	27,555,546 (18.4%)	2,353 (0.009%)
.net	15,008,406	2,741,550 (18.3%)	238 (0.009%)
.de	15,089,860	2,244,139 (14.9%)	254 (0.011%)
.org	11,330,764	2,021,630 (17.8%)	145 (0.007%)
.info	6,524,248	1,309,323 (20.6%)	77 (0.005%)
.ru	5,480,467	998,422 (20.0%)	593 (0.059%)
.nl	5,360,173	880,122 (16.4%)	191 (0.022%)
.top	4,024,497	788,748 (19.6%)	19 (0.002%)
.br	3,813,745	383,910 (10.1%)	185 (0.048%)
.fr	3,449,775	567,887 (16.5%)	133 (0.023%)
.pl	2,621,515	523,497 (20.0%)	81 (0.015%)
.us	2,409,802	472,323 (19.6%)	2 (0.000%)
.loan	2,228,165	445,749 (20.0%)	0 (0.000%)
.es	2,010,710	327,810 (16.3%)	110 (0.036%)
.online	1,105,999	219,447 (19.8%)	67 (0.031%)
.pro	295,201	58,999 (14.2%)	32 (0.054%)
.space	268,846	53,363 (20.0%)	19 (0.036%)
.website	276,063	54,704 (19.8%)	21 (0.038%)
.xxx	93,101	91,877 (98.7%)	1 (0.001%)
Total		**48,948,669**	**5,190 (0.011)%**

8 Discussion

Crawling the Internet inevitably comes with its shortcomings. Limitations in the crawler implementation, network used and analysis can produce both false positives and negatives. The latter category can occur for example when extreme obfuscation is used, as we have seen in Section 6. However, we believe that due to our double crawling strategy, based on both WebAssembly and code signatures, this could not have happened very often. Finally, the use of worldwide NetFlow traffic from a Tier 1 network operator allowed us to analyze the popularity of cryptojacking services in a revolutionary way, although BGP policies, and a specific PoP and IXP footprint could lead to a bias of certain autonomous systems just as some discrepancies might arise due to 1:8192 random sampling. Additionally, since the NetFlows do not reveal the actual contents of the connection, we can never be sure about the contents. However, during our crawls we could confirm the mining applications to contact the WebSocket proxy servers in question, and passive DNS lookups did not show any other domains pointed to that IP. Furthermore, the NetFlows both revealed no traffic to other ports than those seen from our crawlers and packet sizes resembling those observed in our crawls, thus the methodology should provide valid results.

Future work The additional angle provided by the NetFlow data allowed us to study the evolution of cryptojacking over a longer period of time, something which has not been done be-

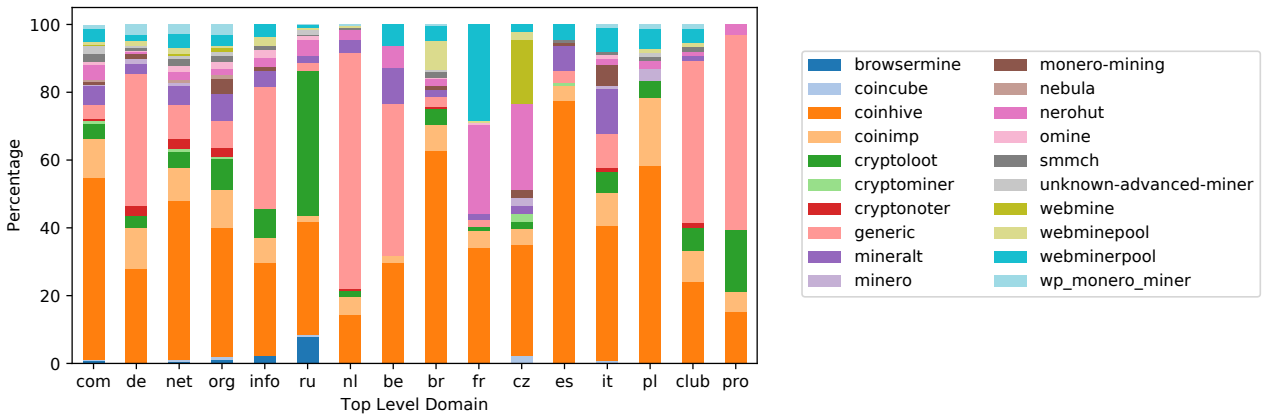

Figure 9: The distribution of used mining applications in various TLDs

fore. Regular crawls of the Internet, especially of the already identified cryptojacking domains gives more insight in this practice, as will it increase the innovation of defense mechanisms. The most influential defense against cryptojacking will nonetheless be frequent patching, as most cryptominers are installed exploiting known vulnerabilities. CMS providers, such as Drupal or WordPress, have shown agility in patching vulnerabilities, but the responsibility of installing these patches remains with the website owner. Finally, as we have seen a decline in the price of Monero (-85% in 2018), we believe that cryptojacking infections on individual websites will decrease, but that cyber criminals will search for other possibilities to exploit cryptojacking at an even larger scale. As we have mentioned in Section 3, the most effective method of collecting large groups of miners is by launching a MITM attack. Investigating the prevalence of this attack vector for cryptomining is something we preserve for future work.

9 Conclusions

In this paper, we have studied the prevalence of cryptojacking as well as of cryptojacking campaigns on the Internet. We have performed multiple large crawls, each with a different focus. In our first crawl, we have analyzed the 1.7M most popular domains to identify organized campaigns. We found 204 campaigns, from which we conclude that the size of cryptojacking campaigns is heavily underestimated by current academic research. Additionally, using solely the Alexa Top 1M shows significantly different results in terms of the size of organized activity and infection rate, which we found to be almost 6 times lower in a random sample compared to the Alexa Top 1M, hence overestimating the problem. Third-party software is often used by attackers to spread cryptojacking scripts over a large number of domains. The share of domains serving advertisements injected with cryptojacking scripts is lower compared to previous work, most likely because of stricter monitoring by advertisement networks. We have seen that obfuscation of cryptojacking scripts is definitely present, but only occasionally used. Comparing our results with data from previous studies (in both February and March 2018) shows that after a year, only 15% of the websites is still actively mining. This, and our novel way of estimating miner application popularity by analyzing NetFlows, led to the conclusion that the cryptojacking landscape is constantly changing and involves a variety of actors.

A second, Internet-scale crawl involving ~20% of 1,136 TLDs (48.9M websites), which represents a truly random sample of the Internet, allows us to conclude that cryptojacking is present on 0.011% of all domains. Not unexpectedly, this percentage increases in the more popular parts of the Internet, because cryptojacking on popular domains is much more lucrative. Both of our crawls have shown that cryptojacking mostly takes place on websites hosting adult content, although the *.xxx* TLD is home to only one cryptojacking website. Based on the applications used within the time span of our analysis, we can conclude that Coinhive was the largest mining application in terms of installation base, but that CoinImp's WebSocket proxy servers were digesting much more traffic in 2018. Looking at the different TLDs, we conclude that Russian, Brazilian and Spanish zones are home to a disproportionate number of cryptojacking domains.

With the discontinuation of Coinhive in March 2019, the landscape of cryptojacking has changed enormously, but based on our results, we are only expecting a further decline in individual cryptojacking activities given that the Monero value keeps diminishing. However, this only stresses the importance of organized cryptojacking campaigns, as cyber criminals will find new ways to spread their cryptojacking infections to still be profitable. Here, campaign analysis will be an important asset: as adversaries are unlikely to develop a unique approach for each infected website, the reuse of resources and methods will provide an effective angle to detect and mitigate these activities.

References

[1] ALEXA. Top 1M sites. http://s3.amazonaws.com/alexa-static/top-1m.csv.zip (December 2018).

[2] CARLIN, D., O'KANE, P., SEZER, S., AND BURGESS, J. Detecting cryptomining using dynamic analysis. In *16th Annual Conference on Privacy, Security and Trust, PST 2018, Belfast, Northern Ireland, Uk, August 28-30, 2018* (2018), pp. 1–6.

[3] CHRISTOPHER, N. Hackers mined a fortune from indian websites, Sep 2018. https://economictimes.indiatimes.com/small-biz/startups/newsbuzz/hackers-mined-a-fortune-from-indian-websites/articleshow/65836088.cms (December 2018).

[4] CISCO. Cisco Umbrella 1 Million. http://s3-us-west-1.amazonaws.com/umbrella-static/top-1m.csv.zip (December 2018).

[5] CLABURN, T. Crypto-jackers enlist google tag manager to smuggle alt-coin miners, Jan 2018. https://www.theregister.co.uk/2017/11/22/cryptojackers_google_tag_manager_coin_hive/ (December 2018).

[6] COIN-HAVE. Coinhave – monero javascript mining. https://coin-have.com/ (December 2018).

[7] COINHIVE. Blog: Discontinuation of coinhive. https://coinhive.com/blog/en/discontinuation-of-coinhive (April 2019).

[8] COINHIVE. Coinhive blog: Authedmine – non-adblocked. https://coinhive.com/blog/en/authedmine (April 2019).

[9] COINHIVE. First week status report, Sep 2017. https://coinhive.com/blog/en/status-report (December 2018).

[10] COINHIVE. Coinhive - monero mining club, Jan 2018. https://coinhive.com/ (December 2018).

[11] CRYPTOLOOT.COM. Cryptoloot - earn more from your traffic. https://crypto-loot.com/ (December 2018).

[12] DOMAINTOOLS.COM. Domain Count Statistics for TLDs. http://research.domaintools.com/statistics/tld-counts/ (January 2019).

[13] ESKANDARI, S., LEOUTSARAKOS, A., MURSCH, T., AND CLARK, J. A first look at browser-based cryptojacking. *2018 IEEE European Symposium on Security and Privacy Workshops, EuroS&P Workshops 2018, London, United Kingdom, April 23-27, 2018* (2018), 58–66.

[14] GITHUB.COM. hoshsadiq/adblock-nocoin-list. https://github.com/hoshsadiq/adblock-nocoin-list (December 2018).

[15] GITHUB.COM. leonidackov901/leonidackov901.github.io. https://github.com/leonidackov901/leonidackov901.github.io (January 2019).

[16] GOOGLE.COM. minerblock. https://chrome.google.com/webstore/detail/minerblock/emikbbbebcdfohonlaifafnoanocnebl?hl=en (January 2019).

[17] GOOGLE.COM. No coin - block miners on the web! https://chrome.google.com/webstore/detail/no-coin-block-miners-on-t/gojamcfopckidlocpkbelmpjcgmbgjcl (January 2019).

[18] HOFFMAN, J. J., LEE, S. C., AND JACOBSON, J. S. 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, May 2015. https://nj.gov/oag/newsreleases15/pr20150526b.html (December 2018).

[19] 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 2018, Toronto, ON, Canada, October 15-19, 2018* (2018), pp. 1701–1713.

[20] JSECOIN. Jsecoin: Digital currency - designed for the web. https://jsecoin.com/ (April 2019).

[21] KEIL, D. Wp monero miner - home. https://www.wp-monero-miner.com/ (December 2018).

[22] 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 2018, Toronto, ON, Canada, October 15-19, 2018* (2018), pp. 1714–1730.

[23] KREBS, B. Krebs on security - who and what is coinhive. https://krebsonsecurity.com/2018/03/who-and-what-is-coinhive/ (December 2018).

[24] LIU, J., ZHAO, Z., CUI, X., WANG, Z., AND LIU, Q. A novel approach for detecting browser-based silent miner. *Third IEEE International Conference on Data Science in Cyberspace, DSC 2018, Guangzhou, China, June 18-21, 2018* (2018), 490–497.

[25] MAJESTIC. Majestic Million CSV now free for all, daily. http://downloads.majestic.com/majestic_million.csv (December 2018).

[26] MCCARTHY, K. Cbs's showtime caught mining crypto-coins in viewers' web browsers, Jan 2018. https://www.theregister.co.uk/2017/09/25/showtime_hit_with_coinmining_script/ (December 2018).

[27] MINERALT. Developer api documentation and reference. https://support.mineralt.io/support/solutions/articles/36000047274-js-miner-usage-and-api-reference (December 2018).

[28] MONERO OCEAN. Monero ocean – faq. https://moneroocean.stream/#/help/faq (May 2019).

[29] MOZILLA FOUNDATION. asm.js - working draft — 18 august 2014. http://asmjs.org/spec/latest/ (November 2018).

[30] MURPHY, M. Youtube shuts down hidden cryptojacking adverts, Jan 2018. https://www.telegraph.co.uk/technology/2018/01/29/youtube-shuts-hidden-crypto-jacking-adverts/ (November 2018).

[31] MURSCH, T. Cryptojacking malware coinhive found on 30,000 websites, Feb 2018. https://badpackets.net/cryptojacking-malware-coinhive-found-on-30000-websites/ (December 2018).

[32] MURSCH, T. Over 100,000 drupal websites vulnerable to drupalgeddon 2 (cve-2018-7600), Jun 2018. https://badpackets.net/over-100000-drupal-websites-vulnerable-to-drupalgeddon-2-cve-2018-7600/ (January 2019).

[33] NAKAMOTO, S. Bitcoin: A peer-to-peer electronic cash system, 2009.

[34] OGONO, U. Monero cryptojacking: Monero cryptocurrency mining malware disrupts government site, Sep 2018. https://smartereum.com/35507/monero-cryptojacking-monero-cryptocurrency-mining-malware-disrupts-government-site-monero-news-today/ (December 2018).

[35] OSBORNE, C. Mikrotik routers enslaved in massive coinhive cryptojacking campaign, Aug 2018. https://www.zdnet.com/article/mikrotik-routers-enslaved-in-massive-coinhive-cryptojacking-campaign/ (December 2018).

[36] PAPADOPOULOS, P., ILIA, P., AND MARKATOS, E. P. Truth in web mining: Measuring the profitability and cost of cryptominers as a web monetization model. *CoRR abs/1806.01994* (2018).

[37] PASTRANA, S., AND SUAREZ-TANGIL, G. A first look at the crypto-mining malware ecosystem: A decade of unrestricted wealth. *CoRR abs/1901.00846* (2019).

[38] PEARSON, J. Starbucks Wi-Fi Hijacked People's Laptops to Mine Cryptocurrency. https://motherboard.vice.com/en_us/article/gyd5xq/starbucks-wi-fi-hijacked-peoples-laptops-to-mine-cryptocurrency-coinhive (February 2019).

[39] RAUCHBERGER, J., SCHRITTWIESER, S., DAM, T., LUH, R., BUHOV, D., PÖTZELSBERGER, G., AND KIM, H. The other side of the coin: A framework for detecting and analyzing web-based cryptocurrency mining campaigns. In *Proceedings of the 13th International Conference on Availability, Reliability and Security, ARES 2018, Hamburg, Germany, August 27-30, 2018* (2018), pp. 18:1–18:10.

[40] RODRIGUEZ, J. D. P., AND POSEGGA, J. RAPID: resource and api-based detection against in-browser miners. *Proceedings of the 34th Annual Computer Security Applications Conference, ACSAC 2018, San Juan, PR, USA, December 03-07, 2018* (2018), 313–326.

[41] RÜTH, J., ZIMMERMANN, T., WOLSING, K., AND HOHLFELD, O. Digging into browser-based crypto mining. In *Proceedings of the Internet Measurement Conference 2018, IMC 2018, Boston, MA, USA, October 31 - November 02, 2018* (2018), pp. 70–76.

[42] SAAD, M., KHORMALI, A., AND MOHAISEN, A. End-to-end analysis of in-browser cryptojacking. *CoRR abs/1809.02152* (2018).

[43] SABERHAGEN, N. V. Cryptonote v 2.0, Oct 2013. https://cryptonote.org/whitepaper.pdf.

[44] SCHEITLE, Q., HOHLFELD, O., GAMBA, J., JELTEN, J., ZIMMERMANN, T., STROWES, S. D., AND VALLINA-RODRIGUEZ, N. A long way to the top: Significance, structure, and stability of internet top lists. *CoRR abs/1805.11506* (2018).

[45] SEGURA, J. A look into drupalgeddon's client-side attacks, Jun 2018. https://blog.malwarebytes.com/threat-analysis/2018/05/look-drupalgeddon-client-side-attacks/ (January 2019).

[46] SLUSHPOOL. Stratum mining protocol. https://slushpool.com/help/topic/stratum-protocol/ (November 2018).

[47] THE MONERO PROJECT. Monero: What is monero (xmr)? https://www.getmonero.org/get-started/what-is-monero/ (December 2018).

[48] THE PIRATE BAY. The pirate bay - miner, Sep 2017. https://thepiratebay.org/blog/242 (December 2018).

[49] WANG, W., FERRELL, B., XU, X., HAMLEN, K. W., AND HAO, S. SEISMIC: secure in-lined script monitors for interrupting cryptojacks. In *Computer Security - 23rd European Symposium on Research in Computer Security, ESORICS 2018, Barcelona, Spain, September 3-7, 2018, Proceedings, Part II* (2018), pp. 122–142.

[50] WEBASSEMBLY.ORG. Webassembly. https://webassembly.org/ (November 2018).

[51] WEBSHRINKER. Webshrinker apis. https://www.webshrinker.com/apis/ (January 2019).

[52] WEBSOCKET.ORG. Html5 websocket - a quantum leap in scalability for the web. http://www.websocket.org/aboutwebsocket.html (November 2018).

[53] WORDFENCE.COM. Wordpress plugin banned for crypto mining, Nov 2017. https://www.wordfence.com/blog/2017/11/wordpress-plugin-banned-crypto-mining/ (January 2019).

[54] XU, J., FAN, J., AMMAR, M., AND MOON, S. B. On the design and performance of prefix-preserving ip traffic trace anonymization. In *ACM SIGCOMM Workshop on Internet Measurement* (2001).

A Search queries for PublicWWW

Table 9: All search queries for the PublicWWW database

Miner	Search term(s)
Coinhive	coinhive.min.js, CoinHive.Anonymous(
JSECoin	load.jsecoin.com
Webmine	webmine.cz
Cryptoloot	/crypta.js, /crlt.js, crlt.anonymous, CryptoLoot.Anonymous
CoinImp	CoinImp.Anonymous, www.hashing.win, hostingcloud.racing
Cryptonoter	minercry.pt/processor.js, cryptonoter
NFWebminer	nfwebminer.com/lib/, NFMiner(
Deepminer	deepMiner
Monerise	monerise_builder, monerise_payment_address(
Coinhave	minescripts.info
Nebula	CoinNebula.Instance
Mineralt	play.gramombird.com/app.js
Munero	munero.me
Minr	cdn.jquery-uim.download, cnt.statistic.date, ad.g-content.bid
Webminerpool	webmr.js
WPMoneroMiner	wp-monero-miner.js
Nerohut	nhm.min.js, nerohut.com/srv
Adless	adless.js
Monero-mining	Perfektstart(
Miscellaneous	function echostat(){var, function printju, pocketgolf.host/start.php async, startMining(, jquory.js

B Added miner applications and their keywords for the campaign crawl

Table 10: The added miner applications and their keywords

Miner	Keywords
Nebula	CoinNebula.Instance
WP Monero miner	wp_js_options | wp-monero-miner
Nerohut	nhm.min.js | NHpwd | nhsrv.cf/srv/serve.php?key=
Webminerpool	webmr.js | startMining(
Minero	minero.cc
Adless	adless.js | adless.io
Monero-mining	PerfektStart | perfekt.js
ProjectPoi	ProjectPoi\b | projectpoi.min.js
Papoto	papoto

C Added miner applications in the Internet scale crawl

Table 11: The added miner applications and their keywords in the latest version of the crawler

Miner	Keywords
SMMCH	simple-monero-miner-coin-hive smmch-public | smmch-mine.js
Webminepool	webminepool.com/lib/base.js
Unknown miner	proofly.date | flightsy.date | gettate.trade alflying.date | flightzy.date | joytate.date zymerget.faith | nflying.win | flightzy.bid flightsy.win | zymerget.bid | nflying.bid baseballnow.press | flightsy.bid
Omine	omine.org
Browsermine	browsermine.com.cc | bmcm.pw | bmnr.pw lm-sdfhfad.ml | new BMCM | asdvhsrtsb.ml

D Human Subjects and Ethical Considerations

For the analysis of cryptojacking usage in the wild, this paper uses NetFlow statistics from a Tier 1 network operator. This data access was cleared by the institutional review board. The research team did not obtain direct access to the NetFlow data containing source and destination IP addresses as personally identifiable information, but instead provided a list of IP addresses of cryptomining proxies and mining pools to the data owner, based on which the corresponding flow records were provided with the connection's source IP protected by a salted hash.