A Recent Year On the Internet: Measuring and Understanding the Threats to Everyday Internet Devices

Afsah Anwar*
Northeastern University,
Florida A&M University
United States
afsah.anwar@famu.edu

Yi Hui Chen
Northeastern University
United States
chen.yihui@northeastern.edu

Roy Hodgman
Rapid 7
United States
roy_hodgman@rapid7.com

Tom Sellers
runZero
United States
tom@fadedcode.net

Engin Kirda
Northeastern University
United States
ek@ccs.neu.edu

Alina Oprea
Northeastern University
United States
a.oprea@northeastern.edu

ABSTRACT

An effective way to improve resilience to cyber attacks is to measure and understand the adversary's capabilities. Gaining insights into the threats we are exposed to helps us build better defenses, share findings with practitioners, and identify the perpetrators to limit their impact. Honeypot interactions have been widely studied in the past to measure cyber attacks, but the focus of more recent honeypot studies has been on IoT-based threats. Hence, classic threats studied by honeypots in depth a decade ago, such as desktop malware and web threats, have lately received much less attention.

In this paper, we perform a measurement study on a large-scale honeypot data collected between July 2020 and June 2021 by a large cybersecurity company. We measure a set of 7 billion connections to extract 806 million alerts raised by 662 endpoints (honeypots) distributed globally. For this study, we create a framework that leverages Open Source Cyber Threat Intelligence (OSCTI) to generate high-level attack classification and malware campaign inferences. One of the main findings of our work is that some networks involved in rogue activities that were reported in literature more than a decade ago [59] are *still* involved in malicious activity. Also, we find that 17 vulnerabilities disclosed more than a decade ago, even as early as 1999, are still used to launch cyber attacks. At the same time, the threat landscape has evolved. We discover that a large fraction of recent campaigns (63.4%) are Stealers or Keyloggers, new attack vectors such as the SMB EternalBlue vulnerability enable rapid self-propagation of malware across the globe, and infection strategies are shared among multiple campaigns (e.g., 10K alerts for Gafgyt, Trickbot, Freakout, and Hajime utilize the infection strategy of Mirai or muBot).

ACM Reference Format:
Afsah Anwar, Yi Hui Chen, Roy Hodgman, Tom Sellers, Engin Kirda, and Alina Oprea. 2022. A Recent Year On the Internet: Measuring and Understanding the Threats to Everyday Internet Devices. In *Annual Computer Security Applications Conference (ACSAC '22),December 5–9, 2022, Austin, TX, USA.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3564625.3564649

1 INTRODUCTION

The increasing Internet-enabled connectivity has incentivized attackers to compromise Internet-connected devices and launch global malware campaigns [32, 43]. The availability of Internet scanning tools such as Zmap, Nmap, Masscan, Shodan, and Censys, offer attackers easy identification of vulnerable endpoints across the globe. For example, recently, the XBash malware scanned the Internet to find potential targets to launch ransomware attacks [19]. Furthermore, with the increasing number of IoT devices, the attack surface and exposure to risk have considerably increased.

Considering these factors, it is essential to develop technologies to understand the threat landscape posed to Internet-connected systems. A variety of approaches have been shown to be effective in observing malicious behavior on the Internet. These techniques capture or monitor malicious activities and consist of approaches such as low or high interaction honeypots [35], Internet telescopes, Darknets, or Blackholes [11, 12], and collecting firewall and IDS logs from a large number of heterogeneous sources [5]. Honeypots are a widely used technology which enable unique access to the behavior of the attackers on the Internet and allows real attack observations. Honeypots have been the focus of malware research in the past and resulted in a number of research papers and systems deployed more than a decade ago [2, 4, 17, 23, 35, 36].

Today, the threat landscape is more diverse and complex than it used to be when previous honeypot papers were published. As a result, the general wisdom suggests that server-side honeypots might not be as valuable as in the past because Internet attacks, and the Internet itself have significantly evolved, and measuring attacker behavior is more challenging [13]. Acknowledging this change, our goal in this work is to revisit honeypot data analysis in order to investigate the threat landscape in an evolved threat ecosystem, re-affirming the usability of distributed honeypot networks. We aim to determine if there are new insights that can be

*Work done while at Northeastern University

ACSAC '22, December 5–9, 2022, Austin, TX, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9759-9/22/12…$15.00
https://doi.org/10.1145/3564625.3564649

gained from honeypot data analysis, and compare these to the previously reported trends from honeypot papers published more than a decade ago.

We gained access to a large honeypot dataset collected by Rapid7 during 2020-2021, consisting of 7 billion connections that raise 806 million network alerts on 662 honeypots. We analyze the largest-to-date honeypot data to identify and measure the different attack vectors that threaten Internet services, and compare them with historical data from previous honeypot papers. Through our analysis, we attempt to investigate the following research questions: (1) What are the attack vectors that threaten the Internet devices today? (2) How is the current threat landscape different from that reported more than a decade ago? (3) What are the newly observed trends in malware campaigns? (4) How are exploits spread and weaponized today?

Towards answering these questions, we design an Open Source Cyber Threat Intelligence (OSCTI) framework to infer the attack and campaign attributes of the alerts. The framework assigns these attributes as high-level descriptions that indicate attacker strategy, target, impact, campaign, and objective. Using our framework, we then analyze in detail the observed attack vectors, their geographical patterns, and malware campaign characteristics. Our analysis of the recent characteristics of malware campaigns shows, as expected, that the majority of malware (63.4%) is involved in stealing sensitive data such as credit cards, or keylogging. During the one year observation period, we determined that a large volume of alerts were attributed to PurpleFox (871K), Android Cerberus (272K), Fileless (1.1M), and Cobalt Strike malware (152K). Additionally, we found a large number of Remote Access Trojans (RATs) (146K), Mirai (93K), Gafgyt (93K), and Crypto-mining (23K) malware, and 8.7K alerts involved in ransomware activities. Moreover, we find eight alerts attributed to the NSO group's spyware. Furthermore, a majority of campaigns (62.4%) are stealers/keyloggers. An interesting finding is imitation of attack strategies among multiple malware campaigns. For instance, around 10K alerts for Gafgyt, Trickbot, Freakout, and Hajime campaigns utilize the infection strategy of the Mirai or muBot campaigns. Additionally, we find hosts employing multiple campaigns (49 subnets involved in two or more campaigns), and evidence of collaborative exploitation (a subnet was seen using 254 out of 256 hosts to exploit the SMBGhost vulnerability).

One of the most important insights of our work is that although the threat landscape has evolved substantially over time, malicious networks and vulnerabilities known for more than a decade ago are still prevalent today. For instance, we find that 74% of the rogue networks detected as malicious more than a decade ago [59] are *still* involved in similar activities today. This is highly concerning and shows the limitations of existing defenses, such as blacklisting and threat intelligence sharing, which do not appear to be sufficient to isolate malicious actors from the Internet. We also discover the continuing exploitation of well-known vulnerabilities that were publicly disclosed between 1999 and 2009, and are still creating significant damage on the Internet. Although widely known, these 38 vulnerabilities are still being actively exploited (17 of them by malware campaigns), affecting a wide range of applications.

Contributions. To summarize, this paper makes the following contributions:

- We analyze a dataset of 806 million alerts from 7 billion connections made by 662 honeypots. To the best of our knowledge, our work is the most extensive study of honeypot data to date, and the first comprehensive look at the threats captured by a honeypot after more than decade.
- We design an Open Source Cyber Threat Intelligence (OSCTI)-based framework to extract actionable intelligence, and gather attack and campaign inferences from the alerts.
- We compare the threats reported in studies conducted more than a decade ago, and investigate how these threats have evolved since then. We find that 17 networks involved in malware campaigns today overlap with the 24 rogue networks that were reported more than a decade ago [59]. We investigate the re-arming of old attack vectors by current malware campaign and find 17 vulnerabilities disclosed more than two decades ago that are still being actively exploited by a variety of campaigns, including APTs, RATs, and Emotet.
- We analyze recent malware threats on the Internet and discover some unique insights. We find that 63.4% of the inferred campaigns are Stealers or Keyloggers, and new attack vectors such as the SMB EternalBlue vulnerability enable rapid self-propagation of malware across continents. We find empirical evidence of shared strategies among campaigns, shared infrastructure being used among various campaigns, and report on collaborative exploitation to amplify a campaign's impact.

2 DATASET AND AUGMENTATION

In this section, we describe the honeypot dataset used for our analysis and our OSCTI alert summarization framework.

2.1 Honeypot Dataset

To understand the current threat behavior, we analyzed the large-scale traffic received by a 662 honeypots over a one-year period. These low-to-medium interaction honeypots are deployed across different geographies. A portion of the 7 billion incoming connections to the honeypots are identified as alerts depending on their interaction. We focus our work on the identified alerts. Overall, our dataset accounts for a wide range of alerts identified over a span of 12 months — July 2020 to June 2021.

Data Collection Infrastructure. The data collection infrastructure comprises of Linux-based honeypots emulating applications for different OS (Windows and Linux) and architectures to facilitate realistic interactions. The infrastructure, deployed by Rapid 7 [1], captures the PCAP files which are transferred into Amazon S3 buckets. These PCAPs are then analyzed by a network security monitoring tool (Suricata) to identify exploitation events and the ruleset is updated daily. Consequently connections or sequences of connections are identified as alerts, if they match the rules.

We conduct the largest study to date, analyzing 806 million alerts raised by 7 billion connections made by the industry-scale globally distributed network of honeypots, covering a large spectrum of services. These alerts are raised from 7 million hosts. Prior works, however, are limited by size and focus. Most of the works deploy research-purpose honeypots on university networks [17, 35], support limited services, or focus on specific infrastructures, e.g., cyber

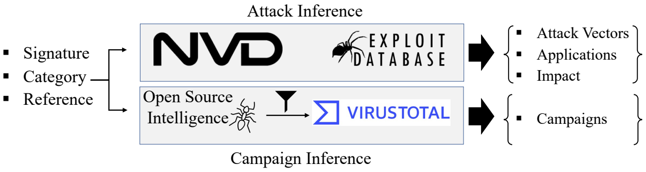


Figure 1: Alert summarization workflow. The attack and campaign attributes of the alerts are summarized using a variety of intelligence resources.

physical systems [27] and IoT [54, 63, 66]. We discuss the characteristics of our dataset, including the targeted services and threat actors in detail in Appendix A.

Although we see alerts raised on all ports, we find that only 20 ports make up for ≈ 60% of the alerts. The alerts are attributed to 2.8 million hosts distributed in 237 out of 250 ISO 3166-1 [3] countries. To investigate this collaborative nature among the threat actors, we map the hosts to the subnets that they belong to. We see that 10% of the alerts are from eight Class C subnets. In addition, 50% and 75% of the alerts originate from only 177 and 1348 subnets, respectively.

2.2 OSCTI for Alert Summarization

Malware campaigns continuously evolve over time, generating a large number of variants. At the same time, newer threats regularly appear, as well. Considering these evolving and diverse threats posed to Internet-connected devices, it is impossible for the threats to be centrally-monitored. However, collaborative efforts of various security organizations and individual researchers can help provide more identification and attribution of the threats. Therefore, we augment the alerts by leveraging public information from OSCTI. The high-level alert descriptions (categories/signatures) do not ascertain the strategy of the adversaries. While ports have been used for attribution and analysis of targeted applications, they often miss on the tunneled abuse of applications. Figure 1 describes the augmentation process. The alerts are first mapped with the National Vulnerability Database (NVD) [44] and exploitDB [57] to find additional details of the attack. Following, we lookup if attack sources have been reported by the community to be involved in malware campaigns. We explain these steps in detail in the following.

2.2.1 Attack Inference. Attack inference aims to identify the causality of the alerts. Towards this, we begin by leveraging the vulnerability and exploit references in the alert signature and/or reference. We design a framework that maps the alerts to the NVD. Additionally, it maps the publicly known exploits in the exploit database exploitDB. These may include exploits for known vulnerabilities and other bugs. Leveraging the NVD and the exploitDB, we extract vulnerability descriptions for the alerts. However, when doing so, we do not consider the indirect exploits, i.e., the exploits for the vulnerabilities in the NVD that are not mentioned in the alert references. The rationality behind this is that a vulnerability can affect multiple products, whereby each of the products will have different exploits, and the raised alert may or may not be targeting the application targeted by the exploits listed in the NVD. We then build heuristics to infer the attack attributes such as attack vectors, targeted application, and alert impact.

2.2.2 Campaign Inference. To identify the campaign an alert is associated with, we utilize the publicly-available indicators of compromise (IoC) shared by security organizations and independent security researchers, namely, Amnesty International [29], CyberMonitor's APT & cybercriminals campaign collection [22], Stampram's malicious trail [22], Executemalware's daily IOCs feed from malware investigations [26], CronUp's malware IoCs [21], Palo Alto Network's Indicators from Unit 42 Public Reports [42], Florian Roth's indicators of compromise feed [55], and ESET Research's IoC's [53]. However, we consider that open source IoCs pose a challenge to their practical applicability. For instance, the IoCs, such as IP addresses are time-sensitive. Therefore, we apply two stages of filtering on the IoCs.

(1) Timeliness. To account for their timeliness, the IoCs should have been observed during our period of observation.
(2) Validation. Apart from timeliness, the crowd-sourced IoCs are also prone to false positives. To remove the false positives, we utilize the VirusTotal service. In particular, we check if the IP addresses has been reported as being involved in malicious activity during our analysis window.

The validated IoCs are then assigned to the campaign they have been reported against.

2.2.3 Campaign Objective. A malware campaign uses targeted methods to achieve its objective. For instance, the SpyNote malware is an Android spyware that logs keystrokes. This module maps the alerts with the campaign objectives through manual lookup using online resources, such as *Malpedia* [52]. Ultimately, the campaigns are labeled with objectives, such as spyware, stealers/keyloggers, financial, critical information, etc.

3 HONEYPOT ATTACK ANALYSIS

In this section, we measure the alerts by utilizing the summaries augmented in Section 2. We identify the different attack vectors that target the devices on the Internet. We then discuss the targeted applications and the impact of the alerts on the devices.

3.1 Attack Vectors

To realize the alert causes and to understand their alert profile, we focus individually on the attack vectors. We investigate four specific attack vectors: Malware, Bruteforce (including Credential), Privilege escalation, and Exploit-based alerts. Apart from the obvious vectors, malware and exploits, the use of default and commonly-used credentials have been popular infection strategy on the Internet [7]. Additionally, privilege escalation is a common method to use infected systems to their full potential and is listed as part of the MITRE ATT&CK framework [39].

3.1.1 Malware. Approximately 2% of the alerts are malware-related. The malware attacks have changed and now pose multi-cornered threats. We identify three different categories of malware attacks, namely: Traditional (the PC and mobile malware), Internet of Things (malware targeting IoT devices), and Cryptojacking (the adversary use the compromised devices to mine cryptocurrencies).

We find that cryptojacking and traditional malware are more concentrated towards a fewer number of countries, compared to the IoT malware (Table 8, in Appendix). For traditional and cryptojacking malware, four countries originate >75% (three for cryptojacking)

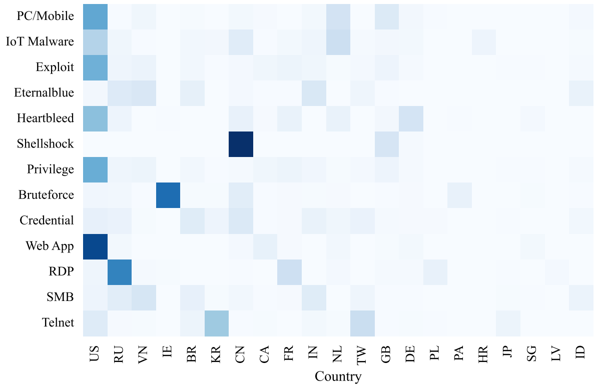


Figure 2: Attack vectors, applications, and impact and their prominent sources. The countries represented are limited to top 5 countries (by #alerts) in their respective categories.

and 12 countries result in >90% of their alerts. Similarly, for traditional malware, we note that nine countries have >75% of the alerts and 21 countries have >90% of their alerts. Additionally, Figure 2 lists the top countries that serve as alert sources for each of the categories. We find that both traditional and IoT malware have United States and Netherlands as the most frequent attack sources, while traditional malware is more skewed towards the United States (45% vs. 26%). Moreover, traditional malware employs a wider range of tactics – such as, targeting RDP and utilizing remote code exploits.

A majority of the cryptojacking alerts (99.97%) occur due to remote code execution exploits. Prior works have shown the ability of adversaries to inject Javascript into their websites that use the computational power of the host machine to mine cryptocurrencies [28]. Bijmans *et al.* [16] showed the exploitation of a vulnerability in Mikrotik OS-based routers (CVE-2018-14847) to mine cryptocurrencies. We find that a majority of the alerts (99.78%) exploit this vulnerability. In addition, we find the cryptojacking malware exploiting two other vulnerabilities, namely, CVE-2017-12615 [46] and CVE-2017-9805 [47]. Moreover, we also find exploits targeting docker images to mine cryptocurrencies.

Campaigns. Through our OSCTI-based campaign inference, we identify malware campaigns that the alerts are associated with. Overall, we identify 118 campaigns, such as Emotet and advanced persistent threats (APTs). The most frequent campaigns include: PurpleFox (871K alerts), Android Cerberus (272K), Mirai (93K), Gafgyt (93K), and crypto miners (23K). We also identify RATs (146K), such as Android Roamingmantis (51K), njRat and Magnetcore (55K), and observede Cobalt Strike (152K) and PowerShell injectors (153K). We discuss some of these campaigns in detail in Appendix C.

Surprisingly, we also identify two spyware campaigns, including FinSpy, the German spyware targeting human right defenders in Uzbekistan. The two spyware campaigns generate eight alerts in total. Additionally, we find hosts that are involved in multiple campaigns. Specifically, we note that there is high infrastructure sharing among the RATs, e.g., njRat and Magentocore. We discuss the attributes of the malware campaigns in Section 4.2.1.

3.1.2 Exploits. More than 8.4% of the alerts arise due to vulnerability exploitation. Adversaries utilize the exploits to abuse various applications, with most of these alerts stemming from network-facing services. Earlier, we saw that IoT malware utilize exploits towards their intent. However, those make up for only 2.6% of the exploits. A vast majority of the exploits target weaknesses in SMB, web applications, and devices (e.g., routers).

Adversaries leverage exploits to attempt at DNS amplification, DNS change, DNS lookup, DNS named authors, and versions attacks. A majority of such attacks target D-Link and ADSL routers (5287 or 0.008% of alerts). We also identify two alerts that exploit weaknesses in device firmware. Apart from weaknesses in the firmware, adversaries also target specific devices to break into them. In fact, 0.27% of the exploits target specific devices.

The secret key compromising vulnerability, Heartbleed, and the arbitrary code execution vulnerability, Shellshock, were disclosed in 2014. Additionally, the exploit targeting SMBv1, EternalBlue, was leaked in 2017. These are critical vulnerabilities, e.g., EternalBlue was actively exploited by the WannaCry and Petya ransomware [10]. Surprisingly, we still observe 5.3 million (7.9%), 10,500 (0.016%), and 1,000 (0.00001%) alerts for EternalBlue, Heartbleed, and Shellshock exploits, respectively.

The exploits originate from 219 countries: Shellshock maps to only three countries (China (84.3%), Britain (14.2%), and Germany (1.5%)), Heartbleed exploits map to 38 countries with >90% traffic originating from 12 countries, and the EternalBlue exploits is spread globally to 195 countries. We further look for the ASNs and the organizations that are responsible for the Shellshock exploits, and observe that they originate from only three organizations, China Telecom, British Telecom, and Telefonica Germany. Figure 2 lists the top countries generating exploit-based alerts. While the exploits mostly originate from the United States, the Shellshock and the EternalBlue exploits frequently originate from smaller countries. United States constitutes 42% and 35% of the overall exploits and Heartbleed, respectively. However, EternalBlue is vastly exploited from Vietnam and India. Additionally, Vietnam is mapped to >5.23% of the exploits, second to only the United States. However, Vietnam does not appear in the countries that exploit the Heartbleed vulnerability. Similarly, although United States appears to be the most common origin for overall exploits and Heartbleed, it ranks as the tenth among EternalBlue exploiters.

3.1.3 Brute Force Attacks. These attacks involve attempts by unauthorized users to gain access to a privileged account. While such attacks mostly involve achieving unauthorized access to a device or service, they can also be used as a means to guess file or directories on a server. A vast majority of the brute force attacks target the SSH protocol (99.93%). We also see such attacks targeting email services (2143 alerts) and MySQL (22464 or 0.2% alerts). The attacks originate from 198 countries. While United States appears as the most frequent source of attack for exploits and malware, as shown in Figure 2, brute force attacks mostly occur from Ireland (>64.5%). We also see that >90% alerts are mapped to only nine countries, and a majority of these attacks originate from smaller countries. Furthermore, 91.8% of the alerts from Ireland originate from the organization Global Layer BV (65.1% of all alerts).

Credentials. In many cases, adversaries use default or common credentials to target poorly-configured devices. Such attacks have been widely known to be employed by Mirai to target routers [6–8]. We find that a large set (84.5%) of alerts arise due to such attempts. In addition, these attempts can be manufacturer specific – Muhstik

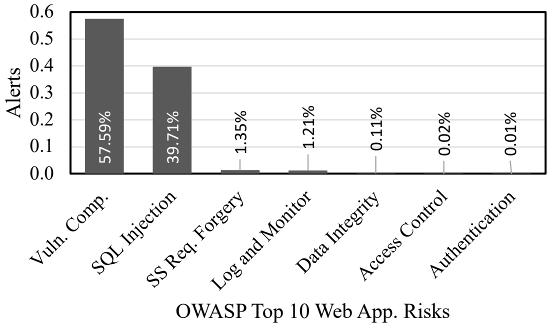


Figure 3: The 6^{th} and the 3^{rd} ranked risk appear as most frequent in our dataset.

botnet targeted the Tomato routers using default credentials. We find alerts targeting the Tomato router with specific credential pairs (admin/admin and root/admin) (3.1%) and Apache Tomcat servers' root account (admin/blank) (0.05%). Moreover, web-based applications such as WebLogic's admin, operator, and monitor login are targeted with default passwords (0.4%). Furthermore, Microhard System's 3G/4G Cellular Ethernet and Serial Gateway is targeted with default credentials (5.5%) and the downloading of their configuration files, leading to information leakage [33].

3.1.4 Privilege Escalation. Approximately 3.1% of the alerts attempt at acquiring additional privileges, targeting different versions of SMB, web applications, and RDP remote code execution exploits. These attempts originate from 218 different countries, with a majority of them being highly localized. Figure 2 shows that most of the alerts originate from the United States, Vietnam, France, Britain, and Russia. More than 50% of the alerts come from only three countries, and >90% of the attacks originate from 36 countries.

The attackers utilize exploits to attain additional privileges. Among these, two SNMP vulnerabilities disclosed in 2002 are the most frequent (CVE-2002-0012 and CVE-2002-0012), followed by the SMBGhost vulnerability (CVE-2020-0796) affecting the SMB version 3, and the directory traversal vulnerability (CVE-2018-14847) in MikroTik RouterOS and remote code execution vulnerability (CVE-2020-14882) in Oracle WebLogic Server.

3.2 Applications

Investigating application-specific alerts provides information about their attack surface. We focus on web applications, RDP, SMB, and Telnet, and investigate the attack vectors being utilized by attackers. Geographically, we observe that web applications are mostly targeted from United States (76.4%), Canada (6.9%), and Netherlands (2.9%), as opposed to RDP (Russia (57.3%), France (17.7%), PL (6.3%)), SMB (Vietnam (13.9%), India (10.0%), Russia (9.3%)), and Telnet (Korea (31.6%), Taiwan (19.2%), United States (10.6%)), which see attacks originating from smaller countries (Figure 2).

3.2.1 Web Applications. Web application attacks involve common attack vectors such as exploits (SQL Injection, Cross-site Injecion). These attacks have been widely studied and have been shown to still be prevalent without much improvement over time despite the existence of many scanning tools today [18, 56]. Additionally, the rise of Internet-connected devices and the use of malware have made the situation worse.

We find that the web application alerts are spread over 22,856 ports. The majority of alerts target ports 8088 (92.5%), 7001 (1.6%), and 80 (1.2%), accumulating more than 95% of all alerts. Web applications make up for approximately 6.2% of all alerts, affected by multiple attack vectors, including malware (0.02%) and exploits (1.5%). 0.04% of the alerts arise due to use of weak credentials. Approximately, 1% of the alerts are remote code executions, 0.002% are brute force attempts, and 0.00001% of the alerts are related to RPC. 0.02% of the alerts are linked to PC or mobile malware, 0.003% are IoT malware-based alerts, and 0.001% are related to cryptojacking malware. Furthermore, 0.03% are target device firmware.

We also find that 1.5% alerts originate from blacklisted hosts and 0.9% are due to invalid communication (invalid header or request to hidden files). Additionally, 1.3% of the alerts are from blacklists, 0.1% from Internet scanners, and 0.004% from Tor exit nodes. Furthermore, other attacks are due to system path and bash shell commands in the URI or user-agent (81% alerts). Finally, the impact of the web app alerts lead to information leak (77.5%) and DoS attacks (0.0002%).

We find a targeted attack on Oracle Weblogic Server through two different attack vectors. The adversaries used two remote code execution vulnerabilities (CVE-2020-14882 or SMBGhost and CVE-2020-2551) in Weblogic servers that were exploited, and admin, monitor, and operator login attempts using default credentials. The default credential-based attacks started in November 2020 and continued until June 2021. However, only 3 subnets were responsible for >85% of the attempts.

We map alerts to OWASP top 10 risks that threaten web applications [60]. To do so, we (i) create heuristics to match them in alerts (enriched with information from NVD and exploitDB), and (ii) match the alert's vulnerability type to the top-10 risks [60]. Figure 3 shows that we successfully mapped seven out of the ten risks. We find that the 6^{th} and the 3^{rd} are most frequent risks in our dataset, accounting for more than 97% of the mapped alerts.

Geographically, web application are targeted from 171 countries. Seven countries are responsible for 94% of the alerts, with 76.3% of the alerts originating from the United States (Figure 2). Additionally, we find that three of the top five ASes belong to the DigitalOcean cloud service provider, accounting for about 51% of the alerts.

3.2.2 RDP. The Remote Desktop Protocol (RDP) allows users to access their computers virtually, having implications on data privacy. RDP is vulnerable to RCE vulnerabilities, such as CVE-2022-21893 [48, 62]. Russia, France, Poland, United States, and Latvia are the most frequent sources of attack.

Among all the alerts, 0.9% alerts target RDP using 14,015 ports, with top ports including: 3389 (99.6%), 3391 (0.001%), 3390 (0.001%), 3395 (0.0009%), and 3393 (0.0008%). Among the RDP alerts, 0.5% exploit vulnerabilities to gain additional privileges. Additionally, 0.05% of the alerts are from PC or mobile malware and 0.0003% from cryptojacking malware. However, other alerts involve connection requests as a user or administrator (0.5% of alerts). Moreover, 20.3% of the alerts are from Internet scanners, 1.3% have communications with invalid header, 0.03% are from Tor exit nodes, and 9.9% of the alerts are from blacklisted hosts. Finally, we see that 0.5% of the alerts lead to information leak (0.3%) and DoS attacks (0.2%).

Approximately, 62% of the RDP alerts originate from Russia (refer to Figure 2). We find that seven countries account for more than 90% of the alerts. Similarly, the top seven ASes lead to 63% of the alerts and five of them belong to Russia. Overall, we find that the alerts are centred around few organizations. For instance, we see that restricting the most frequently observed originating organizations would reduce the RDP alerts by 25%. Additionally, we find that for smaller countries, the alerts are even more centralized. For instance, 99% of the alerts from France originate from a single AS, belonging to IP Oleinichenko Denis.

Tunneling. Attackers create RDP tunnels over other ports to evade firewall and network controls. More than 81% alerts originate from Russia, and 78.2% of alerts originate from AS44477 (IP Oleinichenko Denis) and AS51036 (JSC Zenit Technology). Alarmingly, four class C subnets are responsible for *78.1%* of all alerts targeting RDP.

3.2.3 SMB. The Server Message Block (SMB) is a network protocol that facilitates the utilization of remote computers and servers. However, vulnerabilities, such as EternalBlue, SMBGhost, and SMBleed allow attackers to bypass authentication and spread throughout an affected network. These alerts originate from 229 countries, with Vietnam (13%) and India (10.0%) being the most frequent among them (refer to Figure 2). We find that alerts from Vietnam are concentrated to a handful of ASNs – 93.5% of the alerts originate from only four ASNs. Similarly, we see that 71.4% of alerts from Indonesia originate from only three organizations.

SMB alerts account for 6.6% of all alerts. Of these, 47.9% alerts are from exploits, including 37.8% remote code executions. Among the exploits, 10% of them are attributed to the EternalBlue exploit. Additionally, 0.00003% are tied to cryptojacking malware and the same proportion corresponds to PC or mobile malware. Other weaknesses include abusing null session behavior, lateral movement attempts, or use of powershell scripts over SMB. Additionally, 0.0001% of the alerts are due to network scanners and 0.005% are from Tor exit nodes. Around 7.5% of the alerts are due to improper header, such as invalid header and TCP length.

Apart from known vulnerabilities, the null session behavior also raises substantial alerts. These alerts target the administrative network shares[1]. The IPC$ share is used to administer network servers remotely. It creates pipes between the networked programs for communications. Accessing the IPC$ create a majority of SMB-based alerts, accounting for an impressive 50.4%. The rest of the administrative share accesses amount for 0.05% of RDP alerts.

Lateral Movement. The connection initiation attempts can be utilized to obtain information, resulting in informed attacks. Adversaries initiate SMB sessions with no username or password (Windows attempts implicit credentials [30]). This enables the attackers to propagate in the network, a process known as lateral movement. Additionally, powershell commands or scripts can be seen over SMB. The commands can be hidden or otherwise, without profile, or non-interactive. Such alerts make up for 4.3% of the RDP alerts. We also find alerts corresponding to Windows remote registry service (winreg). The service allows hosts to access the registry across a network. The winreg service is accessed through a specific named pipe (\PIPE\winreg). Particularly, we find 85 alerts corresponding to key creation. However, for an existing key, it opens the key, giving it access to all the keys and subkeys and values. However, functionally, it is similar to the IPC$ share alerts as the winreg can be accessed through IPC$.

3.2.4 Telnet. Teletype Network (Telnet) enables users to access the command line interface of a remote device/server and is commonly used for remote management, e.g., firmware upgrade. However, recently, it has been a target of IoT malware [7–9, 20, 50, 63]. The alerts originate from 209 countries. However, 76.7% of the alerts originate from Netherlands, and 90.5% of the alerts originate from only eight countries. Additionally, we see that 72% of the alerts originate from only one AS (Des Capital B.V. in Netherlands).

A little over 1% of all alerts target Telnet. 98.6% target ports 23 and 1.4% target port 9530. The attack vectors include: malware (2%), privilege gain (3.4%), credential-based (2.1%), and exploits (1.6%). Among the malware alerts, majority of them are attributed to IoT malware, and a very small fraction include cryptojacking (0.0002%) and traditional malware. Among the exploits, 1.4% include RCE.

We also find tunneled alerts targeting applications other than Telnet and other forms of flagged attempts, such as communication from blacklists. We find that 0.00004% (3 alerts) of alerts target RDP, 0.0002% (16 alerts) of the alerts correspond to Internet scanners, 0.04% of the alerts originate from Tor nodes, and 13.8% of the alerts are from blacklisted hosts. Additionally, approximately 74% of the alerts are tied to connections with invalid header.

The attacks targeting port 9530 (1.4%) hunt for devices using the Xiaongmai firmware. This firmware is used by smart devices, such as security cameras, DVRs, and NVRs around the world and allows Telnet communication over port 9530 [65]. The exploited backdoor (released in February 2020) allows the attacker to gain full control over the device through the root shell.

3.3 Impact

We focus our study on three specific impacts of the attack vectors on the applications. Table 1 marks the impacts that affect the applications and the attack vectors that help propel the impact.

3.3.1 Denial of Service Attack. Denial of service attacks result in a victim service such as a web server becoming unavailable. About 5% of all alerts are capable of DoS attacks. Table 1 lists the attack vectors that play a role in DoS attacks and the targeted applications. Particularly, the IoT and PC or mobile malware are the major attack vectors for DoS attacks. Vulnerability exploitations, RPC, and breaking into systems using weak passwords are also employed by the threat actors attempting DoS. The applications facing the majority of DoS attacks include Internet Relay Chat (IRC), SSH, RDP, and web applications. Of the alerts that result in DoS alerts, we find DoS attacks threaten web applications (0.0002%), Internet-connected devices (0.001%), and protocols such as RDP (0.06%) and IRC (0.1%). Additionally, we find that such attacks can be launched by leveraging exploits (99.58%), IoT mawlare (0.1%), and traditional malware (0.0001%). Among the exploits, a majority of them abuse the weak passwords (from a total of 99.63% of the DoS alerts), while only few use RPC (1 alert) and RCE (30 alerts). Additionally, 50.4% of the alerts result in Distributed Denial of Service (DDoS). Among

[1] Administrative shares are network shares in Windows NT that are not visible to non-administrators. These allow system administrators to access the disk volumes in a networked system remotely. Particularly, there are five default shares available — Root partitions or volumes (C$, D$, E$), The system root directory (ADMIN$), Fax share(FAX$), Print share (PRINT$), and inter-process communication (IPC$) [25]

Table 1: Impacts affecting the application and the attack vectors that influence the impact. IoT malware only cause DoS attacks, while other malware have larger impacts.

Impact	Web Apps	Device	Firmware	RDP	IRC	SSH	Traditional	IoT	Priv. Gain	Brute Force	RPC	Exploit	Credentials
DoS	✓	✓	✗	✓	✓	✗	✓	✓	✗	✗	✓	✓	✓
Info. Leak	✓	✓	✓	✗	✗	✓	✗	✗	✗	✓	✓	✓	✓
DNS	✗	✓	✗	✗	✗	✗	✓	✗	✓	✗	✗	✓	✗

the DDoS alerts, 19.9% alerts cause NTP reflection attacks, one is caused by the PC or mobile malware, 30.5% alerts are a result of IoT malware, and 0.03% target web applications.

3.3.2 Information Leak. Observed information leak alerts involve intentional probing to obtain information from devices which is not intentionally accessible through the Internet. For instance, an incoming connection may have a crafted User-Agent to scan SQL injection vulnerabilities. A large fraction (42.3%) of all the alerts result in information leak. Table 1 shows that information leak covers ≈11.5% web-based applications, and 0.04% of devices. The device-based alerts include reconnaissance attacks on AvTech DVRs, Vacron NVR, Hikivision IP camera, VoIP, and Netgear TeadyNAS Surveillance System to gather enough information before an actual attack. Additionally, we find that 3.65% of these alerts are reconnaissance attacks on SSH protocol via brute force SSH attacks. Moreover, we find that 11.9% of the alerts arise due to vulnerabilities, among which 11.8% also execute to additional privileges.

3.3.3 DNS. The DNS protocol helps identify services accessible through the Internet. DNS attacks can target a network with TCP/UDP floods, making the network components inaccessible through the Internet. DNS-based alerts constitute 0.03% of all alerts, of which 64.2% are a result of reconnaissance attacks, and 66.2% are due to vulnerability exploitations which include the 2% that operate with additional privileges. Additionally, 0.006% of the alerts arise from PC or mobile malware-based attacks.

Because information leak alerts make up more than 40% of all alerts, it is intuitive that they are caused by a wide variety of alert types, which might not be the same for every attack. However, we find that these attack alerts influence each other as well. For instance, we observe a high overlap between alerts that cause information leak and alerts that cause DoS. 99.6% of the DoS alerts cause information leak and 11.9% of the information leakage alerts cause DoS attack too. Similarly, 64.2% of the DNS alerts cause information leak, while they account for only 0.05% of information leak alerts. Additionally, 26% of the DNS alerts lead to DoS attacks and 17.1% of the DoS attacks are due to DNS attacks.

4 PERSISTENT AND NEW MALWARE TRENDS

In this section, we first discuss some observed persistent malware trends, including persistence of rogue networks identified more than a decade ago, and continuing exploitation of a set of 17 old vulnerabilities. Then we highlight some new trends on malware and exploits evolution.

4.1 Persistence of Known Malware Behavior

One of the most interesting findings of our work is the persistence of well-known malware behavior over time, suggesting that existing defenses such as blacklisting and threat intelligence sharing are not sufficient at eradicating known malware. We present here our results on the prevalence of known malware behavior after more than a decade.

Table 2: Overlap of rogue networks in our dataset and reported malicious ASes by the Fire system [59]. 17 ASes (71%) reported by Fire more than a decade ago are still active.

AS	Alerts	AS	Alerts	AS	Alerts
AS16276	3.9%	AS174	0.09%	AS10929	0.0002%
AS4134	0.5%	AS26496	0.09%	AS48031	0.0001%
AS4837	0.3%	AS28753	0.01%	AS3595	0.00003%
AS3265	0.2%	AS35908	0.003%	AS44050	0.000004%
AS4812	0.1%	AS27715	0.002%	AS41665	0.000001%
AS36351	0.1%	AS41075	0.002%		

Persistence of Rogue Networks. During our study, we find that the alerts are mostly centralized over geographies, and tend to be concentrated to certain subnets, IP addresses, and ASes, suggesting strong geographical localization of attack vectors. This finding is not novel, and has been reported in previous research. For instance, Stone *et al.* [59] proposed a system to expose organizations and ISPs that persistently involve themselves in malicious activities. The authors use a dataset of botnet C2 traffic, list of servers that host malware executable (drive-by-download), and phish hosting providers. Although the study was conducted 13 years ago and we cover a much wider threat landscape, we find that most of the ASes identified by their Fire system [59] are still active.

In particular, Fire identified and reported a set of 24 ASes involved in malicious activities. We match the ASes mentioned by their systems with the 23,063 ASes that originate the alerts in our dataset. Table 2 lists the overlapping ASes. We find that 71% (17) of the ASes mentioned in the Fire system are found in our dataset as well. We find one AS in the top five, four ASes in the top 100, and eight in the top 150 Ases by the fraction of alerts. This shows that despite the identification of these ASes as malicious more than a decade ago, they are still active and involved in malicious activities.

Persistence of Older Threats. We find that approximately 40.6M alerts (5% of all alerts) arise due to exploitation of vulnerabilities that were disclosed at least a decade ago (before 2011), and 13 vulnerabilities discovered as early as 1999 are still exploited. Despite patches being available, these vulnerabilities are still actively being exploited by numerous malware campaigns, such as PurpleFox, Netwire RAT, Gafgyt, APT 10, and Ransomware. This suggests that systems are not updated regularly and most basic security measures such as patching regularly are not being employed.

We find that 17 out of the 38 older vulnerabilities are exploited by malware campaigns. Table 3 summarizes these 17 vulnerabilities, their details, severities, and malware campaigns exploiting them. These re-defined threats target the SNMP protocol, RDP, firmware applications, TFTP protocol, and PHP-based applications. These 17 exploits utilize remote file inclusion, code or SQL injection, remote code executions, memory exhaustion, etc., to target the applications. We find that the sources involved in exploiting these vulnerabilities

Table 3: Decade-old vulnerabilities exploited by various malware campaigns. A majority of these exploits are exploited by Remote Access Trojans (RATs), such as AgentTesla, Netwire, and Android Spynote.

Vulnerability	Vulnerability Weakness	Product	Severity	Malware Campaign	# Alerts
CVE-1999-0517	Unauthorized Access	SNMP	High	Gafgyt, RATs, Cobalt Strike	43.4K
CVE-2002-0012/13	Privilege Escalation	SNMP	High		
CVE-2001-0540	Memory Exhaustion	RDP - Windows NT	Medium	Fileless, Cobalt Strike, Zeus	2K
CVE-2003-0818	Remote Command Execution	Windows NT 4.0, 2000, and XP	High	Emotet, Qakbot, Trickbot	83
CVE-2002-0953	Code Injection	PHP - PHP Address before 0.2f	High	RATs	43
CVE-2002-1149	Sensitive Information Leak	Invision Board	Medium	PurpleFox	23
CVE-1999-0152	Remote Command Execution	DG/UX finger daemon	High	RATs	8
CVE-2009-2765	Remote Command Execution	cgi-bin - DD-WRT 24 sp1	High	APT 10 - Cloud Hopper	6
CVE-2008-3022	Remote File Inclusion	PHP - PHPortal 1.2	High	RATs	2
CVE-2006-2009	Remote File Inclusion	PHP - phpMyAgenda 3.0 Final	High	RATs	2
CVE-2006-2149	Remote File Inclusion	PHP - Aardvark Topsites	Medium	RATs	2
CVE-2008-2649	Code Injection	PHP - DesktopOnNet 3	High	RATs	2
CVE-2010-0738	Improper Access Control	JBoss EAP	Medium	RATs	2
CVE-1999-0183	Path Traversal	TFTP - Linux	Medium	RATs	1
CVE-2009-0441	Remote File Inclusion	PHP - TECHNOTE 7.2	Medium	RATs	1
CVE-2008-6327	SQL Injection	PHP - ProQuiz 1.0	High	RATs	1

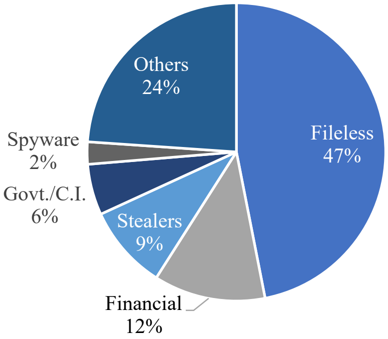


Figure 4: Campaign characteristics and their alert share.

are involved in more than one campaign. For example, the SQL injection vulnerability in ProQuiz's index.php, allows attackers to pass SQL commands with the password parameter (CVE-2008-6327), and even if it maps to only one alert, the source has been reported in multiple RAT campaigns, namely, AgentTesla, Netwire, Adwind, and Nanocore. However, the CVE-1999-0517, CVE-2002-0012, CVE-2002-0012, and CVE-2001-0540 vulnerabilities appear among the most popular exploits that have been exploited by actors, and each of them is involved in at least 10 campaigns.

4.2 New malware trends

We present some novel trends in malware evolution, including the use of fileless malware, the prevalence of stealers and keyloggers, and sharing of malware infrastructure across campaigns.

4.2.1 Campaign Attributes.

The 118 campaigns are mapped to six characteristics, namely: Banking, fileless malware, Stealers/Keyloggers, critical infrastructure, spyware, and government. Overall, of the 2.25 million alerts, we find that 74.3% can be placed in at least one of the six categories. Figure 4 shows the alert distribution of the malware campaigns. While most of the campaigns were attributed as Stealers/Keyloggers, it contributes to only 9% of the alerts.

Fileless Malware. While traditional malware aims to execute malicious functions on the victim machine, they are prone being detected by file-based malware detectors. This malware utilizes scripts that are run on applications, such as powershell.

We find 1.1 million fileless malware-based alerts. We map these alerts to campaigns, including PurpleFox, powershell and python injectors, Freakout, Crypto miners, and Magnetcore malware. However, 77% of the alerts are due to the Purplefox malware. Among the Purplefox alerts, 26.4% exploit the SMBGhost vulnerability. Apart from the SMBGhost exploit, the campaign targets the Eternalblue exploit, MSSQL, and SMB null-session vulnerability. The magnetcore campaigns exploits the decade old SNMP vulnerabilities and the Freakout campaign exploits an OS command injection bug in Realtek SDKs [37]. Additionally, the powershell injector launches brute force attacks on SSH, and exploits an RCE vulnerability in traffic management UI in BIG-IP, the crypto mining campaign exploits the Heartbleed, and Oracle WebLogic vulnerability.

Spyware. We identify one alert from FinSpy, a German-made spyware, targeting human rights defenders (HRDs) in Egypt, Bahrain, Ethiopia, and UAE. This alert arises due to an NTP reflection attack from a Russia-based IP address. Additionally, we identify seven alerts from sources identified for targeting the HRDs and journalists in Uzbekistan. These alerts launch a brute force attack on SSH, and originates from an IP address in the United States.

We also identify 37 campaigns involved in spying. Except for two, we find that all of the campaigns involve keylogging and stealing. Among the prominent campaigns are the RATs Luminousity, Netwire, Sectop, and Rms, and Android Spynote. We find 32 exploits being utilized by spyware, of which 11 do not have a CVE-ID.

Stealers and Keyloggers. We notice that 62.4% of the identified campaigns are Stealers or Keyloggers. Additionally, 50% of the Stealers/Keylogger campaigns are utilized by adversaries for spying. However, 73.6% of the alerts do not overlap with spyware alerts. A majority of these alerts belong to Android Roamingmantis, njRat, Magnetcore, and Zeus campaigns. This category of malware exploited seven vulnerabilities, with SMB being the biggest target.

Critical Infrastructure. We identified 130K alerts from campaigns that target the critical infrastructure (≈128K) and government infrastructure (1.6K). These alerts exploit five exploits, but they cumulatively make up for only nine alerts.

4.2.2 Strategy Mirroring. We find that malware families mirror infection strategies of their peers. Approximately 10K alerts reported for four campaigns imitate the strategies of two other campaigns.

Particularly, we identify 9.6K alerts for the ELF muBoT campaign, but our OSCTI framework identifies them to be a part of the ELF Freakout campaign. Additionally, we identify 338 alerts for the ELF Mirai campaign, which we find to be part of ELF Freakout, Gafgyt, Hajime, and Trickbot campaigns.

Additionally, we find evidence of shared infrastructure among malware campaigns. We see that 49 Class C subnets are being utilized for multiple malware campaigns. We identify one aggressive network that is involved in 21 campaigns (2.1K alerts). However, a majority of these subnets used multiple RATs to their benefit. For instance, we find subnets being used for RATs, Agent Tesla, Netwire, AdWind, Nanocore, and njRat. Additionally, we find 5.4K alerts raised by a single TOR IP address involved in two ransomware attacks, Kronos and Troldesh. These ransomware groups target financial institutions and critical infrastructure.

4.3 New Exploit Tends

We discuss new exploit trends observed from honeypot data analysis, including collaborative exploitation of victims, and rapid geographical spread of exploits.

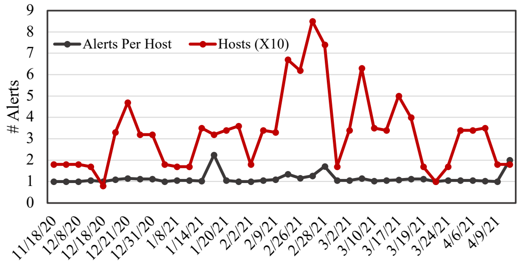


Figure 5: SMBGhost exploitation from a subnet.

Collaborative Exploitation. Among the most frequent exploits, we find that the SMBGhost vulnerability (CVE-2020-0796) was widely exploited. The vulnerability is mostly targeted from Vietnam (13.9%), Russia (9.6%), and India (7.7%). It was publicly disclosed on March 12th, 2020. The remote code execution vulnerability targets Microsoft Windows 10 with SMB protocol (version 3.1.1) and Microsoft Windows server 2016. The attacker connects to a target host and compresses the authentication request with a crafted offset field in the header. During decompression, it leads to buffer overflow, leading to the crashing of the target.

While SMBGhost was universally exploited, we find that 53% of alerts arise from subnets that use multiple hosts for attack. 71% of subnets only use one host each to launch 47% of the alerts. Among the 29% of the subnets, six use 100 or more hosts each. For example, the top subnet (S1) uses 254 of its 256 hosts. The subnet starts exploiting the vulnerability in mid-May and continues until early April, as can be seen in Figure 5. Another example is the subnet (S2) with most associated alerts – it uses 100 hosts to launch 89K alerts throughout our analysis window, with a daily average of 64.5 alerts per host, as opposed to S1's avearge of 1.1. We show S2's exploitation pattern in Figure 13 and discuss it in Appendix C.1.

Geographical Movement of Exploits. Understanding the spread of attack-vectors can help us plan their defenses. For example, a delayed inter-continental spread allows the other countries to prepare defenses against an upcoming attack-vector. This requires for us to focus on the vulnerabilities that were disclosed within our analysis time frame. However, considering the attack volume of the recently disclosed vulnerabilities, we only find the Oracle WebLogic vulnerability (CVE-2020-14882) as a suitable candidate (disclosed on Oct 21, 2020).

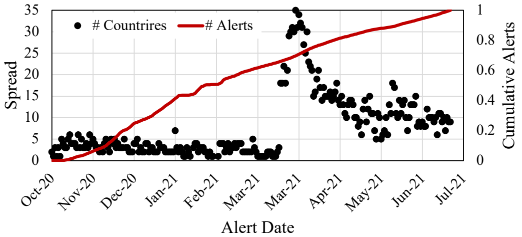


Figure 6: CVE-2020-14882 attack sources over time. The exploits originates from only five countries at the end of 4th day, 10 countries by 2 weeks, and 14 countries in a month. However, March 19th onwards sees increased spread, with March 19th alone adding seven new origins.

The vulnerability grant full control of the WebLogic Server to the attacker. There were in total 322K alerts in a span of 240 days, averaging 1343 exploits per day. Top attack sources include: Russia (73.2%), Lithuania (13.8%), China (2.9%), USA (2.2%) and Japan (1.8%). We observe that the vulnerability is exploited as part of multiple malware campaigns, such as sysrv_hello miner and Netwire RAT.

The attack was first observed on October 31st, 2020 from a Chinese IP address at 8:30 pm. About an hour and half later from its first appearance, we find the vulnerability being exploited from an IP address in Russia. By the third day, it was being exploited from Hong Kong, Germany, and Netherlands. Until the 18th of March 2021, the vulnerability was being exploited by at most five countries per day, with an average of 2.9 countries per day. However, by this date, it accounted for > 65% of all alerts, as can be seen in Figure 6. From March 19th, we observe that the it began being exploited by an average of 15 countries per day.

Overall, the vulnerability was exploited by 85 countries. However, after the fourth day from the first occurrence, we find that it was being exploited by only four countries. And a week later, it was still being exploited by only nine countries. We find that over time, the exploit volume increases with new sources joining in, however, after March 18th, we see a sudden increase in attack sources. We notice that 59 countries joined in within 18 days from March 18th. This suggests that cross-continental intelligence sharing can be effective in limiting the impact of exploits.

5 RELATED WORK

Honeypots have been widely leveraged to understand and investigate the threat landscape. However, the majority of these studies were conducted at least a decade ago. Table 4 lists the prior works that leverage honeypots towards different malware analysis objectives. The first six studies conducted before 2008 are generic, while more recent studies since 2017 are on IoT malware. Out of the six generic threat studies, four focus on the analysis of data collected by the *Leurre.com* project. Leita *et al.* [35] presented the data collection infrastructure of the *Leurre.com* project, launched in January 2003. Until March 2008, the honeypots were distributed in 28 different countries identifying ≈3.7 million sources of attack, with

Table 4: Related Work. After 2008, the focus of the literature have been on IoT-based threats.

Citation	Period	Size	Days	# Conn.	#Src	Scope
Leita *et al.* [35]	Jan '03 - Mar '08	70	-	5.2 M	3.5 M	Generic
Mcgrew [36]	2006	1	101	26.9K	-	Generic
Dacier *et al.* [23]	Sep '06 - Nov '08	40	800	3.5 M	2.5 M	Generic
Bloomfield *et al.* [17]	May '07 - July '07	150	62	2.9 M	-	Generic
Abbasi and Harris [2]	Sep '08 - Nov '08	1	60	34.3K	615	Generic
Thonnard and Dacier [61]	Sep '06 and Jan '08	44	486	-	1.7 M	Generic
Vervier and Shen [63]	Aug '17 - Feb '18	7	218	37.3 M	1.6 M	IoT
Zhang *et al.* [66]	Nov '18 - Dec '18	2	7	-	332	IoT
Metongnon and Sadre [38]	Sep '18 - Feb '19	15	127	68.0 M	2.4K	IoT
Peinert and Giset [51]	Mar '20 - Apr '20	3	28	486.2K	13.4K	IoT
This work	Jul '20 - Jun '21	662	365	7.0 B	7.0 M	Generic

United States, China, and Canada being the most frequent sources of attacks. Dacier *et al.* [23] proposed an attack attribution method for different attack phenomena, utilizing the attacks collected by the *Leurre.com* project from September 2006 to November 2008. They analyzed 800 days worth of data, accumulating 3.5 million connections. They reported that the NetBios and Windows DCOM Service are the most targeted services. Bloomfield *et al.* [17] compared different honeypots, including *Leurre.com* and studied their differences in interaction. They found that attacker traffic on the honeypots differed, with *Leurre.com* receiving the highest volume. Abbasi and Harris [2] deployed a Linux-based honeypot for 60 days in 2008 and found that attacks mostly targeted the SSH port (≈ 98% of probes). Mcgrew [36] deployed low- and high-interaction honeypots on a university network for 101 days in 2006. Ports 1433, 1025, and 80 were most targeted and Germany, United States, and Italy were the most frequent sources of attack. Thonnard and Dacier [61] analyzed 486 days worth of honeypot data collected by *Leurre.com*, and found a total of 1.74M actors. The rise of IoT-based attacks, such as the DoS attack on GitHub, resulted in the investigation of threats posed to IoT devices after 2017. Since 2017, this trend manifested into the utilization of honeypots to investigate IoT threats.

6 KEY INSIGHTS FROM THE STUDY

Our study is the largest study of honeypot-collected traffic to date (7 billion connections, 803 million alerts, 7 million sources), and is the first generic honeypot study after more than a decade. Compared to previous studies, we show the recent evolution of malware attacks, revealing some novel trends in the use of fileless malware, Spyware/Keyloggers, malware infrastructure sharing, and collaborative exploitation. Interestingly, we discover the persistence of rogue networks identified more than a decade ago, and the continuing exploitation of 17 vulnerabilities found between 1999 and 2009. Below, we summarize the key lessons that the reader can distill from this analysis.

Protecting against older threats. Surprisingly, rogue networks discovered more than a decade ago are still active on the Internet, and clearly, we need to develop better global cyber defenses to counteract these persistent adversaries. Traffic originating from ASes with known malicious activities should be blocked at different levels in an organization, and multiple threat intelligence sources should be used by firewalls, web proxies, and endpoint agents to protect against these well-known IP ranges that are certainly not trustworthy. Also, obviously, legacy systems should be updated with latest operating systems and software to prevent the exploitation of vulnerabilities discovered before 2009 – but our work shows that this is not always happening.

Protecting against new vulnerabilities. An interesting observation was the widespread presence of recently disclosed vulnerabilities as attack vectors. We find four CVEs disclosed in 2021 and 21 CVEs from 2020 were actively exploited – among these the SMBGhost (20.3 M alerts) and the Oracle Weblogic (322.3K alerts) vulnerabilities. Among the 31 vulnerabilities from 2021, 2020, and 2019, we find that all the vulnerabilities are remotely-exploitable, have low attack complexity, and require no user interaction. We find that the exploited vulnerabilities are categorized as Critical, High, or Medium severity on the Common Vulnerability Scoring System (CVSS). We can use these insights from our analysis to prioritize the patches to cover newly discovered exploits that have a higher chance of being actively used.

Addressing global threats. The spread of recent exploits over geography suggests that the initial spread might be slower, but at later stages, the attack spread accelerates. For example, by the end of the first week (2.5 weeks since disclosure), we notice that Oracle WebLogic vulnerability was limited to 4 countries and was limited to 7.67 alerts per day. Clearly, cross-continental intelligence sharing can help limit the impact of global threats. In 2015, a cyber intelligence sharing act was passed in the U.S. [49], but there is definitely more room for improvement here and similar global initiatives need to be started.

7 CONCLUSION

We investigate 806 million alerts generated from industry-scale globally distributed honeypots. Although honeypots have been leveraged to understand cyber threats, previous studies have been carried out more than a decade ago. In view of heightened Internet penetration in this decade, we investigate the threats posed to users in this evolved threat landscape. We begin by developing a framework for performing a high-level summarization of the alerts and then measuring the attack vectors. We then investigate the attack-vectors that threaten well known applications and the impact that they cause. We show the persistence of rogue networks that were identified in the past decade and the continuing exploitation of old vulnerabilities. Additionally, we find a geographical movement of exploits over time and collaborative efforts among adversaries to coordinate their campaigns.

Acknowledgement:. This work has been partially supported by NSF Grant 2127200.

REFERENCES

[1] Developers @ Rapid 7. 2022. An Introduction to Project Heisenberg. Available at [Online]: https://www.rapid7.com/research/project-heisenberg/. (2022).

[2] Fahim H Abbasi and RJ Harris. 2009. Experiences with a generation iii virtual honeynet. In *2009 Australasian Telecommunication Networks and Applications Conference (ATNAC)*. IEEE, 1–6.

[3] ISO 3166 Maintenance Agency. 2022. ISO 3166 Country Codes. Available at [Online]: https://www.iso.org/iso-3166-country-codes.html. (2022).

[4] Eric Alata, Vincent Nicomette, Mohamed Kaâniche, Marc Dacier, and Matthieu Herrb. 2006. Lessons learned from the deployment of a high-interaction honeypot. In *2006 Sixth European Dependable Computing Conference*. IEEE, 39–46.

[5] Mark Allman, Ethan Blanton, Vern Paxson, and Scott Shenker. 2006. Fighting coordinated attackers with cross-organizational information sharing. *IRVINE IS BURNING* 121 (2006).

[6] Omar Alrawi, Charles Lever, Kevin Valakuzhy, Ryan Court, Kevin Snow, Fabian Monrose, and Manos Antonakakis. 2021. The Circle Of Life: A Large-Scale Study of The IoT Malware Lifecycle. In *30th USENIX Security Symposium (USENIX Security 21)*.

[7] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai Botnet. In *USENIX Security Symposium*. 1093–1110.

[8] Afsah Anwar, Hisham Alasmary, Jeman Park, An Wang, Songqing Chen, and David Mohaisen. 2020. Statically Dissecting Internet of Things Malware: Analysis, Characterization, and Detection. In *International Conference on Information and Communications Security*. Springer, 443–461.

[9] Afsah Anwar, Jinchun Choi, Abdulrahman Alabduljabbar, Hisham Alasmary, Jeffrey Spaulding, An Wang, Songqing Chen, DaeHun Nyang, Amro Awad, and David Mohaisen. 2021. Understanding Internet of Things Malware by Analyzing Endpoints in their Static Artifacts. *arXiv preprint arXiv:2103.14217* (2021).

[10] Avast. 2022. EternalBlue Exploit | MS17-010 Explained | Avast. Available at [Online]: https://www.avast.com/c-eternalblue. (2022).

[11] Michael Bailey, Evan Cooke, Farnam Jahanian, Andrew Myrick, and Sushant Sinha. 2006. Practical darknet measurement. In *2006 40th Annual Conference on Information Sciences and Systems*. IEEE, 1496–1501.

[12] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, David Watson, and others. 2005. The internet motion sensor-a distributed blackhole monitoring system.. In *NDSS*. Citeseer.

[13] World Bank. 2021. The World Bank: Indicators. (2021). https://data.worldbank.org/indicator/IT.NET.USER.ZS.

[14] Ivan Belcic. 2022. The Zeus Trojan — What It Is, and How to Remove and Prevent it. Available at [Online]: https://www.avast.com/c-zeus. (2022).

[15] Greg Belding. 2020. Purple Fox malware: What it is, how it works and how to prevent it. Available at [Online] : https://resources.infosecinstitute.com/topic/purple-fox-malware-what-it-is-how-it-works-how-to-prevent-it/. (2020).

[16] Hugo LJ Bijmans, Tim M Booij, and Christian Doerr. 2019. Just the tip of the iceberg: Internet-scale exploitation of routers for cryptojacking. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*. 449–464.

[17] Robin Bloomfield, Ilir Gashi, Andrey Povyakalo, and Vladimir Stankovic. 2008. Comparison of empirical data from two honeynets and a distributed honeypot network. In *2008 19th International Symposium on Software Reliability Engineering (ISSRE)*. IEEE, 219–228.

[18] Ahmet Salih Buyukkayhan, Can Gemicioglu, Tobias Lauinger, Alina Oprea, William Robertson, and Engin Kirda. 2020. What's in an Exploit? An Empirical Analysis of Reflected Server {XSS} Exploitation Techniques. In *23rd International Symposium on Research in Attacks, Intrusions and Defenses (RAID 2020)*. 107–120.

[19] Catalin Cimpanu. 2019. New XBash malware combines ransomware, coinminer, botnet, and worm features in deadly combo. Available at [Online]: https://tinyurl.com/3d9wczsr. (2019).

[20] Emanuele Cozzi, Mariano Graziano, Yanick Fratantonio, and Davide Balzarotti. 2018. Understanding Linux Malware. In *IEEE Symposium on Security & Privacy*.

[21] CronUp. 2022. Malware: Repositorio de Indicadores de Compromiso. (2022). https://github.com/CronUP/Malware-IOCs.

[22] CyberMonitor. 2022. APT & Cybercriminals Campaign Collection. (2022). https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections.

[23] Marc Dacier, Van-Hau Pham, and Olivier Thonnard. 2009. The WOMBAT Attack Attribution method: some results. In *International Conference on Information Systems Security*. Springer, 19–37.

[24] Danielle Desfosses David Pany, Steve Miller. 2019. Bypassing Network Restrictions Through RDP Tunneling. Available at [Online]: https://www.mandiant.com/resources/bypassing-network-restrictions-through-rdp-tunneling. (2019).

[25] Microsoft Documentation. 2021. IPC$ share and null session behavior in Windows. Available at [Online]: https://docs.microsoft.com/en-us/troubleshoot/windows-server/networking/inter-process-communication-share-null-session. (2021).

[26] Executemalware. 2022. IOCs from malware investigations. (2022). https://github.com/executemalware/Malware-IOCs.

[27] Claude Fachkha, Elias Bou-Harb, Anastasis Keliris, Nasir D Memon, and Mustaque Ahamad. 2017. Internet-scale Probing of CPS: Inference, Characterization and Orchestration Analysis.. In *NDSS*.

[28] Geng Hong, Zhemin Yang, Sen Yang, Lei Zhang, Yuhong Nan, Zhibo Zhang, Min Yang, Yuan Zhang, Zhiyun Qian, and Haixin Duan. 2018. 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*. 1701–1713.

[29] Amnesty International. 2021. Forensic Methodology Report: How to catch NSO Group's Pegasus. (2021). https://www.amnesty.org/en/latest/research/2021/07/forensic-methodology-report-how-to-catch-nso-groups-pegasus/.

[30] James Kehr. 2020. SMB and Null Sessions: Why Your Pen Test is Probably Wrong. (2020). https://techcommunity.microsoft.com/t5/storage-at-microsoft/smb-and-null-sessions-why-your-pen-test-is-probably-wrong/ba-p/1185365.

[31] Brian Krebs. 2022. Conti Ransomware Group Diaries, Part II: The Office. (2022). https://krebsonsecurity.com/2022/03/conti-ransomware-group-diaries-part-ii-the-office/.

[32] KrebsOnSecurity. 2016. Hacked Cameras, DVRs Powered Today's Massive Internet Outage. Available at: https://krebsonsecurity.com/2016/10/hacked-cameras-dvrs-powered-todays-massive-internet-outage/. (2016).

[33] Gjoko 'LiquidWorm' Krstic. 2020. Microhard Systems 3G/4G Cellular Ethernet and Serial Gateway Configuration Download . Available at [Online]: https://www.exploit-db.com/exploits/45036. (2020).

[34] Don Ovid Ladores, Ian Kenefick, and Earle Maui Earnshaw. 2022. New Nokoyawa Ransomware Possibly Related to Hive. (2022). https://www.trendmicro.com/en_us/research/22/c/nokoyawa-ransomware-possibly-related-to-hive-.html

[35] Corrado Leita, VH Pham, Olivier Thonnard, E Ramirez-Silva, Fabian Pouget, Engin Kirda, and Marc Dacier. 2008. The leurre. com project: collecting internet threats information using a worldwide distributed honeynet. In *2008 WOMBAT Workshop on Information Security Threats Data Collection and Sharing*. IEEE, 40–57.

[36] Robert McGrew. 2006. Experiences with honeypot systems: Development, deployment, and analysis. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*, Vol. 9. IEEE, 220a–220a.

[37] Metasploit. 2015. Realtek SDK - Miniigd UPnP SOAP Command Execution. (2015). https://www.exploit-db.com/exploits/37169.

[38] Lionel Metongnon and Ramin Sadre. 2018. Beyond telnet: Prevalence of iot protocols in telescope and honeypot measurements. In *Proceedings of the 2018 workshop on traffic measurements for cybersecurity*. 21–26.

[39] MITRE. 2022. ATT&CK Matrix for Enterprise. (2022). https://attack.mitre.org/.

[40] Abedelaziz Mohaisen and Omar Alrawi. 2013. Unveiling Zeus: automated classification of malware samples. In *the 22nd International World Wide Web Conference, WWW*. 829–832.

[41] Elizabeth Montalbano. 2021. Purple Fox Malware Targets Windows Machines With New Worm Capabilities. Available at [Online] : https://threatpost.com/purple-fox-malware-windows-worm/164993/. (2021).

[42] Palo Alto Networks. 2022. PAN Unit 42 iocs: indicators related to Unit 42 Public Reports. (2022). https://github.com/pan-unit42/iocs.

[43] Lily Hay Newman. 2018. GitHub Survived the Biggest DDoS Attack Ever Recorded. Available at [Online] : https://www.wired.com/story/github-ddos-memcached/. (2018).

[44] NIST. 2022. National Vulnerability Database. (2022). https://nvd.nist.gov/.

[45] NVD. 2001. CVE-2001-0540 Detail. Available at [Online]: https://nvd.nist.gov/vuln/detail/CVE-2001-0540. (2001).

[46] NVD. 2022. CVE-2017-12615. Available at [Online]: https://nvd.nist.gov/vuln/detail/CVE-2017-12615. (2022).

[47] NVD. 2022. CVE-2017-9805. Available at [Online]: https://nvd.nist.gov/vuln/detail/CVE-2017-9805. (2022).

[48] NVD. 2022. CVE-2022-21893 Detail. Available at [Online]: https://nvd.nist.gov/vuln/detail/CVE-2022-21893. (2022).

[49] The Department of Justice. 2015. Cybersecurity Information Sharing ACT Of 2015 Procedures And Guidance. (2015). https://www.cisa.gov/publication/cybersecurity-information-sharing-act-2015-procedures-and-guidance.

[50] Yin Minn Pa Pa, Shogo Suzuki, Katsunari Yoshioka, Tsutomu Matsumoto, Takahiro Kasama, and Christian Rossow. 2016. IoTPOT: A Novel Honeypot for Revealing Current IoT Threats. *Journal of Information Processing* 24 (2016), 522–533.

[51] Trine Cecilia Peinert and Ingvild Bye Giset. 2020. *Analyzing the IoT Threat Landscape Within University Network Environments Using Honeypots*. Master's thesis. NTNU.

[52] Daniel Plohmann and Steffen Enders. 2022. Malpedia. (2022). https://malpedia.caad.fkie.fraunhofer.de/.

[53] ESET Research. 2022. Malware Indicators of Compromise. (2022). https://github.com/eset/malware-ioc.

[54] Christian Rossow. 2014. Amplification Hell: Revisiting Network Protocols for DDoS Abuse.. In *NDSS*. 1–15.

[55] Florian Roth. 2022. Signature-Base: YARA signature and IOC database for our scanners LOKI and THOR Lite. (2022). https://github.com/Neo23x0/signature-base.

[56] Theodoor Scholte, William Robertson, Davide Balzarotti, and Engin Kirda. 2012. An empirical analysis of input validation mechanisms in web applications and languages. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*. 1419–1426.

[57] Offensive Security. 2022. Exploit Database. Available at [Online]: https://www.exploit-db.com/. (2022).

[58] Joshua Shilko, Zach Riddle, Jennifer Brooks, Genevieve Stark, Adam Brunner, Kimberly Goody, and Jeremy Kennelly. 2021. FIN12: The Prolific Ransomware Intrusion Threat Actor That Has Aggressively Pursued Healthcare Targets. (2021). https://www.mandiant.com/resources/fin12-ransomware-intrusion-actor-pursuing-healthcare-targets.

[59] Brett Stone-Gross, Christopher Kruegel, Kevin Almeroth, Andreas Moser, and Engin Kirda. 2009. Fire: Finding rogue networks. In *2009 Annual Computer Security Applications Conference*. IEEE, 231–240.

[60] OWASP Top 10 team. 2022. Welcome to the OWASP Top 10 - 2021. (2022). https://owasp.org/Top10/.

[61] Olivier Thonnard and Marc Dacier. 2008. A framework for attack patterns' discovery in honeynet data. *digital investigation* 5 (2008), S128–S139.

[62] Lisa Vaas. 2022. Widespread, Easily Exploitable Windows RDP Bug Opens Users to Data Theft. Available at [Online]: https://tinyurl.com/2u88758p. (2022).

[63] Pierre-Antoine Vervier and Yun Shen. 2018. Before toasters rise up: A view into the emerging iot threat landscape. In *International Symposium on Research in Attacks, Intrusions, and Defenses*. Springer, 556–576.

[64] VT. 2022. VirusTotal Detection. Available at [Online]: https://www.virustotal.com/gui/ip-address/5.188.87.51/. (2022).

[65] YourChief. 2020. Full disclosure: 0day vulnerability (backdoor) in firmware for Xiaongmai-based DVRs, NVRs and IP cameras. (2020). https://habr.com/en/post/486856/.

[66] Weizhe Zhang, Bin Zhang, Ying Zhou, Hui He, and Zeyu Ding. 2019. An IoT honeynet based on multiport honeypots for capturing IoT attacks. *IEEE Internet of Things Journal* 7, 5 (2019), 3991–3999.

A DATASET DESCRIPTION

To understand the current threat behavior, we analyzed the traffic received by a group of honeypots. These low-to-medium interaction honeypots are deployed across geographies. The incoming connections are then analyzed to identify incoming attacks.

Each honeypot employs a set of rules which are applied against the interaction of the hosts with the attacker. Upon matching of a rule, it triggers an alert. The alerts have the following artifacts: (i) signature description describes why an alert was raised, (ii) category explains a high level categorization of the alert, (iii) reference lists additional information about the alert, such as vulnerability information (CVE identifier) and link to an exploit, (iv) attacker host, port number, country, Autonomous System (AS), and organization, and (v) targeted host and port number. Further, the signature description, category, and reference provides us the mean to annotate the techniques adopted by the adversaries (we describe the annotation process in section 2).

A.1 General Overview

A portion of the incoming connections to the honeypots are identified as alerts depending on their interaction. We focus our work on the identified alerts. Overall, our dataset accounts for a wide range of alerts identified over a span of 12 months — July 2020 to June 2021. Figure 7 show the alerts generated on each day during the 12 months period. The alert over time shows that alerts aren't highly skewed towards a specific time and also reflects at the continuity in alert identification. We observe an average of 2,208,216.425 alerts every day. The alerts range between 501,077 and 8,534,969, with the day with highest alerts making up for ≈1% of the alerts.

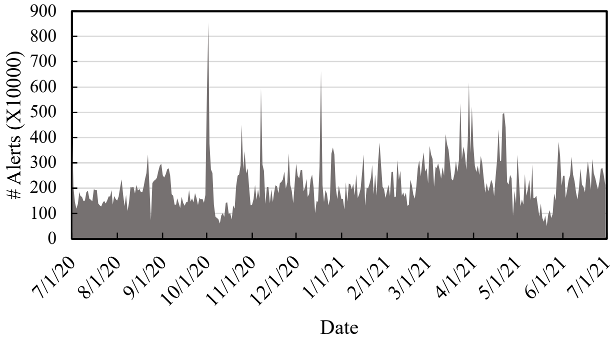


Figure 7: Distribution of alerts through the analysis times frame. The data is distributed through time.

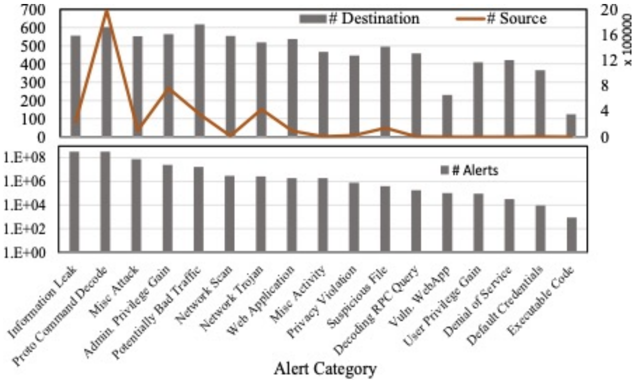


Figure 8: Alert categories and their size. The plot below shows the number of alerts in each of the alert categories. The alerts are highly skewed towards the information leak and protocol command decode categories. The Y-axis of the lower plot is therefore logarithmic (base 10). The upper figure plots the unique hosts and targets in each of the categories labelled in the shared X-axis.

The identified alerts range from vulnerability exploitations to the interactions originating from blacklisted IP addresses. These alerts are then grouped under high level categories defined by their intent. Figure 8 shows the alerts in each of the high level categories, with information leak being the most common category. Meaning that a vast majority of the alerts attempt for information leakage. Statistically, information leak alerts amount for 42.3% of overall alerts and protocol command decode amount for 41.1% of alerts, making for 83.4% of the alerts. The protocol command decode category include alerts, such as SYN resend with different sequence, while the information leak alerts include alerts such as a bruteforce attack on SSH.

A.2 Targeted Services

A quarter of the alerts (25.4%) target the File Transfer Protocol (FTP), with a total of around 204 Million alerts. These alerts include connection from a blacklisted IP address or an exploit. Figure 9 shows the most frequently targeted services. Additionally, figure 10

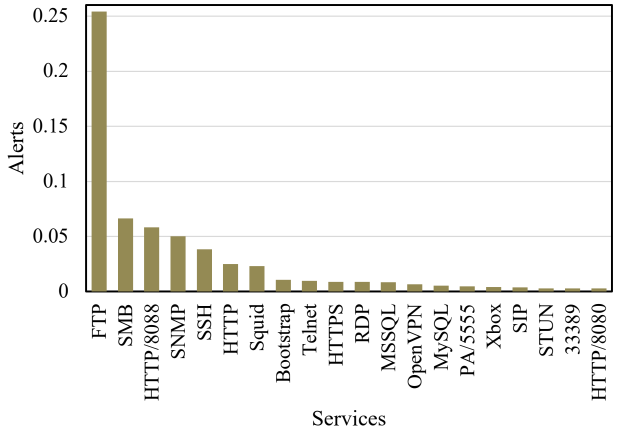


Figure 9: Frequently targeted services. The File Transfer Protocol is the most targeted service. We limit the plot to the 20 most frequent ones.

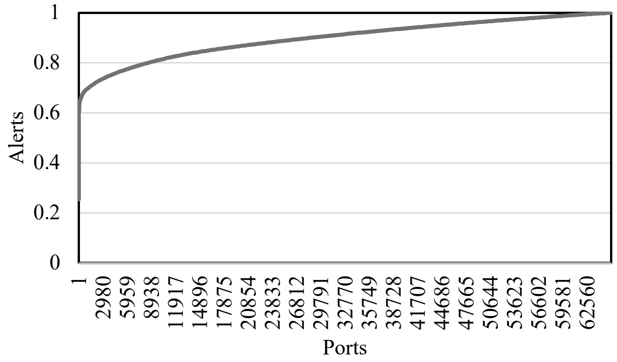


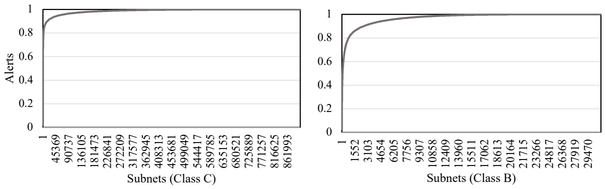
Figure 10: CDF of alerts on each port. We note that only 22 ports have more than 60% of the alerts, but the next 40% are distributed on 65514 ports.

shows the overall set of alerts and their corresponding alerts. We observe that the 20 most frequent ports make up for ≈ 60% of the alerts.

We previously noted that information leak and protocol decode alerts were most frequent and made up for ≈ 83.4% of all alerts. However, emphasised study of the services shows that they exhibit distinct profiles. For instance, we note that the SMB protocol faces the most network Trojans, making up for more than 87% of all network Trojan alerts. Similarly, we observe that only Telnet receives attacks involving the use of default user names and passwords. This has been observed the prior IoT malware based studies, and thus can be attributed to the increasing IoT malware-based attacks. We dig deeper into this cause-based study of the alerts in the next section.

Table 5: Most frequent countries of alert sources. The 20 listed countries, out of 237, make up for 88.4% of the alerts. Smaller countries have a significant share of alerts.

Country	Alerts (%)	Country	Alerts (%)	Country	Alerts (%)
USA	19.1 (23.6)	Hong Kong	1.8 (2.2)	Poland	0.9 (1.2)
Russia	16.2 (20.2)	Brazil	1.6 (2.0)	India	0.9 (1.2)
Germany	5.6 (7.0)	Canada	1.3 (1.7)	Bulgaria	0.9 (1.1)
Netherlans	4.6 (5.6)	S. Korea	1.3 (1.6)	Oman	0.9 (1.1)
France	4.4 (5.4)	Latvia	1.2 (1.5)	Vietnam	0.8 (1.0)
China	3.1 (3.8)	Ireland	1.2 (1.5)	Iran	0.7 (0.9)
Britain	2.8 (3.4)	Singapore	1.0 (1.3)	Others	10.1 (12.6)



(a) Subnet: Class B

(b) Subnet: Class C

Figure 11: Alerts originating from Class B and Class C subnets. Actors collaborate among themselves towards their intent.

A.3 Actor and Victim

The alerts are attributed to 2.8 million hosts distributed in 237 out of 250 ISO 3166-1 [3] countries. We call the alert origins as *actors* while the targeted are referred to as *victims*. These actors, however, often do not act independently. To investigate this collaborative nature among the actors, we map the hosts to the subnets that they belong to. Figure 11 shows that alerts per Class C and Class B subnets. In Figure 11a we see that 10% of the alerts from eight Class C subnets, while 50%, 75%, and 90% of the alerts originate from only 177, 1348, 15533 subnets. Similarly, Figure 11b shows that, 10%, 50%, 75%, and 90% of the alerts belong to only 5, 88, 532, and 2618 subnets. These show that the actors collaborate among themselves towards their motives.

Similar to the subnets, not all the countries have equal share of alerts. Table 5 shows the most frequent sources of alerts. United States (23.6%) and Russia (20.2%) account for ≈ 44% of all alerts, making them the largest actors. The third country in order is Germany, accounting for approximately one-third of the alerts originating from Russia.

We note that 10 countries make up for > 75% of the alerts and 24 countries are responsible for > 90% of the alerts. While it is somewhat expecyed for the United States and China to appear in list (given their large address space allocation), the other countries, such as France, Netherlands, Hong Kong, India, and Vietnam are surprisingly in the list.

We find that the alerts originate from 23,063 ASNs. Figure 12 shows the alerts covered by the ASNs. We have limited the ASNs to 1000 so that the cumulative growth is visible. We observe that 70% and 90% of the alerts originate from 104 and 358 ASNs, respectively. We focus on the smaller countries to identify the actors. For Ireland, we find that all the alerts originate from 51 ASNs, of which 12 ASNs originate >99.9% of alerts. Similarly, for Latvia, we observe 54 ASNs in total and 11 ASNs cause >99.9% of the alerts. Table 6 lists five frequent organizations in Ireland and Latvia.

We further analyze the alert sources in Ireland. More than 85% of the alerts originating from Ireland can be pinned down to three

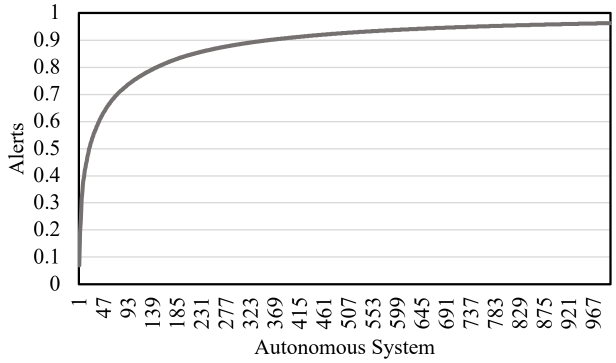


Figure 12: Proportion of alerts covered by ASes. Alerts corresponding to 1000 ASNs (96.3% of alerts) have been plotted.

Table 6: Top AS sources for smaller countries with significant alerts. Note that the alerts are concentrated to a very few ASes.

Ireland	Latvia
Global Layer B.V. (78.1%)	2 Cloud Ltd. (62.2%)
Amazon.com (12.8%)	Sia IT Services (34.2%)
Microsoft Corporation (5.2%)	Dedipath LLC (1.5%)
Digiweb Ltd. (2.5%)	Sia Nano IT (1.3%)
Liberty Global B.V. (0.5%)	Sia Tet (0.5%)

subnets. The most abused subnet generated ≈60.48% of alerts and 24 networks generated >98% of alerts. Manually searching for the actors belonging to these networked revealed that they have been involved in bruteforce attack on SSH in 2021 and have been identified as malicious by at least one vendor on VirusTotal [64].

B ALERT SUMMARIZATION

While ports have attributed to analyze targeted applications, they often miss on the tunneled abuse of applications. For example, a tunneled abuse of the RDP [24] through the SSH is categorized as a potential bad traffic.

Alerts may be due to multiple causes, which overlap across categories. We create heuristics to identify the causes of alerts and applications affected by the alerts. Table 7 shows the spread of causes across alert categories. For instance, we see that a wide range of alerts result from vulnerability exploitations. However, there seems to be no exploits in the "Executable Code" category – which seems counter-intuitive. Upon further investigation, we notice that the code executions, such as Remote Code Execution (RCE), require additional privileges, and thus have been categorized under user/admin privilege gain category. However, we see that the alerts summarized as remote codes also cover the alerts categorized as Executable Codes.

C MALWARE CAMPAIGNS

Through campaign inference, we identify alerts corresponding to malware campaigns, such as fileless malware and advanced persistent threats (APTs). Surprisingly, we also identify multiple spyware campaigns, including FinSpy, the German spyware targeting human right defenders in Uzbekistan. In this section, we discuss a few of these campaigns. It is important to note that we only focus on the initial infection strategies adopted by the campaigns. Additionally, we find hosts that are involved in multiple campaigns. Therefore, to focus on campaign-specific strategies, we limit our analysis to actors that are only involved in a single campaign.

Ransomware. Although, analyzing the deeper techniques adopted by threat actors is out of our scope, we identify 8.7K ransomware alerts, of which 1.9K are attributed to the Nemty ransomware group. Apart from being involved in web scanning activities, the Nemty ransomware is seen exploiting the Heartbleed vulnerability. Additionally, we find that almost 84% of the those alerts originate for Tor nodes. Moreover, ransomware groups, such as Conti and FIN12, use various attack vectors, such as Trickbot, Emotet, and Citrix exploits [31, 58], for their propagation. Overall, including the Emotet and Trickbot alerts, the ransomware attacks make up to 8.7K alerts.

Zeus. Zeus is an old malware that targets the sensitive personal information, such as banking credentials, which was first seen in 2007, but it has persisted over time [14, 40]. First, we see that all the alerts originate from Netherlands. It belongs to AS 202425 and hosted by *IP Volume Inc.* The malware targets victims through phishing, social media messages, and pay-per-install advertisements. However, we find the Zeus exploiting the null vulnerability to identify vulnerable NETBIOS, and then exploit the SMBGhost vulnerability (we will focus on the vulnerability in the next section). Additionally, we also observe traffic targeting the remote desktop service on non-standard ports until it finds the service. It then attempts to cause memory leak by exploiting CVE-2001-0540 [45].

Purplefox. Purplefox appeared in 2018, but rose to prominence in 2020. Prior independent efforts have shown the ability of this family to exploit different vulnerabilities, remote code executions (CVE-2020-0674) and privilege escalation bugs (CVE-2019-1458, CVE-2019-1458). However, they state that the infection begins with phishing and ads [15]. However, our study shows the presence of more offensive approach adopted by the malware. As such, it attempts at exploiting the vulnerabilities proactively to execute scripts that trigger the vulnerabilities.

We observe that the family exploits a wide range of vulnerabilities, such as SMBGhost, Eternalblue, PHP cgi query string vulnerability (CVE-2012-1823), and SMB null session vulnerability. Apart from exploits, we also see wide usage of port scanning and scanning for MSSQL, mySQL, and MS terminal server on different ports. Additionally, we identify brute force attempts on SSH and mySQL. We note that except for brute-force attempts targeting the SMB [41], none of the other strategies have been identified in the prior efforts.

C.1 SMBGhost Exploit

An interesting subnet (S2) exploits the SMBGhost vulnerability by using 100 out of its 256 allocated hosts. This subnet is singularly responsible for 44% (89K) of the SMBGhost-related alerts. As can be seen in Figure 13, on the first day it generates 554 alerts from eight hosts. Additionally, the network uses an average of a little over 4 hosts per day, with an average of 306.2 alerts per day. While subnet S1 displays characteristics of persistence, S2 presents itself as more

Table 7: Presence of our high level alert summaries in the different alert categories. We map the summaries with the categories that they are present in. For instance, Brute force alerts are present in the following categories, Protocol Command Decode, Privilege Gain, Information Leak, Web Applications, and Network Scans.

Alert Summary	Proto. C.	Misc	User/Admin Priv.	Info. Leak	Bad Traffic	Net. Trojan	WebApp	Net. Scan	Creds.	Exec. code
Exploit	✓	✓	✓	✓	✓	✓	✓	✓	✓	✗
Malware	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
Remote Code	✓	✓	✓	✓	✗	✓	✓	✓	✗	✓
Blacklist	✗	✓	✓	✗	✗	✓	✗	✗	✗	✗
Privilege Gain	✗	✗	✓	✗	✗	✗	✗	✗	✗	✗
Default Creds.	✗	✗	✓	✓	✗	✗	✗	✗	✓	✗
Bruteforce	✓	✗	✓	✓	✗	✗	✓	✓	✗	✗

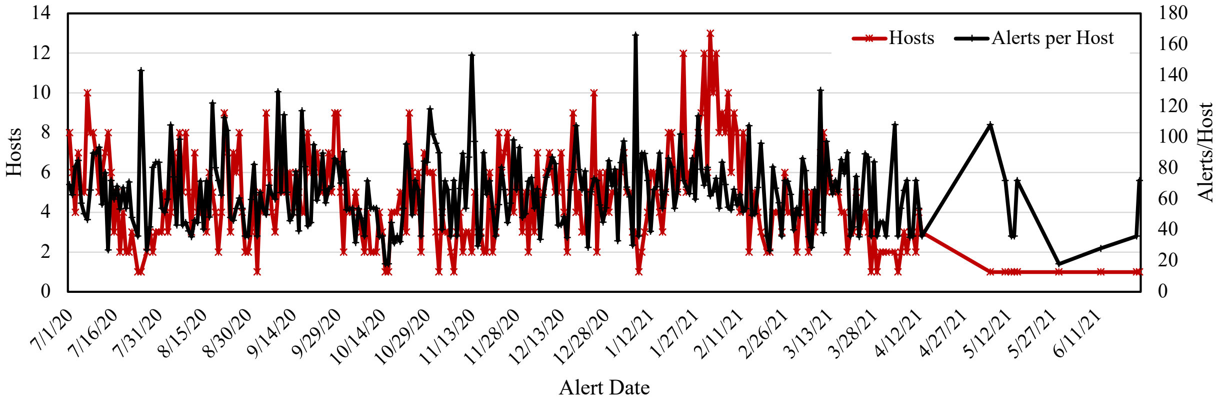


Figure 13: SMBGhost exploit: Subnet with most alert association.

Table 8: Top sources of malware alerts. The traditional and the IoT malware have overlapping origins, but Cryptojacking alerts have highly dissimilar origins.

Malware	Traditional	IoT	Cryptojacking
USA (34.1%)	USA (45.4%)	USA (25.7%)	Egypt (53.5%)
NL (17.4%)	NL (15.9%)	NL (18.4%)	Russia (17.4%)
Britain (6.3%)	Britain (11.5%)	China (9.7%)	S. Korea (4.4%)
China (6.1%)	Vietnam (3.7%)	Croatia (4.4%)	Luxembourg (3.0%)
India (3.2%)	Czech (2.5%)	India (4.1%)	Bulgaria (2.3%)

aggressive in exploiting the vulnerability continuously throughout our analysis window.

D LIMITATIONS

Our OSCTI framework covers a variety of sources, including, daily feeds, independent researchers, and security vendors. However, it is not an all-encompassing list of sources. We argue that the key insights shown in Section 6 hold true with the current sources. Additionally, adding new sources will only strengthen the insights.

The malware campaigns identified as mirroring strategies of others are not aliases. At the same time, it is impossible to disregard one or more of the shared threat intelligence reports to label IoCs to just one campaign. This is specifically difficult in the scenario when it has been shown that the adversaries utilize different families to achieve their stages of impact. Moreover, campaigns have been shown to share their distribution infrastructure [34].

Table 9: List of all exploited vulnerabilities identified by Suricata.

Exploited Vulnerabilities					
CVE-1999-0152	CVE-2002-0013	CVE-2010-0738	CVE-2017-12617	CVE-2019-10758	CVE-2020-26919
CVE-1999-0183	CVE-2002-0421	CVE-2010-1097	CVE-2017-12635	CVE-2019-11581	CVE-2020-3452
CVE-1999-0278	CVE-2002-0953	CVE-2010-3055	CVE-2017-18377	CVE-2019-12725	CVE-2020-5410
CVE-1999-0386	CVE-2002-1149	CVE-2012-0209	CVE-2017-5521	CVE-2019-12780	CVE-2020-5902
CVE-1999-0407	CVE-2002-1436	CVE-2012-1823	CVE-2017-5638	CVE-2019-1653	CVE-2020-6287
CVE-1999-0509	CVE-2003-0042	CVE-2013-0229	CVE-2017-8917	CVE-2019-16759	CVE-2020-7961
CVE-1999-0517	CVE-2003-0528	CVE-2013-2135	CVE-2017-9791	CVE-2019-16920	CVE-2020-8193
CVE-1999-0531	CVE-2003-0605	CVE-2013-3623	CVE-2017-9805	CVE-2019-19781	CVE-2020-8195
CVE-1999-0532	CVE-2003-0715	CVE-2013-3815	CVE-2018-1000861	CVE-2019-7256	CVE-2020-8196
CVE-1999-0736	CVE-2003-0818	CVE-2014-0160	CVE-2018-10561	CVE-2019-9621	CVE-2020-8209
CVE-1999-0737	CVE-2004-1776	CVE-2014-6271	CVE-2018-10562	CVE-2019-9978	CVE-2020-8515
CVE-1999-1376	CVE-2006-2009	CVE-2014-8361	CVE-2018-11776	CVE-2020-0796	CVE-2020-9054
CVE-1999-1538	CVE-2006-2149	CVE-2015-1427	CVE-2018-13379	CVE-2020-10148	CVE-2020-9484
CVE-2000-0071	CVE-2007-0631	CVE-2015-1635	CVE-2018-14847	CVE-2020-10204	CVE-2021-2109
CVE-2000-0126	CVE-2007-0676	CVE-2015-3337	CVE-2018-19276	CVE-2020-11651	CVE-2021-21978
CVE-2000-0630	CVE-2008-2639	CVE-2016-3088	CVE-2018-20841	CVE-2020-13942	CVE-2021-22986
CVE-2000-0778	CVE-2008-2649	CVE-2016-6563	CVE-2018-2628	CVE-2020-14181	CVE-2021-25646
CVE-2000-0868	CVE-2008-3022	CVE-2017-0143	CVE-2018-7600	CVE-2020-14882	CVE-2021-27561
CVE-2001-0540	CVE-2008-4250	CVE-2017-1000353	CVE-2018-9866	CVE-2020-15227	
CVE-2001-0876	CVE-2008-6347	CVE-2017-10271	CVE-2018-9995	CVE-2020-15505	
CVE-2001-0877	CVE-2009-0441	CVE-2017-12149	CVE-2019-0708	CVE-2020-1938	
CVE-2002-0012	CVE-2009-2765	CVE-2017-12615	CVE-2019-1003000	CVE-2020-2551	