Unearthing malicious campaigns and actors from the blockchain DNS ecosystem

Fran Casino[a,b], Nikolaos Lykousas[a], Vasilios Katos[c], Constantinos Patsakis[a,b]

[a]*Department of Informatics, University of Piraeus, 80 Karaoli & Dimitriou str., 18534 Piraeus, Greece*
[b]*Information Management Systems Institute, Athena Research Center, Artemidos 6, Marousi 15125, Greece*
[c]*Bournemouth University, Poole House P323, Talbot Campus, Fern Barrow, Poole, Dorset, BH12 5BB, UK*

Abstract

Blockchain DNS has emerged as an alternative solution to traditional DNS to address many of its inherent drawbacks. In this regard, a blockchain DNS approach is decentralised, resilient, provides high availability, and prevents censorship. Unfortunately, despite these desirable features, the major blockchain DNS solutions to date, Namecoin and Emercoin have been repeatedly reported for malicious abuse, ranging from malware distribution to phishing. In this work, we perform a longitudinal analysis of both these chains trying to identify and quantify the penetration of malicious actors in their ecosystems. To this end, we apply a haircut blacklisting policy and the intelligence collected from various engines to perform a taint analysis on the metadata existing in these blockchains, aiming to identify malicious acts through the merge of identifying information. Our analysis provides an automated validation methodology that supports the various reports about the wide-scale abuse of these solutions showing that malicious actors have already obtained an alarming and extensive share of these platforms.

Keywords: Blockchain, Blockchain Forensics, Cybercrime, DNS, Malware, Decentralised DNS

1. Introduction

With the continuous digitisation of procedures, services, and products, crime has been shifting towards the same direction. Despite the continuous evolution of artificial intelligence techniques such as machine learning, pattern recognition and natural language processing, which are capable of ingesting terabytes of unstructured data to enhance response times, and expand the capacities of security

Email addresses: `francasino@unipi.gr` (Fran Casino), `nlykousas@unipi.gr` (Nikolaos Lykousas), `vkatos@bournemouth.ac.uk` (Vasilios Katos), `kpatsak@unipi.gr` (Constantinos Patsakis)

operations, attackers tend to be always a step ahead. The latter is directly related to the appearance of novel technologies, industrialisation processes, the difficulty to collect data from diverse sources in orchestrated campaigns and their timely detection, and the lack of proactive security mechanisms. As a result, cybercrime is predicted to be the third-largest economy in 2021[1].

Meanwhile, there have been systematic efforts to address the security and privacy issues of the Domain Name System (DNS). The DNS is one of the oldest yet critical Internet application-level protocols. In this regard, recommendations and approaches for security improvements such as DNSSEC, DNSCurve, and DNS over TLS/HTTPS are hindered by the lack of adoption [2], which leave DNS exposed to several threats, including man-in-the-middle attacks, passive eavesdropping and data injection. Moreover, the hierarchical design of DNS makes it prone to particular types of attacks such as poisoning, as well as amplification type of denial of service attacks [3]. For instance, due to the lack of authentication in the traditional DNS protocol, a DNS server cannot authenticate whether a response originates from a valid DNS resolver, which is ranked higher in the DNS hierarchy. Therefore, an attacker may query a DNS server for a known website XYZ and then send a spoofed response which falsely claims that the IP of XYZ is an attacker controlled host. However, for efficiency, DNS servers store the responses from DNS resolvers in their cache. Thus, the spoofed response will be cached in the DNS server. As a result, all users who later ask for the IP of XYZ will be redirected to host controlled by the attacker. Furthermore, freedom of speech is hard to accomplish given the actual design of DNS, since, e.g. authoritative regimes can manipulate them to block traffic and censor everything that may question them.

Recently, with the exploitation of decentralised, immutable data structures such as blockchain, several industries have found a way to promote their services and enhance their features, including security, privacy, traceability, and verifiability [4, 5]. Nevertheless, the inherent immutability of such systems paired with design flaws prevent illegal and undesired content from being modified or taken down [6, 7]. In this context, novel decentralised applications such as decentralised DNS systems are not an exception [8, 9]. Therefore, despite the potential of BDNS systems to disrupt traditional DNS models, their inherent design flaws can be used to leverage resilient malware campaigns.

Motivation and main contributions: The threat landscape has changed considerably since the introduction of DNS, urging the community to seek alternatives for this service. These alternatives are served in two main flavours: 1) Security improvements of the existing DNS using approaches like DNS over HTTPS [10] and DNS over TLS [11], and 2) Decentralisation of DNS, with blockchain as the enabling technology. In the latter case, several approaches are already functional, with Namecoin and Emercoin being the most mature and used ones. In addition, other approaches seem to perpetuate the blockchain

DNS trend, such as Handshake[1], while some registered patents by, e.g. Alibaba[2] and with upcoming projects (e.g., both EXIP [12] and Butterfly [13] projects launched and ICO in 2021), which aim to extend the foundational properties of BDNS, highlight the importance of a proper design of such systems. In addition, novel browsers like Brave [14] are rapidly gaining attention due to their privacy properties, as well as other potential benefits for the users. Brave already adopts several similar mechanisms like Unstoppable domains and the Ethereum name service (ENS).

Despite the research leveraged by the community towards more secure and resilient DNS systems, adversaries are expected to opportunistically take advantage of such changes by exploiting both the technology in its early stages, as well as the lack of knowledge and experience of the end-users and system administrators. For instance, well-known malicious campaigns are still exploiting such systems. For example, BazarLoader struck again in April 2021, showcasing new specific attack patterns similar to these of Trickbot [15], as also claimed in the past [16]. It is therefore imperative to raise awareness on the emerging security threats

This work extends the initial findings of [8, 9] and provides a automated and comprehensive approach towards discovering illegal activities related to blockchain DNS services to the one described in [17]. In the latter, the authors captured malicious traffic originating from blockchain DNS resolved sources and conducted a binary classification approach between benign traffic (traditional) and malicious blockchain DNS traffic. Nevertheless, the size of their dataset and the fact of differentiating between disparate types of traffic (i.e. traditional and blockchain-based) requires further research to provide more extensive and statistically sound outcomes.

In this work, we analyse the corpus of domains registered in Namecoin and Emercoin and their registered IPs. Moreover, we provide evidence of the connection between a subset of such domains and illegal activities, as reported and corroborated by several individual sources. To this end, we adapt the blacklisting poison and haircut policy of Möser et al. [18] to a blockchain DNS context. This approach enables an investigator to identify strong connections among IPs and wallets that are validated by existing attack patterns, e.g. BazarLoader [19]. Moreover, we identify traces of active attacks and campaigns and several correlations on the metadata used in both chains, namely wallets, IPs, domains and emails. In addition, by analysing the malicious IPs used by several subsets of wallets and domains, we identify potentially malicious IPs that have not been reported yet. For each investigation phase we provide a detailed description of the procedures, and a comprehensive representation of the outcomes, which prove that the existing blockchain DNS systems are far from delivering the evangelised features. To the best of our knowledge, this is the first piece of

[1] https://www.coindesk.com/handshake-goes-live-with-an-uncensorable-internet-browser

[2] `https://domainnamewire.com/2019/08/15/alibaba-files-blockchain-domain-name-patent-application/`

research that provides detailed and documented proof of the malicious activities carried out in both Namecoin and Emercoin by automating the analysis of internal blockchain data as well as correlated data from external intelligence. Moreover, we provide several automated mechanisms to leverage proactive measures and detect cybercriminal campaigns orchestrated in the core of blockchain DNS systems. Finally, our methodology illustrates how blockchain forensics can be performed beyond the cryptocurrency ecosystem, where the actual evidence are not limited to the data existing in the chain itself.

The rest of the article is organised as follows. In Section 2, we provide a general background on blockchain DNS and explore the related work. In Section 3, we describe the methodology adopted in terms of data collection and analysis, and in Section 4 we provide a thorough analysis of the registered domains in Namecoin and Emercoin, as well as the identification of the illegal activities leveraged by such domains. Finally, in Section 5, we discuss the findings of our experiments and conclude the article by providing some threads for future research.

2. Related Work

As studied in the current literature [20, 21, 22, 23, 24, 25, 9, 26, 27], the main features that decentralised systems can potentially provide are availability, robustness, censorship resistance, as well as other managerial improvements. Table 1 summarises the main characteristics and features of blockchain DNS systems according to the literature.

Property	**Description**
Availability	The availability of the system depends on multiple peers and not on a single entity.
Automated Management	Auctions to register domain names, fast and transparent ownership control
Censorship-resistance	Domain name resolution services and information are not subject to borders or bans
Decentralisation	The network is completely distributed with no central entities
Namespace Freedom	Registration of new SLDs and TLDs
Robustness	Resilient to attacks that affect centralised DNS systems such as MiM, spoofing, cache poisoning, cracking.
Trust	Through verifiable and robust consensus mechanisms
Unlimited Resources	A high number of simultaneous users sharing their assets.

Table 1: Main characteristics of blockchain DNSs.

The early strategies adopted to create decentralised DNS systems focused on the development of specific TLDs such as in the case of the Dot-P2P project (with the .p2p TLD) [28]. However, the inherent performance bottlenecks contributed to adoption delays and diminished the functionality of such systems. Only recently, and due to the progressive adoption of blockchain-based distributed DNS systems [29], the idea of functional and real-world distributed DNS systems is showing clear signs of a comeback.

There exists a set of functional approaches to blockchain-based DNS according to the scientific literature. Hari et al. [30] provided a thorough discussion about the limitations of traditional practices and the benefits of using blockchain for the development of a DNS infrastructure. In [31], Benshoof et al. proposed D^3NS, which integrates a distributed hash table and domain name ownership implementation based on the Bitcoin blockchain. One of their aims is to replace the top-level DNS and certificate authorities, offering increased scalability, security and robustness. Gourley and Tewari [32] proposed the use of blockchain to improve the main drawbacks of DNSSEC in the certificate validation procedure, creating an enhanced DNS security extension. With a similar aim, Guan et al. [33] presented AuthLedger, blockchain-based system that provides efficient and secure domain name authentication. Liu et al. [34] proposed a blockchain-based decentralisation DNS resolution method with distributed data storage to mitigate single points of failure and domain name resolution tampering. BlockZone, proposed by Wang et al. [35], uses a replicated network of nodes to offer efficient name resolution supported by improved Practical Byzantine Fault Tolerance (PBFT) consensus mechanism. Yu et al. [36] proposed the use of a consortium blockchain to establish a DNS cache resources trusted sharing model, which improves the credibility of DNS resolution results by establishing a complete chain of trust.

In the IoT communications domain, some authors have developed specific blockchain-based solutions to enhance domain name resolution and management. For instance, Duan et al. [37] presented DNSLedger, a decentralised, hierarchical multi-chain structure to provide domain name resolution services. BlockONS, proposed by Yoon et al. [38], described a robust and scalable object name service appropriate for an IoT ecosystem with the aim to overcome classical problems related to DNS resolution, namely DNS cache poisoning, spoofing, and local DNS cracking. ConsortiumDNS, presented by Wang et al. [39] is a three-layer architecture composed by a consortium blockchain, a consensus mechanism and external storage. The authors claim that their approach is more efficient compared to other well-known approaches such as Namecoin or Blockstack. Finally, a set of patented designs of Blockchain-based DNS systems can be found in [40, 41].

The first system to reach a certain level of maturity was Namecoin[3], which is a cryptocurrency based on Bitcoin, with additional features such as decentralised name system management, mainly for the `.bit` domain. Moreover, it was the

[3] `https://www.namecoin.org/`

first project to provide security, decentralisation and human-meaningfulness, as required to address Zooko's triangle[27]. Nevertheless, due to the lack of support and adoption, Namecoin's main drawback is its insufficient computing power, which makes it more vulnerable to the 51% attack than other similar systems. Blockstack [42] is a blockchain-based naming and storage system that separates control and data planes, enabling seamless integration with the underlying blockchain. EmerDNS[4], more commonly known as Emercoin, is a blockchain DNS system which supports a wide range of DNS records. EmerDNS operates under the "DNS" service abbreviation in the Emercoin NVS. Handshake [5] is one of the most widely supported technologies, which aims to offer an alternative to existing certificate authorities. Therefore, Handshake aims to replace the root zone file and the DNS name resolution and registration services worldwide.

In addition to the above systems, there are two approaches that are based on the Ethereum blockchain, the Ethereum name service[6] (ENS), and Nebulis[7]. The former uses smart contracts to manage the `.eth` registrar through bids. Moreover, ENS recently added the support for `.onion` addresses. The latter is a globally distributed directory that relies on the Ethereum ecosystem and smart contracts to store, update and resolve domain records. Moreover, Nebulis uses decentralised storage technologies such as IPFS as a replacement for HTTP. Table 2 summarises the main features of the discussed DNS approaches.

Finally, OpenNIC[8] is a unique case, since it is a hybrid approach in which a group of peers manages namespace registration, yet the name resolving task is fully decentralised. OpenNIC provides DNS namespace and resolution for an extensive set of domains, including those managed by EmerDNS, and New Nations[9]. In addition, OpenNIC resolvers have recently added access to domains administered by ICANN. Notably, OpenNIC has dropped the support for `.bit` domains due to malware abuse [10]. As stated in the corresponding voting:

> "*Over the past year .bit domains have started being used as malware hubs due to their anonymous nature. Since there is no way to contact the owner of those domains, it creates a backscatter effect, and a number of people running public T2 servers have seen domains blacklisted, emails blocked, and shutdown notices from their providers.*"

Currently, several malicious campaigns are exploiting the features of the blockchain DNS ecosystem. Setting aside the massive cybersquatting attacks [9]

[4] `https://emercoin.com/en/documentation/blockchain-services/emerdns/emerdns-introduction`

[5] `https://handshake.org/`

[6] `https://ens.domains/`

[7] `https://www.nebulis.io/`

[8] `https://www.opennic.org/`

[9] http://www.new-nations.net/

[10] `https://wiki.opennic.org/votings/drop_namecoin`

and hosting of malicious marketplaces, e.g. Joker's Stash [43, 44], the blockchain DNS approach has been exploited by many malware families as it provides *bulletproof hosting* [45]. The latter cannot be considered a recent development as reports about the abuse of `.bit` domains date back to 2013 [46]. From that point onward a number of regular reports emerged on specific malware families exploiting the blockchain DNS ecosystem. For instance, Fbot botnet used domains resolved by Emercoin to communicate with its *command and control* (C2) servers [47] and the same approach was used by Cerber [48]. In general, as reported by FireEye [49], blockchain DNS domain have been used for hosting C2 servers of many malware families, including but not limited to Necurs, AZORult, Emotet [50], Terdot, Gandcrab [51], SmokeLoader [52], and very recently Trickbot [19].

Table 2 summarises the main features of the most relevant Blockchain-DNS systems.

Method	**Pedigree Platform**	**Registrar & Management**	**Resolution**	**TLD Examples**
ICANN	Network of servers and resolvers	Centralised		`.com .net .org`
Namecoin	Bitcoin and Peercoin	Decentralised		`.bit`
Emercoin	Bitcoin	Decentralised		`.coin .bazar .emc`
ENS	Ethereum	Decentralised		`.eth .onion`
Handshake	Bitcoin	Decentralised		unrestricted
Blockstack	Blockchain agnostic	Decentralised		`.id .podcast .helloworld`
OpenNIC	Decentralised servers	Hybrid		`.bbs .pirate .libre`

Table 2: Main characteristics of the most relevant DNS systems. Although Blockstack is blockchain agnostic, it is mainly used with Bitcoin blockchain.

Internet users can reach the TLDs offered by Namecoin, OpenNIC, New Nations, and EmerDNS (e.g. `.coin`, `.emc`, `.lib` and `.bazar`) through various browser extensions such as `peername`, `blockchain-DNS` and `friGate` [53]. The domain name resolution procedure is outlined in Figure 1.

Finally, despite the theoretical and desired features previously described, blockchain DNS systems have several drawbacks, which can be exploited by malicious actors [54, 9, 55]. Patsakis et al. [9] explored the main blockchain DNS systems and identified a set of challenges and threats related to their underlying registrar mechanisms, malware and phishing campaigns, and the immutability of data residing in such systems. Similarly, Xia et al. [56] performed a qualitative analysis of the Ethereum Name Service and discussed their challenges. Recently, Huang et al. [17] explored the traffic generated by sites resolved by blockchain DNS systems and analysed its patterns. Despite the fact that their dataset contains few benign samples, their outcomes showed that they could differentiate between traditional domains and blockchain DNS domains that were known to leverage malicious activities, according to VirusTotal.

Following an analysis of the literature, the main drawbacks identified by researchers to detect malicious activities in blockchain DNS systems are (i) the lack of automated tools to pair the activities performed in the blockchain with

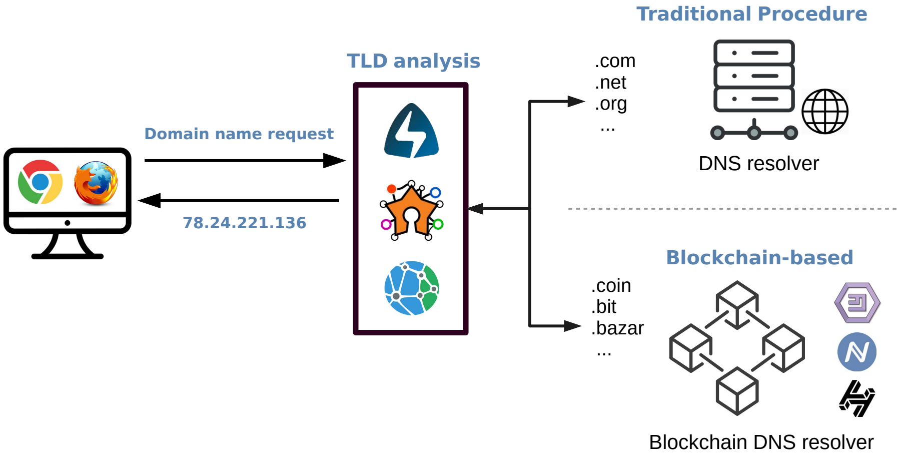


Figure 1: Workflow of the domain name resolution procedure. The extension analyses the TLD of the requested domain and directs the query to the corresponding DNS system.

external intelligence tools, (ii) the difficulty to extract interoperable metrics (e.g., behavioural indicators) to identify malicious behaviours, and (iii) the unstructured nature of data, which prevents the application of policies extendable to other frameworks. To the best of our knowledge, our work is the first to propose a fully automated pipeline leveraging a structured data analysis and feature collection, which is used to correlate blockchain data with external intelligence sources and apply proactive policies to effectively detect malicious behaviours as well as cybercrime campaigns.

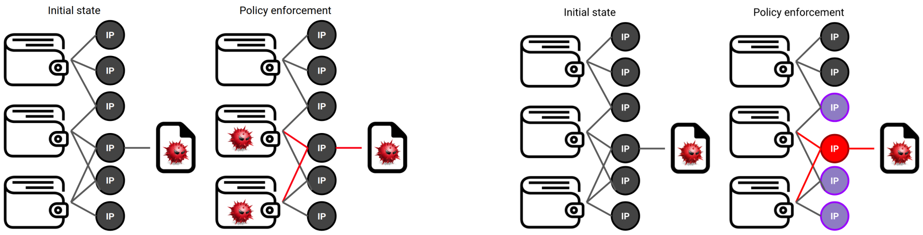
3. Methodology

As already discussed, in a blockchain DNS system, one registers a domain by paying through the corresponding cryptocurrency, e.g. Namecoin, Emercoin, etc. Setting aside the monetary transactions which may hinder money-laundering acts, the maliciousness stems from the content that such a domain has. Currently, we are well aware that blockchain DNS systems have been exploited by malicious actors for several malware campaigns or black marketplaces, as discussed in Section 2. One may ponder about the extent of this exploitation, as it is infeasible to collect all the content, and even if it were possible, it would be impossible to collect the content that existed and was flagged malicious.

To alleviate this challenge and create a ground truth, we base our analysis on the domains and IPs that are registered in these blockchains. To this end, we initially perform a dump of these blockchains to collect all the domain names and the IPs that have been used by them. Contrary to traditional DNS systems, in blockchain-based DNS all the history of a domain, including the IPs that were used to provide the content is recorded and publicly accessible.

Additionally, we aim to establish a baseline approach to perform blacklisting and use it to measure the number of malicious wallets and domains. A straightforward process is to use an intelligence engine to query these domains. However, taking into consideration only the domains is not very effective as most of these domains are not indexed, and only a few intelligence engines collect data about them. Moreover, it is highly possible that the logs that they have may not refer to the domains per se, but their IP addresses. This can be attributed to the fact that the DNS query is performed to a non-standard TLD and the engine drops it. Nonetheless, the connection to the IP is recorded. Therefore, one has to consider whether the IP has been used for other malicious activities, e.g. spamming, phishing etc.

We argue that the blacklisting policies of Möser et al. [18] that were applied in Bitcoin to trace money laundering can be adopted in the blockchain DNS chains to identify malicious activity. To this end, we adapt the poison and haircut policies as follows. Let us assume that wallet W_1 has registered a domain D_1 which is mapped to IP_1. If IP_1 is flagged as malicious, then the wallet is flagged as malicious. Similarly, if wallet W_2 has registered a domain D_2 which is also mapped to IP_1, then wallet W_2 is also flagged as malicious. In essence, a malicious IP "poisons" all the wallets that are attached to it. Nonetheless, once we have a malicious IP in a wallet, it taints the rest of the IPs of the wallet. Using the haircut policy of Möser et al. we consider the rest of the IPs as *suscpicious*. Therefore, poisoning is applied to domains and wallets, while haircut is applied to IPs. The two policies are illustrated in Figure 2.



(a) Poison blacklisting policy.

(b) Haircut blacklisting policy.

Figure 2: Wallet and IP blacklisting with the poison (a) and haircut (b) policies.

Based on the above, we first need to look for the domains and then extract intelligence about the IPs that are used. Using the above, we attempt to identify any emerging patterns and whether the tainting approach provides any insight regarding upcoming threats.

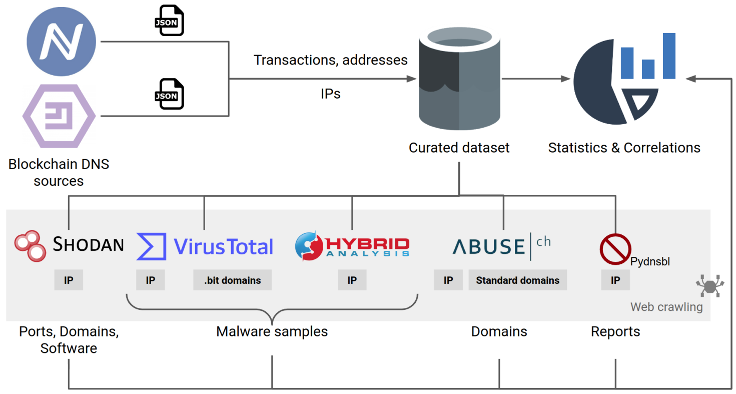


Figure 3: Outline of the methodology for analysing blockchain DNS data.

4. Experimental Setup

To investigate malicious activities related to the use of blockchain DNS platforms, we analysed the contents of both Namecoin[11] and the Emercoin[12] blockchains. Namecoin was the first widely used Blockchain DNS, becoming a reference point for more recent approaches such as Emercoin and Blockstack. This blockchain manages the registrar of the `.bit` TLD through a straightforward procedure, in which a registrant specifies the SLD that they wish to register (which is subsequently appended with the `.bit` TLD), as well as the resolving IP and other secondary parameters. The Emercoin blockchain is one of the most well-known services for domain registration. Surprisingly enough, although the naming requirements of Emercoin specify that only lowercase alphanumeric ASCII characters are allowed, the chain contains case sensitive domains not only for the advertised TLDs but for traditional TLDs like `.com`. In the following sections, we describe the details of each phase of our approach, which are detailed in Figure 3.

4.1. Data collection and dataset structure

For the purposes of this research, we downloaded all the data from the two most widely used chains supporting blockchain DNS, which at the time of writing are Emercoin and Namecoin, in the form of JSON files. From these files, we extracted a subset of relevant information, namely domain names, IPs and emails (by using the *value* field), and the wallets associated to each domain, to create a curated dataset. Based on this, our dataset consists of a set of unique 5985 IP addresses. Note that the set of IP addresses consists of the public IPs as there were many occurrences of private IPs. Most likely, the private IP

[11] https://www.namecoin.org/
[12] `https://emercoin.com/`

addresses are acting as placeholders for future record updates. We also noted invalid IPs or containing typos, for instance, one of the four integers of an IPv4 address contained a number greater than 255. These IP addresses were pruned as they provided no tangible value from an investigation perspective. Therefore, we ended up with 5130 public IPs being used in Namecoin, 919 in Emercoin, and 55 IPs are in both chains.

In addition, the dataset contains 2469 Emercoin wallets and 61357 Namecoin wallets, which are related to these IPs in distinct ways. Finally, the number of domains related to these IPs are 4452 in the case of Emercoin, and 27403 in the case of Namecoin. Nonetheless, not all of them are valid domain names. There are multiple domains which do not conform to the DNS format, e.g. they contain non allowed characters, have registered the same domain with combinations of upper and lower case characters etc. As a result, the resulting numbers of domains are 2675 for Emercoin and 27261 for Namecoin.

The first step in our intelligence collection was to query the registered domains in the available engines. Due to the fact that these TLDs are not widely available, only a few engines provide actual information. In our research, we used VirusTotal, which at the time of writing supports only queries for `.bit` domains. From the 27261 domains that were queried, only 661 were recorded in VirusTotal, 195 of which were reported malicious. Notably, these malicious domains were associated with 576 unique public IP addresses, implying that almost all of them have been updated several times. The fluxing rate of these IP addresses will be discussed in Section 4.2.2. Based on our poison blacklisting policy, since these domains are reported as malicious, the associated wallets that have registered them, and the IPs that have hosted them are poisoned, hence flagged as malicious.

Next, we submitted all the extracted IPs from Namecoin and Emercoin to VirusTotal, Hybrid Analysis, and Shodan, and collected the information that each platform has about them. We queried the 5985 unique IPs to which domains have been mapped in *VirusTotal* and *Hybrid Analysis* to determine how many of them are linked with malware samples that they have analysed. Notably, 1550 (25.9% of the total) IPs are reported malicious in the two platforms as they are correlated with 32340 unique samples. Moreover, using intelligence from the different sources provided by Abuse[13], we identified some more IPs being malicious, reaching to 26.18% of the total. Merging the latter with the reports of VirusTotal for the `.bit` domains we have 1926 malicious IP addresses. Finally, we queried VirusTotal for the rest of IPs for other malicious activity, e.g. spamming, phishing etc. Of the remaining 4062, 131 were flagged as malicious, raising the total to 2057 IPs. Practically, more than a third (34.32%) of the IPs to which domains backed by blockchain DNS are redirecting are known to be malicious.

Notably, these IPs are linked with several malware families including, but not limited to, Emotet, AZORULT, Feodo, Cerberus, GermanWiper, and Gand-

[13] `https://abuse.ch/`

Crab. A more comprehensive list is presented in Table 3.

Type	Families
Banking malware	Ursnif, Chthonic, Dridex, Panda, BankBot, ClipBanker, Cerberus, Feodo, Geodo, heodo, Gozi, Vawtrak, Qbot
Ransomware	Buran, GlobeImposter, GermanWiper, GandCrab, Hermes, Phobos, Paradise, Troldesh, Sigma, maze, locky, zerber
Loader	hancitor
Trojan	Bifrost, emotet, DanaBot, PsiXBot
Stealer	AZORULT, Valak
Miners	xmrig, minergate, acruxminer
Botnet	Gafgyt, Mirai, Ramnit
RAT	agent tesla, quasar, ghøst, imminent monitor rat

Table 3: Identified malware distributed by IPs where Emercoin and Namecoin map their domains.

Moreover, we used Pydnsbl[14], an aggregator of blacklists of IPs to determine how many of the IPs have been blacklisted. In total, 1629 of the IPs in our dataset are blacklisted. Purging the duplicate reports of the IPs, the malicious reported IPs are 3039, representing the 50.78% of the total.

Next, we correlated these IP addresses with information from Shodan. While only 2493 of the IP addresses had been monitored and indexed by this tool, we nevertheless can extract valuable intelligence. In Tables 4a and 4b we report the ten most common ports these devices are using and the ten most common identified products by Shodan, respectively. The results indicate that most of the servers are providing web hosting, file sharing, DNS, and mail services, with a preference to Linux-powered servers, implied by the use of SSH.

Port	Count	Common service
80	1690	Web server
443	1411	Web server over SSL/TLS
22	1068	SSH
53	888	DNS
21	386	FTP
25	381	SMTP
993	380	IMAP over TLS/SSL
587	342	SMPT
143	334	IMAP
995	320	POP3 over SSL/TLS

(a) Used ports

Software	Installations
OpenSSH	1123
Apache httpd	729
nginx	681
Exim smtpd	276
MySQL	200
Postfix smtpd	178
Pure-FTPd	141
MS IIS httpd	54
ProFTPD	53
Microsoft HTTPAPI httpd	28

(b) Identified software

Table 4: Statistics from Shodan

[14] https://github.com/dmippolitov/pydnsbl/

4.2. Blockchain DNS analysis and correlation

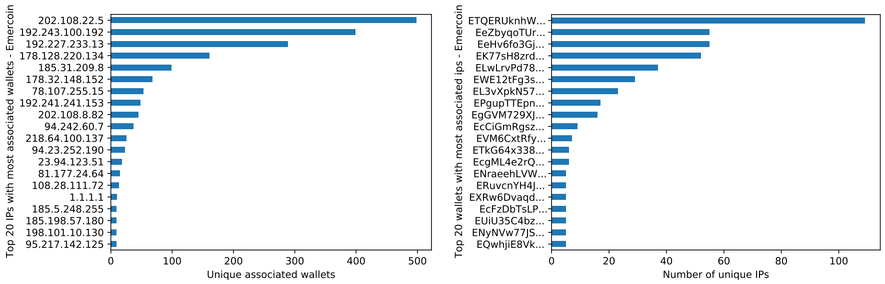
In what follows, we provide a detailed analysis of both Emercoin and Namecoin blockchains. First, we provide an exploratory analysis to highlight the most active IPs and wallets of each system and their ties with malicious activities, as reported by external intelligence sources. Second, we provide a geographical coverage of the IPs of each system. Next, we focus on the potential threats of such systems and apply our blacklisting policy, namely a hop-based approach, to analyse the links between IPs, wallets, domains, and e-mails and categorise their threat level. Finally, we analyse the user's behaviour according to some features to discover patterns that could indicate potential harm, and provide a statistical analysis by correlating them with maliciously reported IPs.

4.2.1. Emercoin

In the case of Emercoin, we created several data structures to establish associations between wallets, IP addresses and domains. First, we collected some statistics regarding the IP addresses found in Emercoin, and how different wallets used them to update the *value* field of one or several domain names. In this regard, Figure 4a provides an overview of the top 20 Emercoin IPs in terms of the number of wallets using them. As it can be observed, a small subset of IPs are associated with more than 100 wallets, yet the vast majority of IPs have only one wallet associated with them, as it can be understood by observing the decreasing pace of the values. For instance, looking at the top five, the most used IP (`202.108.22.5`) has been reported as malicious. In the case of the runner up `192.243.100.192`, although it has not been reported as malicious, it directs to a "boutique" [15] for selling Emercoin domains. The IP `192.227.233.13` is found in many expired domains and was reported as malicious, yet it is not resolving to any site at the time of writing. The IP address `178.128.220.134` is resolving to emerAPI, an Emercoin related software, which includes links to the official site, yet there is no proof of its authenticity. Finally, `185.31.209.8` is an IP announced in several Eastern Europe sites [57] to be used when registering Emercoin domains. In the latter case, several users have used it as a default option. It is worth noting that, although there are only two IPs reported as malicious in this top five, our hop-based association approach, later described in this section, flagged IPs `192.243.100.192` and `185.31.209.8` as suspicious. The latter means that, a) the intelligence available for these sites is insufficient, b) that such IPs are not being used with malicious intentions yet, or c) that malicious users, like benign ones, initially used them when setting up their wallets or d) as a means to temporarily hide their activity and redirect incoming traffic.

Next, we computed the same statistics this time considering each wallet. Figure 4b shows the amount of IPs used by the top 20 Emercoin wallets in their registered domain(s). We can observe that several wallets contain more than

[15] https://www.ecwid.com/store/cantdoevil/Existing-Invincible-EmerDNS-Domains-Contact-p155967426



(a) Top 20 most used IPs in Emercoin and the number of wallets using them.

(b) Top 20 Emercoin wallets and the corresponding number of IPs found in their domains.

Figure 4: Statistics about the most used IPs and biggest wallets of Emercoin.

50 IPs related to them. In this regard, a clear example of the extent to which Emercoin is being used for malicious purposes is given by observing, e.g. the top three wallets, since these are associated with several malicious IPs. Moreover, the wallet `ETQERUknhW2A5cBmfHN4VBqL7VGiFnKQRh` has been related with the DGA of BazarLoader [19] (also known as BazarBackdoor).

In addition, we depicted in Figure 5 the geographical coverage of the Emercoin IPs, and we compare it with the reported malicious activities collected in Section 4.1. As identified in the maps, there is a direct correlation between the number of hosts and the malicious IPs reported. It is worth to mention that, in proportion to the amount of hosted IPs, there are less malicious IPs located in Russia and China than in other areas such as North America and Australia, according to the intelligence reports.

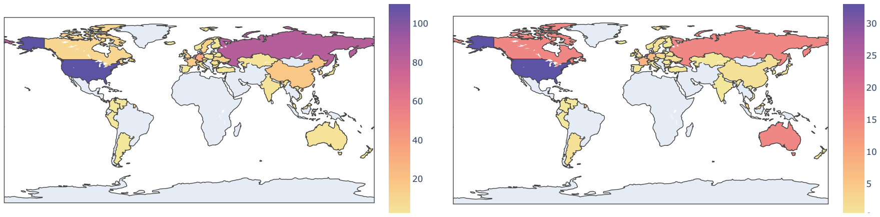


Figure 5: Geographical coverage heatmap of IPs mapped in Emercoin (left) and the corresponding malicious reports (right).

The next phase of the analysis focused on the identification of possible relationships between the different objects existing in these blockchain systems. More precisely, we analysed the correlations between wallets, as reported in the previous experiments, the set of *"apparently"* benign IPs, and the domains used in Emercoin. For this purpose, following the methodology described in Section 3, we developed a hop-based association approach, as described in Algorithm 1. More concretely, if a wallet or a domain contains a malicious IP, we tag the rest of the IPs associated with such wallet or domain as suspicious. Moreover, we

Algorithm 1 Hop-based Association

```
function ComputeSuspiciousIPs( Dict $ip\_to\_wallet$, Dict $wallet\_to\_mail$, Dict $ip\_to\_domain$, List $malicious\_ips$)
    Dict $status\_ips$ = { };
    while ($ip$ in $malicious\_ips$) do
        $status\_ips[ip] = malicious$    ▷ Store $\{key, value\}$ pair.
        $wallet\_list$ = GetWallets ($ip\_to\_wallet$, $ip$)    ▷ Wallets associated with malicious $IP$.
        $domain\_list$ = GetDomains ($ip\_to\_domain$, $ip$)    ▷ Domains associated with malicious $IP$.
        $associated\_ips$ = GetIPs ($wallet\_to\_mail$, $wallet\_list$, $domain\_list$)    ▷ Get IPs of associated wallets and domains
        $status\_ips$ = UpdateDict ($associated\_ips$)    ▷ Update benign IPs with suspicious value
    end while
return status_ips    ▷ Dict with classified IPs
end function
```

use additional information from the *value* field of the curated dataset to find further relationships between such domains and wallets (e.g. wallets using the same email). In this case, we add the IP addresses of the additional wallets to the suspicious list. Following our methodology, we assume that if a wallet has used an IP reported in a malicious campaign, the rest of the associated IPs can potentially be used for similar purposes. Note that a suspicious state can only be updated by a malicious one if a specific IP is found to be malicious according to our ground truth, and that suspicious IPs do not spread their status further.

Concerning the detailed procedures and computational cost of our hop-based approach, the first step is to collect a snapshot of the whole blockchain and parse it into a structured JSON file, which is updated at regular intervals Since this activity is performed offline, we consider this cost negligible. Next, the hop-based approach is applied to both Namecoin and Emercoin data in the order of seconds, even without parallelisation. More concretely, the cost of exploring all the IPs of a given blockchain system and, in the case they are reported as malicious, marking as suspicious the rest of IPs of the wallets containing it, is upper bounded by $O(n^2)/2$ in the case of a fully connected undirected graph. Given n nodes, the number of edges in a fully connected undirected graph is $n(n-1)/2$. As previously seen in Section 4.2, the connectivity of both Namecoin and Emercoin is far from a fully connected undirected graph, and thus the cost in such cases is much lower than $O(n^2)/2$. Moreover, note that the computational cost is also tied to the amount of dangerous IPs of the network. In other words, we only explore the wallets associated with an IP if the latter is marked as dangerous. Finally, the cost of identifying whether an IP is malicious is linear and is proportional to the time it takes to query a threat intelligence engine like VirusTotal that we used in this work.

In the case of Emercoin, our hop-based association found 280 new potentially malicious IPs, in addition to the 502 malicious IPs confirmed by the intelligence collected. Therefore, by revising our initial statistics, 74 IP addresses were found to be benign (only 8% of the IPs did not present any connection with malicious activities).

Further analysis was conducted on the intelligence collected in Section 4.1. In this regard, we use the list of IP addresses and the classification (i.e. benign, malicious and suspicious) provided by the hop-based approach. Thus, we deploy a graph-based visualisation of Emercoin (see Figure 6a), in which nodes represent IPs, and the edge connecting two IPs represents a commonly shared interrelation in the form of, e.g. a wallet, an email, a domain or a combination of them. In the case of benign IPs, we can observe that they are mostly isolated (cf Figure 6a), since they have a very small representation in Figure 6b. In the case of malicious and suspicious clusters, we can clearly identify their connections and all the associations, showcasing the relevance of the hop-based procedure to find new, potentially malicious groups of IPs. The average clustering coefficient of the network represented in Figure 6a is 0.701 and in the case of Figure 6b (discarding the isolated nodes) is 0.831. These numbers denote the high degree of connectivity between the nodes when they belong to a cluster, exhibiting highly interconnected communities. Figure 7a shows the Complementary Cumulative Distribution Function (CCDF) of Emercoin. It can be clearly observed, by merging the data represented in Figure 7a with the visual information of Figure 6a, that there are specific peaks corresponding to high degree clusters. The number of clusters appears to be similar regardless of their degree, for clusters with more than 10^2 elements. The latter denotes specific malicious behaviours (note that high degree clusters exist only in a malicious context as seen in Figure 6a), which can be understood as outliers (they do not follow the initial data distribution, in which the higher the degree, the lower the amount of clusters). This malicious clusters can be potentially related to a specific campaign, orchestrated by one or several users using a closed set of IPs, wallets, emails and domains. As an additional outcome, we depicted the distribution of Eigenvector centrality in Figure 7b. It can be observed that we have a cluster of nodes close to zero (corresponding to isolated nodes with few or none connections with highly connected nodes), and another cluster with a value above 0.08. The latter means that the nodes of the malicious clusters are highly interconnected between them and, in some cases, to other clusters. Therefore, in some occasions, the same assets (i.e. wallets, emails, IPs, or domains) have been used in more than one campaign, probably triggered by the same entities.

Finally, to identify additional relevant features, we explored the amount of updates that each domain had. In the analysed blockchain DNS systems, a domain can be updated by several reasons, such as renewing its time to live, assigning a new IP to it, or changing the *value* field to add extra options or information [58]. Our hypothesis was that highly active domains could be associated with malicious activities. In this regard, Figure 8 shows the top 20 most active domains in terms of updates. For instance, the most updated domains are `everypony.emc` and `mymonero.coin` and in both cases these domains are associated with malicious IPs. Nevertheless, since the vast amount of Emercoin domains only contain one interaction (corresponding to their creation operation) we went a step further and explored if the combination of updates and the number of different IPs associated to each domain over time, could be used to indicate the goodness of a domain name. Therefore, we computed a ratio considering the number of IPs and the number of updates for each domain as described in (1).

$$Ratio_{IPs,updates} = \frac{\text{Number of unique IPs}}{\text{Number of updates}} \tag{1}$$

Next, we selected a range from 1 to 10 to represent the number of updates and, for each value, we computed the average $Ratio_{IPs,updates}$ for the set of benign domains and the set of malicious ones (i.e. domains were tagged as malicious if they contain a malicious IP in their records). Note that we considered values equal or above a specific number of updates to compute each average. The values, as well as the associated t-test outcomes, are shown in Table 5.

Domain Type	**≥1**	**≥2**	**≥3**	**≥4**	**≥5**	**≥6**	**≥7**	**≥8**	**≥9**	**≥10**
benign	0.928	0.492	0.409	0.298	0.257	0.234	0.234	0.228	0.212	0.200
malicious	0.864	0.761	0.773	0.737	0.72	0.722	0.723	0.729	0.730	0.748
***t*-test values:** statistic = -5.5507 — p-value = 0.0002										

Table 5: Different $Ratio_{IPs,updates}$ average values considering a range of update values, and the corresponding t-test outcomes. Note that the column "≥1" considers all the domains existing in the blockchain.

As observed from the t-test outcome, the IP address updates are significantly higher for the domains engaging in malicious activity than the benign ($p = 0.0002$), where a malicious domain is expected to have twice as many IP updates as a benign one. This can be used as a composite indicator of compromise and tactics, techniques and procedures.

Note that the fact that most of Emercoin domains do not have more than one update hinders the classification for domains if we consider only such case. Nevertheless, the more updates, the more evident is the difference in the behaviour between benign and malicious domains.

The latter means that malicious domains use more IPs per update than benign ones, on average. Note that the insight provided by the t-test can be complemented with the total number of IPs registered in a domain.

4.2.2. Namecoin

In the case of Namecoin, we computed the same set of data structures as with Emercoin, to analyse the different relations between IP addresses and wallets. Figure 9a shows the correlation between the number of unique wallets and the top 20 IPs existing in Namecoin. It is noteworthy that, from the top five IPs, four were malicious except `91.250.85.116`, which was found to be suspicious by our hop-based association approach.

The next data structure, graphically depicted in Figure 9b, reports the correlation between the top wallets and the number of IPs related to each of them. It can be observed that the amount of IPs associated with these wallets is far lower than the numbers seen in Emercoin. Nevertheless, the latter is not related to a decrease in the number of malicious wallets. This is supported by the fact that, e.g. the top five wallets depicted in Figure 9b used malicious IP addresses in their domains.

Similar to Emercoin, we depicted in Figure 10 the geographical coverage of the Namecoin IP addresses and the malicious activities collected in Section 4.1. In this case, we observe a stronger correlation between the amount of IP addresses hosted and the reported malicious activities than in the case of Emercoin.

Next, by using Algorithm 1, we computed the set of suspicious IP addresses contained in Namecoin. In this case, in addition to the 2577 malicious reported IPs, we classified 1118 as suspicious ones, leaving 1431 as benign ones (i.e. only a 28% of the IPs were not connected to maliciously reported IPs). After computing such statistics, we depicted the graph representation of the Namecoin ecosystem in Figure 11a. As in the case of Emercoin, nodes represent the IPs, and edges represent a common value (e.g. wallet, email, domain) shared between them. If we compare the representations depicted in Figure 11a and Figure 11b, we can observe a substantially reduced number of benign nodes in the latter, since most of them appear to be isolated. In the case of suspicious nodes, they are correlated with malicious ones, exhibiting clearly identifiable clusters. Moreover, there are different sizes of malicious clusters, yet well represented due to the high connectivity between malicious IPs. In addition, we computed the CCDF and the eigenvector distribution and depicted them in Figure 12a and 12b, respectively. In the former case, we can observe a similar behaviour than the one discussed in Emercoin Section. That is, a set of malicious (according to the visual analysis of Figure 11a) high degree clusters is represented, breaking the data distribution into two identifiable subsets (i.e. the data follows a completely different distribution below and above 10^2). In addition, Figure 12b shows the eigenvector distribution of Namecoin. Again, there are two identifiable types of nodes in terms of centrality relevance, being the ones close to 0.05 the ones which denote higher connectivity, linking different malicious clusters. The average clustering coefficient of the network represented in Figure 11a is 0.446 and in the case of Figure 11b (discarding the isolated nodes) is 0.694. These numbers are lower than in the case of Emercoin due to the high amount of isolated nodes existing in Namecoin. Nevertheless, we can observe a rapid growth when we discard these isolated nodes. The latter means that, despite having some clusters which are not fully interconnected (especially small-sized ones), the average connectivity of the nodes when they belong to a cluster is high.

Next, we extracted the most updated domains in Namecoin and depicted them in Figure 13. It is worth to note that, for instance, in the case of the two most updated domains, the users always used a private IP (`127.0.0.1`). In this regard, the behaviour of apparently benign users is not always expected by the network in terms of information updates. Since in both cases the owner updated the domain with the same information that it previously had (i.e. without the need to do it nor any other justifiable reason). Next, we used Equation (1) with the benign and malicious subsets of Namecoin domains to compute the values for the same range than the one used in Emercoin, and depicted the results in Table6.

Domain Type	**≥1**	**≥2**	**≥3**	**≥4**	**≥5**	**≥6**	**≥7**	**≥8**	**≥9**	**≥10**
benign	0.396	0.304	0.278	0.249	0.225	0.216	0.194	0.146	0.137	0.121
malicious	0.504	0.410	0.380	0.354	0.345	0.328	0.344	0.353	0.339	0.343
***t*-test values:** statistic = -4.5437 — *p*-value = 0.0003										

Table 6: Different $Ratio_{IPs,updates}$ average values considering a range of update values, and the corresponding t-test outcomes. Note that the column "≥1" considers all the domains existing in the blockchain.

The values obtained in Namecoin denote the same behaviour than the ones observed in Emercoin, yet this time with lower average values. The latter is a consequence of the Namecoin renewal requirement, which translates into a higher number of updates per domain to overcome their expiration time. Therefore, malicious domains tend to have more IPs per update, provably to keep malicious campaigns alive during longer periods and avoid security measures such as blacklisting.

In addition to the previous experiments, we extracted the common public IPs in both Emercoin and Namecoin and found that a total of 55 IPs are shared between such systems (we did not consider public nor IPs used in well-known services or traditional DNS servers), from which 32 are malicious. The latter exhibits the possibility that the same actors are perpetrating malicious activities in both blockchains.

4.3. Use Case Example

To showcase some of the functionalities of the proposed correlation analysis approach, we extracted a set of malicious domains reported back in 2018 by FireEye in several campaigns, namely Gandcrab ransomware, CHESSYLITE, Neutrino and other samples [59]. First, we computed some basic statistics for each domain by querying our curated dataset. In this regard, several of the domains did not resolve to any IP (`bleepingcomputer.bit`, `nomoreransom.bit`, `esetnod32.bit`, `emsisoft.bit`, and `gandcrab.bit`), and some others (`brownsloboz.bazar`, `brownsloboz.lib`, and `brownsloboz.emc`) only contained private IPs so were not considered further. The rest of the domains were studied and their main statistics are described in Table 7. The Namecoin domains reported exhibit specific behaviours that are aligned with the outcomes reported in Section 4.2.2. With the exception of `flashupd.bit` and `cyber7.bit`, the domains used a set of different IPs which were associated with a large number of different wallets (i.e. several wallets were managing such IPs and used them in another domains as well, as reflected in Table 7, column *'Related Wallets'*). Moreover, the $Ratio_{IPs,updates}$ value of such domains (i.e. considering the total number of IPs and the number of updates), is aligned with the malicious behaviour observed in Namecoin. Next, we analysed which of these subset of domains were related in terms of IPs, wallets, or emails, and we observed that leomoon.bit lookstat.bit sysmonitor.bit volstat.bit and xoonday.bit shared common information. Moreover, we extended our search to find other domains that were correlated with these ones and we found the following list of domains: typeme.bit, browbaseis.bit, silikat.bit, vedixme.bit, testikname.bit, delix.bit, cash-money-analitica.bit, fooming.bit, firststat.bit, skildexin.bit, glesifax.bit, stamexis.bit, flexz.bit, checkxod.bit, money-cash-analitica.bit. Finally, we extended the list of suspicious IP addresses by using our hop-based association approach.

			IP classification breakdown			
Domain	**Updates**	**Related Wallets**	**Benign**	**Malicious**	**Suspicious**	**Total**
leomoon.bit	17	71	0	9	3	12
lookstat.bit	11	35	0	3	4	7
sysmonitor.bit	15	52	0	6	5	11
volstat.bit	16	48	0	7	3	10
xoonday.bit	15	76	0	10	0	10
flashupd.bit	1	2	0	1	0	1
cyber7.bit	1	1	0	1	0	1
brownsloboz.bit	6	14	0	4	1	5

Table 7: Statistics and IP classification of the studied Namecoin domains. Domains coloured in red denote a malicious clustered group.

4.4. Evaluation of the hop-based policy

Further to our initial experiments, we also evaluated the efficacy of our hop-based policy. To achieve this one would have to determine whether IPs that were classified as suspicious from our algorithm would be later identified by threat intelligence platforms. Note that platforms such as VirusTotal do not report the first time that an IP was classified as malicious but only the last analysis result and its date.

Leaving a timeframe of approximately six months, we queried VirusTotal for the IPs that our hop-based approach had classified as suspicious. The returned results proved our hypothesis as 47 of these IPs are now reported as malicious, as seen in Table 8. It should be noted, that our approach identifies sources from which an adversary may launch an upcoming attack. Therefore, our approach correctly identified such IPs in a predictive security manner.

5. Discussion

One of the conclusions that can be extracted from the outcomes discussed in the previous sections is that Namecoin and Emercoin are currently primarily used for malicious purposes since a huge share of the IPs registered in Emercoin and Namecoin are directly associated with malicious activities. Such statistics

185.117.119.190	192.241.241.153	54.37.229.180	89.223.88.183	185.86.148.137
91.235.129.241	210.16.101.109	108.167.140.18	185.222.202.206	193.106.31.146
51.89.177.5	192.3.12.121	5.34.180.226	185.101.105.232	111.90.149.240
45.141.84.190	5.252.176.7	45.153.184.158	185.14.187.128	209.141.36.7
23.239.84.135	31.220.23.1	192.99.178.153	95.217.74.220	172.82.152.132
45.32.236.82	185.147.14.237	145.239.47.64	185.13.36.121	64.44.51.117
195.123.237.156	93.115.28.9	185.107.94.36	5.83.163.2	51.81.112.135
194.5.249.247	138.68.149.171	185.82.202.123	109.201.133.111	104.203.229.17
23.92.93.233	107.174.86.134	108.170.40.59	173.249.5.248	5.182.210.180
109.234.35.166	104.161.32.111			

Table 8: Originally classified suspicious IPs for which VT reports malicious activity.

hinder the adoption of blockchain DNS systems and the trust of the community towards them. Therefore, the emergence of novel solutions overcoming the main drawbacks of blockchain DNS is required. After exploring the state-of-the-art and analysing the actual status of Emercoin and Namecoin, we identified different subsets of challenges applicable to these and other blockchain DNS systems. These challenges can be mainly classified into (i) the registration procedure and users behaviour, (ii) the extraction of information flows and their links with external threat analysis systems, and (iii) the security of the underlying blockchain platform and proactive measures.

There is an urgent need to improve the robustness and security of the registration procedures in blockchain DNS systems. One clear example relies on Emercoin registrar, which allows the use of case sensitive, non UTF-8, and other forbidden patterns and characters, as well as invalid domains according to RFC 1123 [60]. Furthermore, strategies to avoid, e.g. cybersquatting, are required, such as the one implemented by Handshake, which reserved the top 100k Alexa domains. In terms of user behaviours, specific control of the amount and speed of domains registered could help in detecting and reducing several campaigns. In this regard, we studied the behaviour of users and their strategies to avoid being linked or related to other activities in both Emercoin and Namecoin. While there exist several wallets containing a vast number of IPs in both systems, most malicious users follow the strategy of one-wallet one-IP. That is, to avoid being tracked, users often use different wallets with a low time-to-live (e.g. only for one IP update). The latter hinders the task of identifying malicious wallet-to-IP connections, especially since most of the interactions in the blockchain are of this nature. Nevertheless, our methodology is able to unveil these internal relationships by exploring the correlations in different dimensions, namely wallets, IPs, domains, and further information stored in the *value* field. For instance, we can leverage proactive security in blockchains, with, e.g., active checks focusing on the behaviour of the users, as well as the information associated with each wallet. As observed in the studied BDNS systems and due the possibility of having other potential indicators, we believe that exploring and assessing the different data managed by such systems is crucial to design the proper mitigation strategies. For example, parameters such as the amount of suspicious domains registered (e.g., domain squatting [56], or artificially generated domains [61]), the number of wallet updates, the IPs and domains registered, and the connectivity of the nodes are features that can be used to identify potentially harmful user behaviours. The latter can be augmented by our hop-based approach as well as similar methods following blacklisting policies, enhancing the reliability and trust of blockchain DNS while reducing the impact of malicious campaigns. Therefore, it is imperative to establish a holistic end-to-end approach, possibly through integrating smart contracts with revocation mechanisms [62, 63], to manage the registration procedure as well as to protect blockchain DNS system from misuse. Moreover, while we have to support security and privacy initiatives, the accountability perspective, especially when it comes to critical Internet infrastructures such as DNS must also be taken into consideration.

Another issue that we encountered during our investigation is that the bulk of threat intelligence sources lack information regarding blockchain DNS systems. Moreover, the intelligence collected from the sources used in this article is disparate and not homogeneous. For instance, only VirusTotal keeps track of requests to `.bit` domain but not to `.coin, bazar, .lib` and `.emc` domains. Hybrid Analysis does not keep track of any such requests. Notably, other platforms do not keep track of these domains, nor of their updates but monitor each connected IP individually. With the continuous rise of such schemes, the quest for information about such domains and their interconnections becomes even more necessary.

The timely collection of quality intelligence is crucial to detect cybercriminal campaigns and may lead to their prevention since methodologies like the one proposed in this article rely on such information to establish ground truth. Therefore, more efforts should be devoted to the active monitoring of the blockchain DNS ecosystem, including both their domains and IPs, in an automated way.

In the case of blockchain features, they are often recalled in their beneficial form, yet some of them can leverage malicious opportunities. The clearest example of this is immutability. In this regard, the impossibility of deleting records guarantees traceability and auditability of malicious campaigns, their modus operandi, and enables mitigation actions. For instance, we can leverage proactive security in blockchains, with, e.g. active checks focusing on the behaviour of the users, as well as the information associated with each wallet. The latter can be used to detect future campaigns by using, e.g. our hop-based approach as well as similar methods following haircut blacklisting policies, enhancing the reliability and trust of blockchain DNS while reducing the impact of malicious campaigns. Nevertheless, the impossibility of deleting, e.g. malicious records or illegal information, is a clear disadvantage. In this regard, there is still much work ahead to enable efficient blockchain deletion mechanisms [64, 7], since actual practices mainly rely on forks, and long block consolidation mechanisms, which add prohibitive overhead to blockchain systems. Aligned with the idea of forks, well-known systems such as Bitcoin and Ethereum have opted for forks as a solution to security issues or required protocol changes to enable further functionalities [65, 66]. Therefore, fork-based strategies, including novel and robust functionalities, could help in recovering the trust in Namecoin and Emercoin.

In principle, blockchains are considered to provide some form of privacy. While there is no transaction privacy, users through the use of multiple wallet addresses may enjoy some privacy guarantees. Hence, blockchain DNS approaches, beyond decentralisation, immutability, and resilience may provide some privacy guarantees to the owner of the domains, through, e.g. pseudoanonymisation. Notably, in our research we observe that even though both chains have several thousands of wallet addresses, users have opted to share self-identifying information such as emails allowing the linking of their wallets, defying the very scope of using different wallets for registering their domains. In fact, as discussed in the previous section, this behaviour is frequent, indicating the lack of understanding of how blockchains work from the users' perspective.

6. Conclusions

In this article, we provided a thorough analysis of the most mature blockchain DNS systems, namely Namecoin and Emercoin. In addition to reviewing the actual state-of-the-art of blockchain DNS systems, we proposed a sound and automated methodology to retrieve, process, and analyse the data stored in such systems. Thereafter, we recalled a set of blacklisting policies, namely blacklisting and haircut, and used the latter in our investigation to provide an insight into how Namecoin and Emercoin are used. The outcomes of our analysis, which includes internal correlations and external intelligence linked to several campaigns, concluded that the actual blockchain DNS ecosystems are being used for malicious purposes since more than 50.7% of the IPs used by the domains registered has been reported as malicious. Moreover, we developed a predictive association method to identify suspicious IPs (more than 24% of all IPs were tagged as suspicious), enabling proactive measures.

Finally, we identified and discussed the main challenges and proposed several ways to overcome them, according to the knowledge extracted from our analysis and the well-known flaws of blockchain DNS systems. Future work will focus on exploring other blockchain DNS systems and studying further proactive strategies to prevent malicious activities in blockchain ecosystems. Moreover, we will explore other strategies to identify malicious behaviour considering e.g., time-based thresholds, to capture potential active threats.

Acknowledgments

This work was supported by the European Commission under the Horizon 2020 Programme (H2020), as part of the projects *CyberSec4Europe* (Grant Agreement no. 830929) and *LOCARD* (Grant Agreement no. 832735).

The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

References

[1] W. E. Forum, The global risks report 2020, http://www3.weforum.org/docs/WEF_Global_Risk_Report_2020.pdf (2020).

[2] C. Shar, State of dnssec deployment 2016, Internet Society Report (2016).

[3] M. Anagnostopoulos, G. Kambourakis, P. Kopanos, G. Louloudakis, S. Gritzalis, Dns amplification attack revisited, Computers & Security 39 (2013) 475–485.

[4] F. Casino, T. K. Dasaklis, C. Patsakis, A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telematics and Informatics 36 (2019) 55 – 81.

[5] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, M. H. Rehmani, Applications of blockchains in the internet of things: A comprehensive survey, IEEE Communications Surveys & Tutorials 21 (2) (2018) 1676–1717.

[6] G. Ateniese, B. Magri, D. Venturi, E. Andrade, Redactable blockchain–or–rewriting history in bitcoin and friends, in: 2017 IEEE European Symposium on Security and Privacy (EuroS&P), IEEE, 2017, pp. 111–126.

[7] E. Politou, F. Casino, E. Alepis, C. Patsakis, Blockchain mutability: Challenges and proposed solutions, IEEE Transactions on Emerging Topics in Computing (2019) 1–1.

[8] H. A. Kalodner, M. Carlsten, P. Ellenbogen, J. Bonneau, A. Narayanan, An empirical study of namecoin and lessons for decentralized namespace design, in: WEIS, 2011.

[9] C. Patsakis, F. Casino, N. Lykousas, V. Katos, Unravelling ariadne's thread: Exploring the threats of decentralised dns, IEEE Access 8 (2020) 118559–118571.

[10] P. Hoffman, P. McManus, DNS queries over HTTPS (DoH), https://tools.ietf.org/html/rfc8484 (2018).

[11] Z. Hu, L. Zhu, J. Heidemann, A. Mankin, D. Wessels, P. Hoffman, Specification for dns over transport layer security (tls), https://tools.ietf.org/html/rfc7858 (2016).

[12] The exip project, https://exip.live/ (2021).

[13] The butterfly protocol, https://www.butterflyprotocol.io/ (2021).

[14] The brave browser, https://brave.com/ (2021).

[15] Andrew Brandt, Bazarloader deploys a pair of novel spam vectors, https://news.sophos.com/en-us/2021/04/15/bazarloader/ (2021).

[16] Cyware, Links discovered between bazar and trickbot, https://cyware.com/news/links-discovered-between-bazar-and-trickbot-2909546d (2020).

[17] Z. Huang, J. Huang, T. Zang, Leopard: Understanding the threat of blockchain domain name based malware, in: A. Sperotto, A. Dainotti, B. Stiller (Eds.), Passive and Active Measurement, Springer International Publishing, Cham, 2020, pp. 55–70.

[18] M. Möser, R. Böhme, D. Breuker, Towards risk scoring of bitcoin transactions, in: International conference on financial cryptography and data security, Springer, 2014, pp. 16–32.

[19] J. Bader, The domain generation algorithm of bazarloader, https://johannesbader.ch/blog/the-dga-of-bazarbackdoor/ (2020).

[20] E. S.-J. Swildens, R. D. Day, et al., Domain name resolution using a distributed dns network, uS Patent 7,725,602 (2010).

[21] C. Cachin, A. Samar, Secure distributed dns, in: International Conference on Dependable Systems and Networks, 2004, 2004, pp. 423–432.

[22] V. Ramasubramanian, E. G. Sirer, The design and implementation of a next generation name service for the internet, in: Proceedings of the 2004 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, SIGCOMM '04, Association for Computing Machinery, New York, NY, USA, 2004, p. 331–342.

[23] M. Wachs, M. Schanzenbach, C. Grothoff, A censorship-resistant, privacy-enhancing and fully decentralized name system, in: International Conference on Cryptology and Network Security, Springer, 2014, pp. 127–142.

[24] Z. Qiang, Z. Zheng, Y. Shu, P2pdns: A free domain name system based on p2p philosophy, in: 2006 Canadian Conference on Electrical and Computer Engineering, 2006, pp. 1817–1820.

[25] M. Abu-Amara, F. Azzedin, F. A. Abdulhameed, A. Mahmoud, M. H. Sqalli, Dynamic peer-to-peer (p2p) solution to counter malicious higher domain name system (dns) nameservers, in: 2011 24th Canadian Conference on Electrical and Computer Engineering (CCECE), IEEE, 2011, pp. 001014–001018.

[26] F. Casino, E. Politou, E. Alepis, C. Patsakis, Immutability and decentralized storage: An analysis of emerging threats, IEEE Access 8 (2020) 4737–4744.

[27] S. Al-Mashhadi, S. Manickam, A brief review of blockchain-based dns systems, International Journal of Internet Technology and Secured Transactions 10 (4) (2020) 420–432.

[28] D. Storm, P2p dns to take on icann after us domain seizures, `https://www.computerworld.com/article/2469753/p2p-dns-to-take-on-icann-after-us-domain-seizures.html` (2010).

[29] E. Karaarslan, E. Adiguzel, Blockchain based dns and pki solutions, IEEE Communications Standards Magazine 2 (3) (2018) 52–57.

[30] A. Hari, T. Lakshman, The internet blockchain: A distributed, tamper-resistant transaction framework for the internet, in: HotNets 2016 - Proceedings of the 15th ACM Workshop on Hot Topics in Networks, 2016, pp. 204–210.

[31] B. Benshoof, A. Rosen, A. Bourgeois, R. Harrison, Distributed decentralized domain name service, in: Proceedings - 2016 IEEE 30th International Parallel and Distributed Processing Symposium, IPDPS 2016, 2016, pp. 1279–1287.

[32] S. Gourley, H. Tewari, Blockchain backed dnssec, Lecture Notes in Business Information Processing 339 (2019) 173–184.

[33] Z. Guan, A. Garba, A. Li, Z. Chen, N. Kaaniche, Authledger: A novel blockchain-based domain name authentication scheme, in: ICISSP 2019 - Proceedings of the 5th International Conference on Information Systems Security and Privacy, 2019, pp. 345–352.

[34] J. Liu, B. Li, L. Chen, M. Hou, F. Xiang, P. Wang, A data storage method based on blockchain for decentralization dns, in: Proceedings - 2018 IEEE 3rd International Conference on Data Science in Cyberspace, DSC 2018, 2018, pp. 189–196.

[35] W. Wang, N. Hu, X. Liu, Blockzone: A blockchain-based dns storage and retrieval scheme, in: Artificial Intelligence and Security, Springer International Publishing, Cham, 2019, pp. 155–166.

[36] Z. Yu, D. Xue, J. Fan, C. Guo, Dnstsm: Dns cache resources trusted sharing model based on consortium blockchain, IEEE Access 8 (2020) 13640–13650.

[37] X. Duan, Z. Yan, G. Geng, B. Yan, Dnsledger: Decentralized and distributed name resolution for ubiquitous iot, in: 2018 IEEE International Conference on Consumer Electronics, ICCE 2018, Vol. 2018-January, 2018, pp. 1–3.

[38] W. Yoon, I. Choi, D. Kim, BlockONS: Blockchain based Object Name Service, in: ICBC 2019 - IEEE International Conference on Blockchain and Cryptocurrency, 2019, pp. 219–226.

[39] X. Wang, K. Li, H. Li, Y. Li, Z. Liang, ConsortiumDNS: A distributed domain name service based on consortium chain, in: Proceedings - 2017 IEEE 19th Intl Conference on High Performance Computing and Communications, HPCC 2017, 2017 IEEE 15th Intl Conference on Smart City,

SmartCity 2017 and 2017 IEEE 3rd Intl Conference on Data Science and Systems, DSS 2017, Vol. 2018-January, 2018, pp. 617–620.

[40] H. Li, H. Ma, L. Haopeng, Z. Huang, X. Yang, K. Li, H. Wang, Blockchain-based domain name resolution system, uS Patent App. 15/768,833 (May 30 2019).

[41] H. Li, X. Wang, Z. Lin, J. Wu, X. Si, K. Li, X. Yang, H. Wang, Systems and methods for managing top-level domain names using consortium blockchain, uS Patent App. 10/178,069 (2019).

[42] M. Ali, J. Nelson, R. Shea, M. J. Freedman, Blockstack: A global naming and storage system secured by blockchains, in: 2016 USENIX Annual Technical Conference (USENIX ATC 16), USENIX Association, Denver, CO, 2016, pp. 181–194.

[43] Seize and desist? the state of cybercrime in the post-alphabay and hansa age (2017).

[44] KrebsOnSecurity, Carders park piles of cash at joker's stash, `https://krebsonsecurity.com/2016/03/carders-park-piles-of-cash-at-jokers-stash` (2016).

[45] Abuse.ch, .bit - the next generation of bulletproof hosting, `https://abuse.ch/blog/dot-bit-the-next-generation-of-bulletproof-hosting/` (2017).

[46] T. Micro, .bit domain used to deliver malware and other threats, `https://www.trendmicro.com/vinfo/us/security/news/cybercrime-and-digital-threats/bit-domain-deliver-malware-and-other-threats` (2013).

[47] H. Wang, Fbot, a satori related botnet using block-chain dns system, `https://blog.netlab.360.com/threat-alert-a-new-worm-fbot-cleaning-adbminer-is-using-a-blockchain-based-dns-en/` (2018).

[48] S. Pletinckx, C. Trap, C. Doerr, Malware coordination using the blockchain: An analysis of the cerber ransomware, in: 2018 IEEE Conference on Communications and Network Security, CNS 2018, 2018, pp. 1–9.

[49] K. G. Randi Eitzman, J. Valdez, How the rise of cryptocurrencies is shaping the cyber crime landscape: Blockchain infrastructure use, `https://www.fireeye.com/blog/threat-research/2018/04/cryptocurrencies-cyber-crime-blockchain-infrastructure-use.html` (2018).

[50] C. Patsakis, A. Chrysanthou, Analysing the fall 2020 emotet campaign (2020). arXiv:2011.06479.

[51] L. Abrams, Gandcrab ransomware distributed by exploit kits, appends gdcb extension, https://www.bleepingcomputer.com/news/security/gandcrab-ransomware-distributed-by-exploit-kits-appends-gdcb-extension/ (2018).

[52] Microsoft, Behavior monitoring combined with machine learning spoils a massive dofoil coin mining campaign, https://www.microsoft.com/security/blog/2018/03/07/behavior-monitoring-combined-with-machine-learning-spoils-a-massive-dofoil-coin-mining-campaign/ (2018).

[53] Emercoin, Emercoin links & resources, https://emercoin.com/en/documentation/links-resources (2019).

[54] L. Böck, N. Alexopoulos, E. Saracoglu, M. Mühlhäuser, E. Vasilomanolakis, Assessing the threat of blockchain-based botnets, in: 2019 APWG Symposium on Electronic Crime Research (eCrime), IEEE, 2019, pp. 1–11.

[55] K. Perlow, Mapping out decentralized namecoin and emercoin infrastructure, in: Black Hat USA, Las Vegas, 2018.

[56] P. Xia, H. Wang, Z. Yu, X. Liu, X. Luo, G. Xu, Ethereum name service: the good, the bad, and the ugly, arXiv preprint arXiv:2104.05185 (2021).

[57] S. Null, In response to the ukrainian (and not only) prohibitions: decentralized emerdns system against site blocking, https://sudonull.com/post/70519-In-response-to-the-Ukrainian-and-not-only-prohibitions-decentralized-EmerDNS-system-against-site-blo (2019).

[58] Emercoin, The emercoin nvs, https://emercoin.com/en/documentation/blockchain-services/emernvs (2021).

[59] J. V. Randi Eitzman, Kimberly Goody, How the rise of cryptocurrencies is shaping the cyber crime landscape: Blockchain infrastructure use, https://www.fireeye.com/blog/threat-research/2018/04/cryptocurrencies-cyber-crime-blockchain-infrastructure-use.html (2018).

[60] R. Braden, Rfc1123: Requirements for internet hosts-application and support (1989).

[61] F. Casino, N. Lykousas, I. Homoliak, C. Patsakis, J. Hernandez-Castro, Intercepting hail hydra: Real-time detection of algorithmically generated domains, arXiv preprint arXiv:2008.02507 (2020).

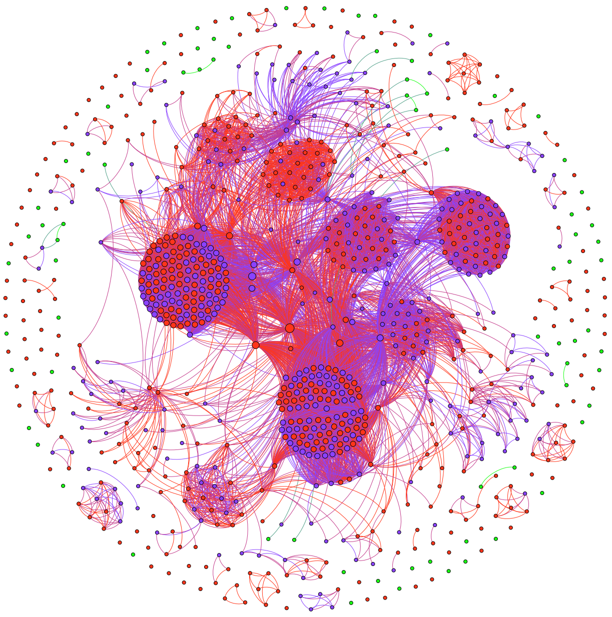
[62] Y. Yu, Y. Zhao, Y. Li, X. Du, L. Wang, M. Guizani, Blockchain-based anonymous authentication with selective revocation for smart industrial applications, IEEE Transactions on Industrial Informatics 16 (5) (2020) 3290–3300.

[63] J. P. Cruz, Y. Kaji, N. Yanai, Rbac-sc: Role-based access control using smart contract, IEEE Access 6 (2018) 12240–12251.

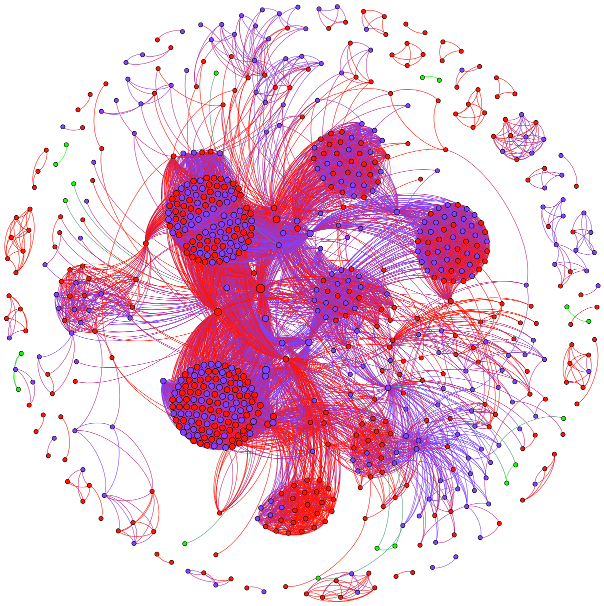
[64] K. Maeda, M. Ohtani, Y. Oishi, C. Yasumoto, J. Zhu, Deletion of blocks in a blockchain, uS Patent 10,739,997 (Aug. 11 2020).

[65] F. Schär, Blockchain forks: A formal classification framework and persistency analysis, Munich Personal RePEc Archive (2020).

[66] T. Neudecker, H. Hartenstein, An empirical analysis of blockchain forks in bitcoin, in: International Conference on Financial Cryptography and Data Security, Springer, 2019, pp. 84–92.

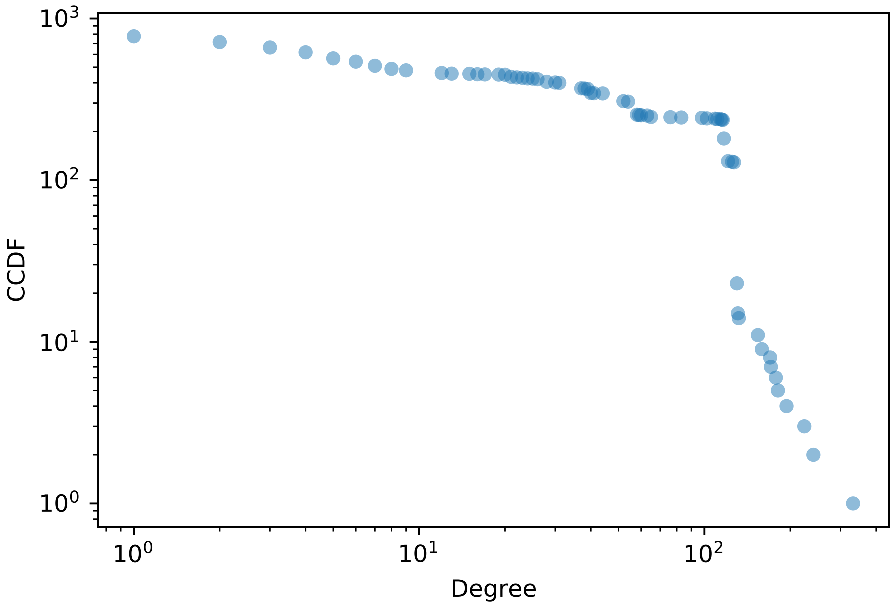


(a) Emercoin representation including all isolated nodes, where each node represents an IP, and their size is weighted according to their connectivity. The edges represent commonly shared data between nodes, such as wallets, emails or domains.

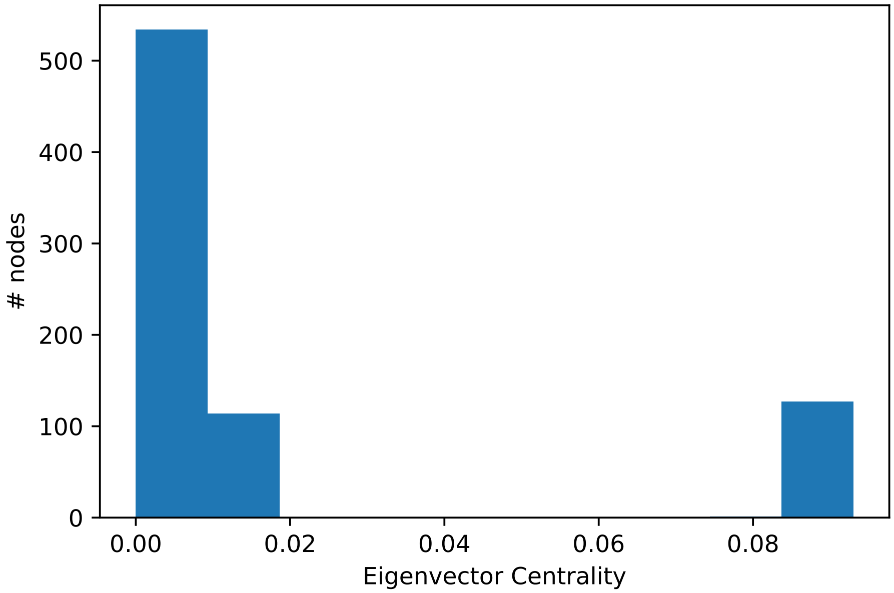


(b) Emercoin graph representation excluding isolated nodes. It can be observed that only a reduced number of benign nodes are present.

Figure 6: Graph-based representation of the Emercoin ecosystem.



(a) CCDF of Emercoin.



(b) Eigenvector centrality distribution of Emercoin

Figure 7: CCDF and eigenvector centrality values of the Emercoin ecosystem.

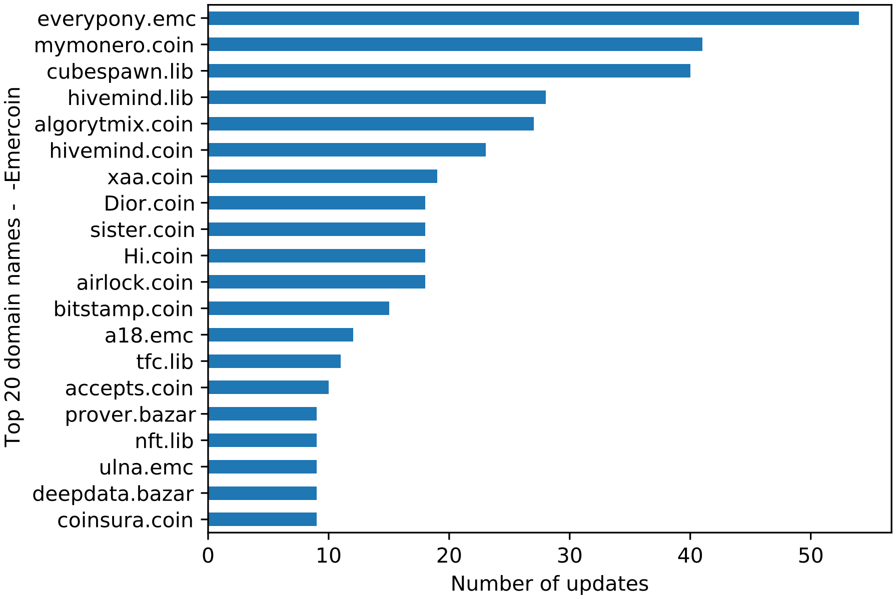
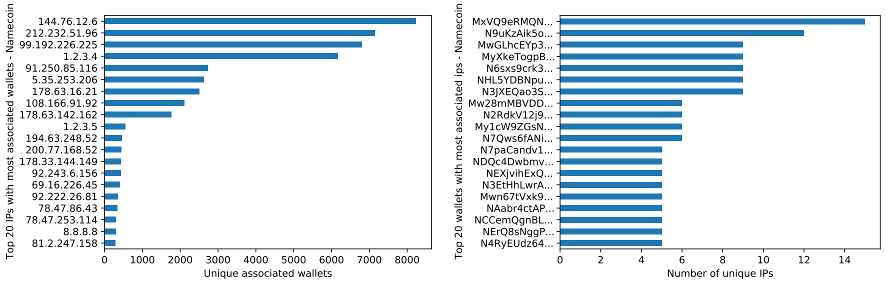


Figure 8: Top 20 most updated domains in Emercoin.



(a) Top 20 most used IPs in Namecoin and the number of wallets using them.

(b) Top 20 Namecoin wallets and the corresponding number of IPs found in their domains.

Figure 9: Statistics about the most used IPs and biggest wallets of Namecoin.

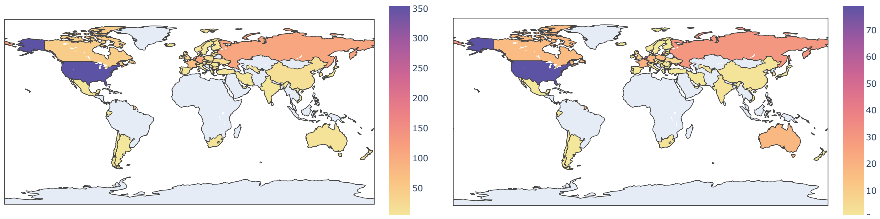
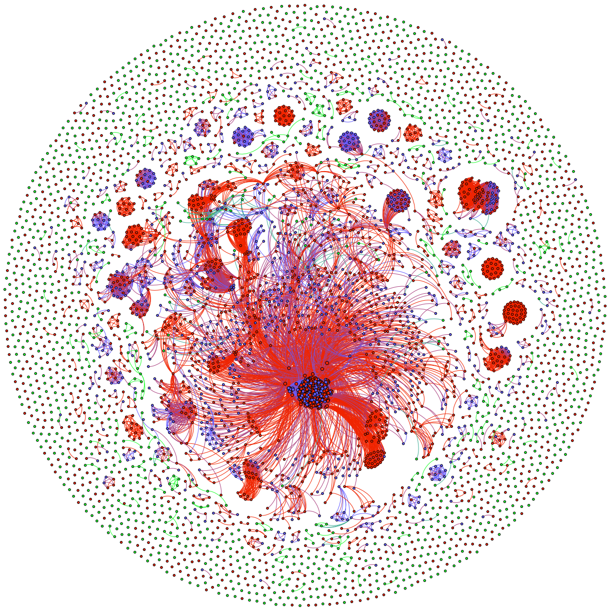
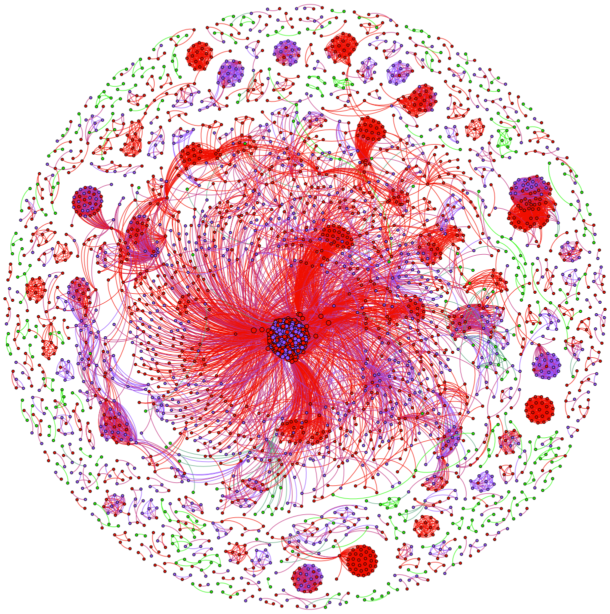


Figure 10: Geographical coverage heatmap of IPs mapped in Namecoin (left) and the corresponding malicious reports (right).

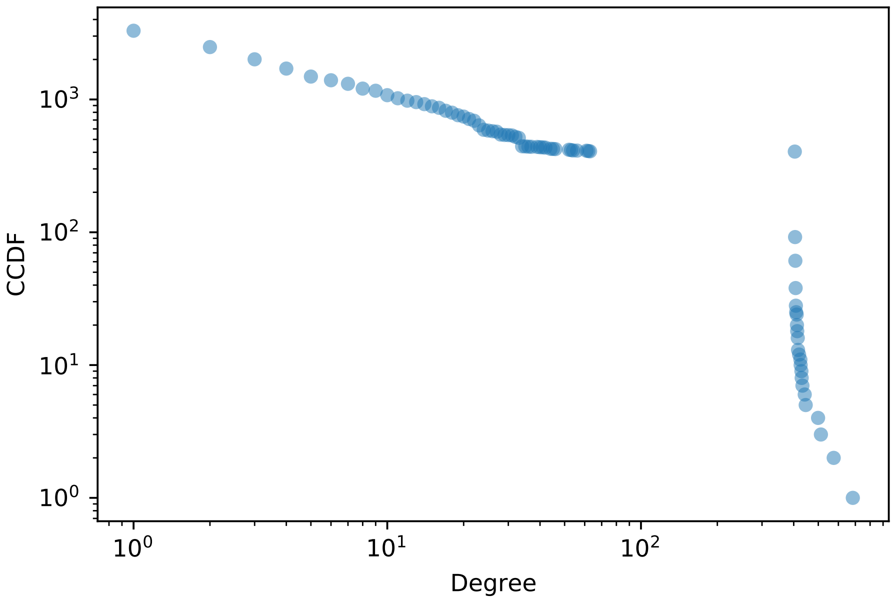


(a) Namecoin representation including all isolated nodes, where each node represents an IP, and their size is weighted according to their connectivity. The edges represent commonly shared data between nodes, such as wallets, emails or domains.

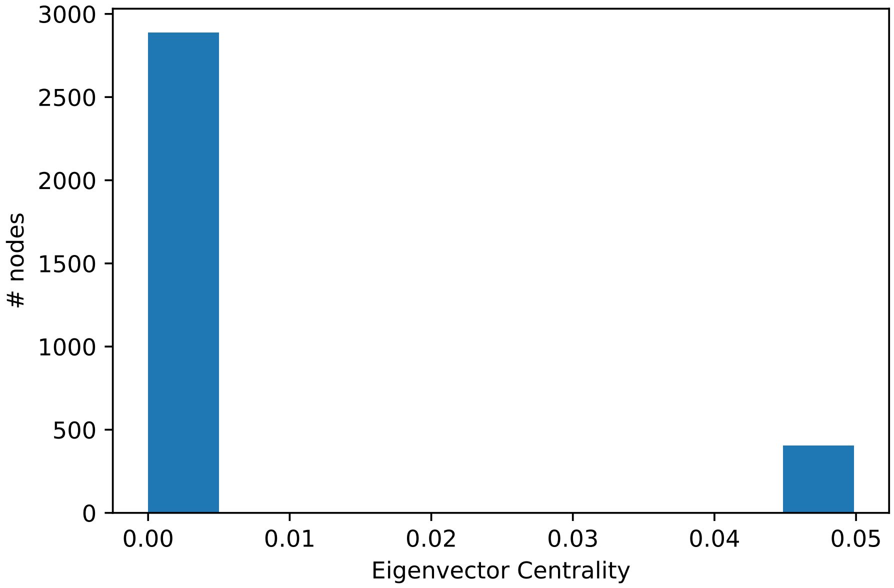


(b) Namecoin graph representation excluding isolated nodes. It can be observed that the amount of benign nodes is substantially reduced.

Figure 11: Graph-based representation of the Namecoin ecosystem.



(a) CCDF of Namecoin.



(b) Eigenvector centrality distribution of Namecoin.

Figure 12: CCDF and eigenvector centrality values of the Namecoin ecosystem.

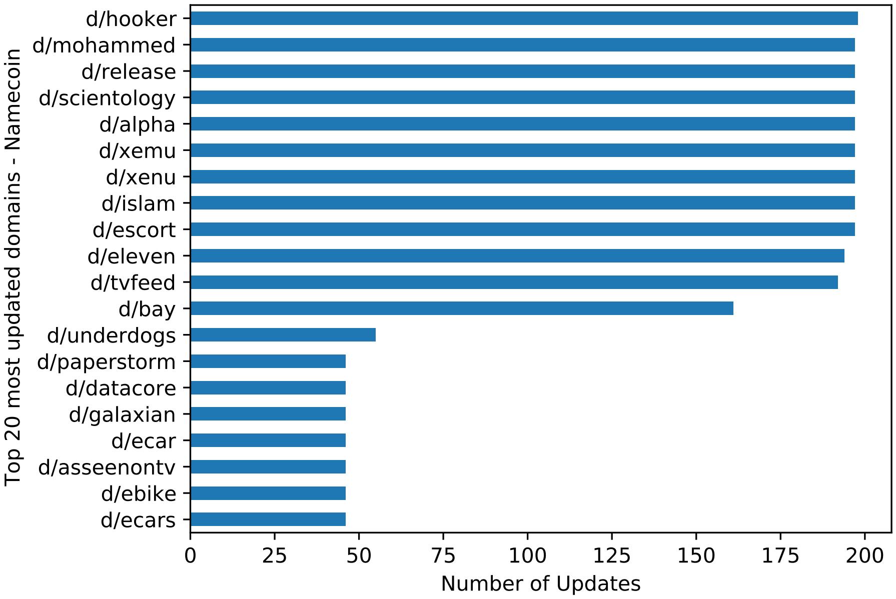


Figure 13: Top 20 most updated domains in Namecoin