Tracing Vendors: A Middlebox-Centric Study of Network Interference

Bulut Ulukapi
University of Twente
Enschede, The Netherlands
b.ulukapi@utwente.nl

Anna Sperotto
University of Twente
Enschede, The Netherlands
a.sperotto@utwente.nl

Ralph Holz
University of Münster & University of Twente
Münster, Germany & Enschede, The Netherlands
ralph.holz@uni-muenster.de

2025 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW) | 979-8-3315-9546-3/25/$31.00 ©2025 IEEE | DOI: 10.1109/EUROSPW67616.2025.00027

***Abstract*—Middleboxes are intermediary network devices that facilitate traffic monitoring, filtering, and modification. They serve a broad spectrum of functions, ranging from benign tasks to highly controversial ones such as censorship. A solid body of work exists that describes methods to probe or identify middleboxes from remote including censorship middleboxes; similarly, much research has gone into fingerprinting network devices. However, there is comparatively little work that aims to understand which type of devices occurs in which networks. In this study, we choose to investigate middleboxes that reside in networks reported for network interference. We use yarrpbox, a scanning tool, to detect middleboxes and map them to vendors utilizing third-party datasets.**

Covering more than 500 Autonomous Systems reported for interference, we identify about 250 middleboxes, which we study in detail. We find that the location of middleboxes across countries does not correlate to the Internet Freedom Index, and we identify a distribution of vendors as well as a distribution across countries that differs markedly from previous reports. Most middleboxes in the reported networks are actually likely to serve multiple purposes, and this complexity calls for new measurement methodologies to determine whether the reported interference is a byproduct of some configuration or the primary purpose of a middlebox. We also identify a number of security issues in a number of devices, lending further support for the hypothesis that middleboxes can increase the attack surface of a network. We conclude with a discussion of directions to understand middlebox deployment with further measurements.

***Index Terms*—Middleboxes, Network Interference, Vendor Mapping, Network Analysis**

1. Introduction

The Internet's unforeseen growth and integration into many aspects of our lives resulted in a complex landscape. Middleboxes have been deployed into networks to manage this intricate ecosystem. These devices perform advanced functions for security enforcement, traffic optimization, performance, and policy control. Thus, they have become fundamental to modern network operations. However, middleboxes are also widely employed for network interference, including deep packet inspection (DPI), traffic filtering, and connection manipulation. Some interference practices are of concern, *e.g.,* when they infringe on privacy or constitute censorship that violates human rights.

Much prior research focused on detecting and characterizing middlebox *interference*, for example by analyzing packet modifications and investigating protocol deviations [1]–[4]. Another body of work focused on *device fingerprinting* [5]–[7]. However, there remains a gap in research that combines these two forms of investigations into middleboxes. This study is a first step to bridge this gap: we leverage active probing techniques to detect middlebox interference and correlate these findings with vendor-specific fingerprints extracted from data sources such as Censys and Shodan. We thereby investigate the feasibility of middlebox vendor attribution and shed light on their distribution, deployment patterns, and potential security risks in network environments associated with interference.

We take a a slightly different approach than previous work. We investigate the paths to target networks reported by OONI or Censored Planet as likely interfering with their users' traffic [8], [9]. The rationale, in short, is that one can expect a significantly higher probability of finding middleboxes being involved in network interference. Note that our definition of middleboxes is not limited to those performing censorship. Our methodology consists of active probing using Yarrpbox, a high-speed stateless scanning tool, to detect the IP addresses that interfere with end-to-end connections. We then use vendor data obtained from Censys [10] and Shodan [11] to identify the devices we detected. We also report on their geographic distribution, presence across various Autonomous Systems (ASes), as well as potential vulnerabilities that have been reported for a given product. Our findings indicate that US-based companies dominate the middlebox market, with Check Point emerging as the leading vendor. We found that the overall presence of middleboxes did not strongly correlate with a country's Internet Freedom Index, suggesting that many middleboxes are deployed possibly for benign purposes such as performance optimization. We also observed a trend towards unified middleboxes with combined functionalities, such as next-generation firewalls (NGFWs), which blurs the line between different types of middleboxes and complicates their classification. Finally,

DOI 10.1109/EuroSPW67616.2025.00027

our vulnerability analysis suggests that most security issues are associated with outdated software versions, legacy protocols, and misconfigurations; however, no verified vulnerabilities were identified among the devices examined.

The remainder of the paper is organized as follows. Section 2 gives background information and related work. Section 3 details our data acquisition methodology and the overall workflow of the study, presents an overview of our datasets, and explain the limitations of our study. Then, we present our results in Section 4, including a vulnerability analysis. Finally, Section 5 provides a discussion of the results.

2. Background and Related Work

Middleboxes are specialized, intermediary devices that perform advanced functions beyond mere packet forwarding. Unlike traditional routers and switches, middleboxes can actively engage in inspecting, filtering, and modifying network traffic for various purposes such as performance optimization, protocol translation, censorship and security [12]. The term hence refers to a broad range of devices, including firewalls, load balancers, proxies, intrusion detection/prevention systems (IDS/IPS), WAN optimizers, network address translators (NATs), and deep packet inspection (DPI) devices.

Middlebox detection and analysis. Researchers have explored various methodologies to detect and analyze middleboxes and characterize their activities. Some of the earliest approaches focused on TCP behavior [1], [2], [13]–[17]. These studies demonstrated that middleboxes can block or modify TCP extensions, disrupt congestion control, and cause compliance issues with the TCP protocol, thereby degrading performance and hindering the adoption of new TCP features. They also documented how middleboxes constrain the evolution of TCP [15], [18].

In a different fashion, a number of tools focus on traffic other than TCP. The authors of [19] investigate changes on web pages via HTTP, reporting that over 1% of web clients received altered pages due to modifications attributed to ISPs, proxies, and malware. They also provided the parties likely responsible for these modifications, including a partial vendor attribution. Kreibich *et al.* [20] proposed Netalyzr to analyze networking inconsistencies, including the ones caused by middleboxes, and presented the identified systematic problems including fragmentation challenges, unreliable path MTU discovery, restrictions on DNSSEC deployment, and deliberate manipulations of DNS results. Sundara *et al.* [21] investigate connection tampering by analyzing Cloudflare's traffic to identify comprehensive tampering signatures that reveal the real-world impact on users worldwide. The authors of [22] introduced Tracebox, an extension of traceroute that detects middleboxes along a path. The tool achieves this by comparing the outgoing packet with the ICMP message returned at each hop, which often contains a (partial) quote of the original TCP segment. This allows to reveal modifications such as TTL changes, header rewriting, or payload alterations. Several follow-up studies investigated middleboxes in more detail. Thirion *et al.* [23] introduced `traceboxandroid`, a tool that tracks middleboxes in mobile networks, demonstrating their significant presence and impact on traffic behavior. Zullo *et al.* [24], [25] extend this work by employing smart traceroute techniques to reveal address translation and proxy phenomena, thereby exposing middlebox behaviors in mobile environments.

A more recent tool by Hilal and Gasser [26] is Yarrpbox, a high-speed network scanner built for large-scale middlebox detection. Overcoming the limitations of Tracebox by employing a stateless architecture and randomized probing, Yarrpbox otherwise builds on Tracebox's techniques. It detects middleboxes at scale by sending specially crafted packets to target IP addresses and comparing the resulting responses. Its speed makes it a suitable choice for our study to identify devices associated with network interference.

Device fingerprinting. Numerous studies present vendor fingerprinting techniques for network devices. Examples include [27], which aims at router signatures generation by analyzing protocol-specific TTL variations, and [7], which relies on SNMPv3 responses. A further study [28] builds on [7] to identify vendors with minimal probing overhead. Finally, the authors of [29] present an active measurement framework for identifying censorship middleboxes and their vendors.

Commercial endeavors such as Censys and Shodan have emerged as well. Among other things, they offer data for vendor attribution at Internet scale. For the fingerprinting stage of our study, we rely on the data that the latter two provide under their academic licences. Although relatively little is known about the precise methodologies they employ for vendor identification, their broad coverage and accessibility make them suitable choices for our analysis.

Censorship platforms. Internet Censorship, which is often implemented with middleboxes, is a complex problem to measure at scale. Two major projects that attempt to do this are the Open Observatory of Network Interference (OONI) [30] and Censored Planet [31]. OONI collects and analyzes censorship, surveillance, and traffic manipulation data from user-run probes, using tests for blocked websites, throttled services, and middlebox interference. Censored Planet is an automated censorship measurement system that continuously monitors network interference in over 200 countries. Unlike OONI, it does not rely on user-supplied data but performs server-side and client-side testing to detect DNS manipulation, IP blocking, and HTTP interference. In our work, we use data from both OONI and Censored Planet to retrieve reports of network prefixes that are possibly associated with network interference. It is known from literature [32] that censorship measurement data is subject to limitations, including incomplete coverage, unreliable metadata, and unexpected interferences which require careful consideration of several factors such as geoblocking practices and potential network fail-

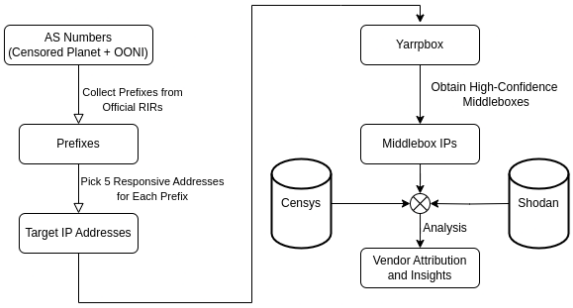

Figure 1. Overall Workflow of the Study

ures. We acknowledge that some of the prefixes that we determine may not actually engage in censorship or network interference. However, the overall sound techniques of these projects make these prefixes good starting points.

3. Methodology

We describe the methodology used to collect the data for our study. We provide an overview of our workflow in Fig. 1, including the steps involved in identifying target IP addresses, detecting middleboxes, and fingerprinting middlebox vendors. Our methodology combines active probing and existing datasets to identify middleboxes within networks previously reported for network interference. This study represents an initial attempt to assess the feasibility of vendor attribution and categorization of middleboxes with an aim of having a deeper understanding of their deployment contexts and functional roles.

IP Target Set and Middlebox Detection. We first identify a target set of IP addresses. In this study, we focus on IPv4 addresses only. We choose our targets by collecting Autonomous Systems (ASes) that have been reported for network interference by OONI or Censored Planet. Our rationale is two-fold here. First, given the nature of these data sets, we assume that the reported ASes have a significantly higher probability of being involved in network interference, and hence we should be able to detect middleboxes on the paths to these ASes or inside them. Second, we assume that traffic leaving these networks will go across the same border devices as traffic entering these networks, at least most of the time. In other words, we assume that ASes where traffic is entirely asymmetric, *i.e.,* it always enters via one path but leaves via another, are rare. This assumption is supported by previous studies that established the prevalence of asymmetric routes in general [33], [34]. In particular, the authors of [34] found that paths tend to be symmetric near the endpoint of a connection and more likely to by asymmetric in the middle. An exception may be IXPs, where a recent study found strong asymmetry in a few IXPs [35]. Hence, while our assumption rests on evidence, our numbers still constitute a lower bound.

Prefixes associated with interference. From OONI, we collect AS numbers (ASNs) where the measurements have been tagged "confirmed" to indicate high confidence in interference occurring on the path. We use the measurements from the "web_connectivity" tests, as these are numerous and have the most confirmed cases. We collect all reports for 2024, resulting in 323 ASNs. Censored Planet does not use such tagging. Therefore, we look for strong indicators of network interference. These are a high number of anomalies (timeouts and content mismatches), TCP reset packets, and known blockpages. We also gathered the data covering 2024 and arrive at 195 further ASNs from Censored Planet. We use `bgpq4` [36] to identify the prefixes for each ASN, using data from Regional Internet Registries (RIRs). We use the ARIN, LACNIC, RIPE, APNIC, and AFRINIC databases. Note that this constitutes another limitation of our study as this mapping occurs after the interference reports, and it is possible that some IP prefixes have changed owners in the meantime.

Middlebox detection on the paths. To identify target IP addresses in the target prefixes, we use *zmap* scans on port 80 to identify responsive IP addresses—this port is one of the most frequently used on the Internet. For each prefix, we randomly select up to five responsive IPs. Where our scans yield fewer than five response IP addresses, we add randomly chosen unique IPs from the same subnet to arrive at five target IP addresses.

For the middlebox detection, we use *yarrpbox*. *yarrpbox* can detect a solid number of traffic modifications while leveraging a stateless, parallelized probing architecture. This enhances scanning speed and effectively circumvents detection-triggered defenses, such as ICMP rate limiting. To identify where on the path a traffic modification occurs, *yarrpbox* uses the message parts that are quoted in ICMP replies. We refer the reader to [26] for details. All scans were carried out on a single machine located within our university in Europe. We empirically evaluated the performance of *yarrpbox* to find optimal scanning parameters, ultimately running the tool with a scan rate of 5 kpps. This moderate rate helps ensure that most probes reach their destination and the responses are reliably captured. Moreover, this rate is low enough to evade overly aggressive blocking or filtering. Given our relatively small pool of target IP addresses, this rate leads to a scan being completed in approximately one hour. We performed our scan on 3 February 2025. We used the TCP SYN option for the scan and set the scanned port to 80. During the preliminary trial scans, we experimented with various port numbers, including 22, 80, 443, and 500. Among these, port 80 yielded the highest number of detected middleboxes. Consequently, we selected port 80 for our comprehensive scan due to its greater likelihood of eliciting middlebox responses.

We found a bug in *yarrpbox*'s analysis tool, which led to the tool mistaking some destination IPs for middleboxes. We reported this to the developers. We check our final dataset for middlebox candidates. Similarly to [26], we employ a "location error" metric to classify middlebox IPs by confidence level. This metric measures the distance from a previous replying hop that quotes at least as much of the original packet

as the candidate middlebox IP. By disregarding prior hops that lack adequate ICMP quoting, we can more effectively pinpoint middlebox interference to a location on the path. We only consider the so-called high-confidence middlebox IPs, *i.e.,* those with location error of at most 1. In the following, we always refer to these when we speak simply of "middleboxes".

Fingerprinting Middlebox Vendors. We use data from Censys and Shodan to determine the vendors of the detected middleboxes. For Shodan, we query the API for the IP address of the presumed middlebox. For Censys, we process a snapshot from 10 September 2024, which was the latest available to us at the time of writing. Since middleboxes can be assumed to be long-lived devices with static IPs, this should not impact our vendor attribution too much.

We aggregate the extracted information, which includes geographic location, open ports and services, service banners, software and operating system vendors, and the respective labels provided by Shodan and API. We further refine the analysis by examining patterns in open ports, and service banners. We give a summary in the appendix A.

When an IP address appears in both Censys and Shodan, we evaluate the consistency between the two. In case of a discrepancy or data being older than two weeks, we perform a manual verification that the hosts are still online (sending ICMP pings) and the reported ports are open (quick port scans). In cases where information is only available from Shodan, we assess the "last update" date. If the date is no older than two weeks, we record the provided data. Most of the data for our middlebox IPs are at most one week old.

Limitations. We documented the assumptions we make above. However, our study still has some further limitations. First, even though a middlebox may be on the same path segment that inbound and outbound traffic takes, it may still only react to outgoing traffic. This means that our method from probing from remote and hence sending incoming traffic may not trigger such middleboxes, and we hence would undercount. Another limitation concerns the choice of target IPs. We choose only five of these. It is possible that networks have further internal segmentation beyond the routable prefix, and a middlebox may only sit on one of the paths to these. We would also miss this. Naturally, if a middlebox does not manipulate traffic in any of the ways that *yarrpbox* detects for TCP and IP, we also miss it.

Concerning vendor attribution, our analysis is based on data from Censys and Shodan, which means we share any limitations their scanning methodologies have. We acknowledge that, in certain cases, the information retrieved from the aforementioned data sources may lack sufficient accuracy to reliably establish one-to-one mappings between the scanned IP addresses and the corresponding data obtained.

Datasets. We summarize the characteristics of our IP targets in Table 1. From OONI and Censored Planet, we collect 518 ASes with ∼380k prefixes. Mapping the target ASNs to countries using `bgpq4` [36], we observe that the prefixes span over 120 different countries. We choose a total of ≈2M target IPs using the method described above.

TABLE 1. Overview of Target Set

	# of ASN	# of Prefix	# of IP
OONI	323	174105	869240
Censored Planet	195	204533	1128568
Total	**518**	**378638**	**1997808**

TABLE 2. Overall Scan Statistics. MB = middlebox, HC = high confidence.

Replies	Traces	Modifications	MB	HC MB
21.8M	2M	20847	1289	258

The overall statistics of our scan can be found in Table 2. Our scan results in 21.8 million replies from 2 million traces. We observe 20 847 modifications in total (for details on the modifications, see Table 8 in the appendix. We obtain in 1289 potential middlebox IPs, of which just 258 are classified as high-confidence.

Ethics. In our measurements, we adhered to best practices as described in [37] by limiting the probing rate and employing a well-established blocklist. We refrained from conducting large-scale scans, and all packets were TCP-compliant to ensure network stability. Additionally, we utilized a dedicated server with an informative reverse DNS name and maintained a website that detailed our measurement process. Contact information was provided to facilitate contact with us in case of any issues or opt-out requests. Throughout the measurement period, we did not receive any complaints or requests for inclusion in our blocklist.

4. Results

We present the results of our analysis of the detected high-confidence middleboxes (hereafter simply called "middleboxes") along with a vendor attribution. We also provide insights into the distribution of middleboxes across different countries and ASes.

4.1. Distribution of Middleboxes

We conduct an analysis to understand the distribution patterns, locations, and potential purposes of middleboxes.

4.1.1. Path-Level Analysis. We determine the distribution of middleboxes on the way to a target address. We achieve this by identifying the ASes of middleboxes and examining the traceroute outputs collected during detection. The majority of them (74.4%) are located in the same AS as the target address. This is in line with studies that focused on general middlebox deployment [18], [26]. Moreover, 90.6% of these middleboxes are located at most 3 hops away from the target address. The remaining 24.6% of detected middleboxes (66 IPs) are found between the source and target ASes. Upon further investigation of the traceroute outputs of middleboxes that are

found between the source and the target AS, 72.7% of those were detected at most 3 hops away from the target address. Overall, this suggests that most middleboxes are deployed close to the target address, which would allow to use them for traffic optimization, performance, security enforcement, and content filtering purposes. Moreover, 18,2% of the middleboxes detected in transit ASes (12 IPs) appear as the last hop in traceroute results. Upon querying those IPs on Censys and Shodan, we find relevant information for six of those middleboxes. Three of the devices are CheckPoint firewalls with combined functionalities including remote access. Three other devices appear to have only one open port each, either 123 (NTP), 161 (SNMP), and 179 (BGP). We speculate the remaining unidentified devices may be reverse proxies or load balancers, which make them the last visible hop. However, we could not confirm this with our approach.

We investigate how many of our scanned paths are affected by at least one middlebox modification. We find that 17% of the paths have at least one modification (including trivial ones such as Type of Service). This is higher than the approximately 10% reported in previous studies [18], [26]. This increase supports our hypothesis that we are more likely to find middleboxes near networks that have been reported for network interference.

4.1.2. Country-Level Analysis. We geolocate the middleboxes across different countries using IP2Location's DB23 [38]. We find at least one middlebox in at least one AS in 51 countries. We calculate the middlebox density for each country by dividing the number of middleboxes by the total number of IP prefixes in that country. Table 3 ranks the top ten countries by their middlebox density together with the Internet Freedom Index and Score. The index is based on the Freedom on the Net 2024 report [39] by Freedom House, which assesses the level of Internet freedom in countries worldwide. The index ranges from 0 (least free) to 100 (most free), with countries classified as Free, Partly Free, or Not Free.

We find no correlation between middlebox density and a country's Freedom Index. South Africa has the highest density of middleboxes (0.25); it is classified as Free. Then several countries follow with similar densities ranges between 0.03 and 0.05. We also see similar densities for Uganda, Australia, and Rwanda, which have been classified as Partly Free, Free, and Not Free, respectively. We see that even the countries that are classified differently may have similar middlebox densities. At the same time, countries with a low Internet Freedom Score like Uganda and Rwanda have a very low middlebox density of 0.01, but this could also be influenced by the low Internet penetration rates for those countries (as documented on the ITU Data Hub (https://datahub.itu.int). We should also note that the Internet Freedom Index is not available for all countries, and some countries such as Belgium, Portugal, Mongolia, and Sweden are not indexed. However, they have Global Freedom Scores (96, 96, 84, and 99, respectively) which is also published by Freedom House. Upon checking the Global Freedom Scores and Internet Freedom Scores of the countries in our top 10 country list, we see a correlation among scores in almost all cases except Uganda. As a result, we supplement our analysis with Global Freedom Scores where Internet-specific metrics are missing.

Our findings indicate that likelihood of censorship is not a determinant for the existence of middleboxes, and vice versa. We should mention this outcome may be affected by several factors: censors might deliberately obscure or mask the behavior of deployed devices to avoid detection, and many middleboxes may serve non-censoring functions such as traffic engineering or performance optimization. We also note that countries with limited freedom may deploy fewer observable middleboxes for many networks, or alternatively rely on other censorship or surveillance techniques not captured by our particular setup. We provide the top 10 countries, based on the number of detected middleboxes, in Table 6.

4.1.3. AS-Level Analysis. We analyze the distribution of middleboxes across target ASes. We use Team Cymru's IP to ASN Mapping Service [40] to determine the country of registration for these ASes as well. Table 4 presents the top ten ASes with the highest density for middleboxes. The ASes with the highest densities are MMS-AS-ID(AS38320) and JHU(AS5723), which are registered in Indonesia and the United States, respectively. They host only one detected device, yet have a density of 0.13——meaning that over ten percent of their prefixes are covered by these middleboxes.

When we compare the country of registration of the ASes with the highest middlebox density to the countries with the highest middlebox density (based on our geolocation method), there is no obvious correlation between the two. It is common that larger ASes have IPs outside their country of registrations. Only Belgium, Australia, and Portugal occur in both lists. Moreover, we can see that all the detected middleboxes reside in the same AS (ISEEK-AS-AP) for Australia.

We offer two main observations. First, higher middlebox densities occur across a wide geographical range, underscoring that middlebox deployment is not isolated to a single region or specific countries. Second, the presence of just a few middleboxes can still cover a significant fraction of a network's IP space.

4.1.4. AS Classification of Middleboxes. We now categorize middleboxes according to the so-called usage type of the AS they are located in. The term usage type refers to the purpose of an AS determined by our geolocation database.

Figure 2 illustrates the distribution of middlebox IPs across different AS categories. Three catgegories stand out: Internet Service Providers (both landline and mobile) and Data Centers. Together, they account for 86% of all middleboxes. Commercial (COM) networks and contribute a moderate number of IPs, highlighting some use business-related services. Notably, government-related networks and educational networks appear less frequently. To put this in the right context, it is important to recall here that we test

TABLE 3. Top 10 Countries by Middlebox Density and Their Internet Freedom Status.

Rank	Country	MB Count	MB Density	Internet Freedom Status	Score
1	South Africa	4	0.25	Free	74
2	Belgium	10	0.05	Not Indexed	-
3	Portugal	2	0.05	Not Indexed	-
4	Zambia	2	0.04	Partly Free	62
5	Mongolia	1	0.04	Not Indexed	-
6	Canada	3	0.03	Free	86
7	Sweden	9	0.01	Not Indexed	-
8	Uganda	1	0.01	Partly Free	53
9	Australia	9	0.01	Free	76
10	Rwanda	1	0.01	Not Free	36

TABLE 4. Top 10 ASes with highest middlebox density

AS Name	MB Count	Density
MMS-AS-ID (AS38320/ID)	1	0.13
JHU (AS5723/US)	1	0.13
KAZAKHTELECOM-AS (AS50482/KZ)	1	0.06
PROXIMUS-ISP-AS (AS5432/BE)	4	0.05
ALTIBOX-AS (AS29695/NO)	3	0.04
TELCONET (AS27947/EC)	1	0.03
TELENET-AS (AS6848/BE)	3	0.03
ISEEK-AS-AP (AS9723/AU)	9	0.03
MEO-EMPRESAS (AS15525/PT)	1	0.02
GMEDIA-AS-ID (AS55666/ID)	3	0.02

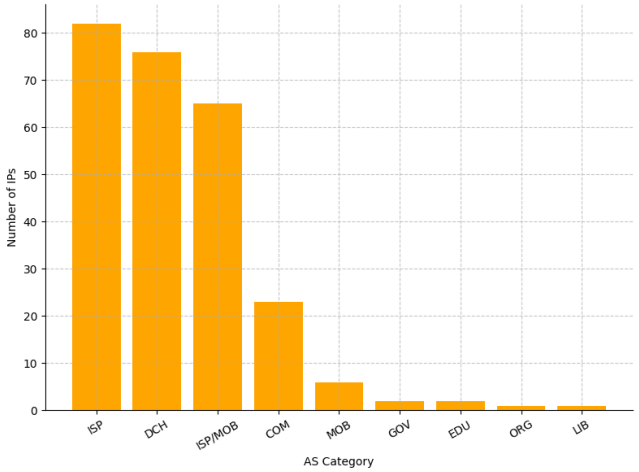

Figure 2. Mapping Middleboxes to AS Usage Type

only for middleboxes in networks that have already been reported for interference. It is quite plausible to assume that government and education networks also deploy many middleboxes, more than what we observed in our analysis due to being less reported for interference.

4.2. Vendor Attribution

We use data from Censys and Shodan to extract vendor information for the identified middlebox IPs. Censys provides data for 125 out of 258 IPs, Shodan for 122. Together, we have data for 139 IPs. Vendor information is available for only 106 of them. For the remaining IPs, we conduct a further investigation to assess their reachability and open ports. We choose the top 5 ports that are commonly open for the above 106 IP addresses but find that all of them are either filtered or closed and no banners can be obtained.

TABLE 5. Top 10 Vendors.

Vendor	Count	Percentage
Check Point	32	30.2%
Palo Alto Networks	21	19.8%
Cisco	19	17.9%
pfSense	4	3.8%
Sophos	4	3.8%
Microsoft	4	3.8%
Mikrotik	3	2.8%
Ubuntu	3	2.8%
Fortinet	3	2.8%
Ruije Networks	3	2.8%

Table 5 lists the top ten vendors in our dataset. The top three are Check Point, Palo Alto Networks and Cisco. The companies are known for firewall and intrusion prevention solutions. Other vendors appear in smaller but still substantive proportions. pfSense is an open-source firewall; Sophos and Fortinet are large commercial enterprises that sells security solutions. Some Fortinet devices have recently been linked to serious vulnerabilities and credential leaks [41], [42].

Comparing this to results from literature, the vendor distribution in our study differs from previous results. The authors of [29] identified censorship devices across four countries and analyzed 19 detected devices. The top vendors were Cisco and Fortinet (together 12 devices). CheckPoint was not idenfied as a vendor at all in [29]. In another relevant study of middlebox detection [26], the authors employed SNMPv3 data to obtain vendor information for 2189 middleboxes, finding that more than half were Cisco devices, about 20% were from, Juniper, and nearly 7% were from Huawei. We note that Censys also scans SNMP banners, so there should not be a massive difference due to our data source.

We attribute the differences to the approach we take, namely taking networks reported for interference as a starting point. This would explain the differences to [26]. We speculate that the differences to [29] can be explained due to the focus of that study on four countries. Further alternative and complementary explanations are that about two years elapsed since the previous studies.

In the following, we give details per vendor and delve into various products we find in our data.

4.2.1. Check Point. Check Point devices represent a significant portion of the identified middleboxes (30.2%). We observed that firewall capabilities are nearly universal across all identified Check Point devices. More than 20 devices also provides VPN functionality. Investigating the open ports and service banners of Check Point devices from correlated Censys and Shodan data, and manually checking product descriptions and release notes, we identified various further functionalities.

One notable example is the Connectra Web Security Gateway, which integrates VPN remote access with endpoint security and intrusion prevention capabilities [43]. We identified 16 devices deploying this solution. We also identified 4 devices with Mobile Access VPN, which is integrated into Check Point Next-Generation Firewalls (NGFW). Furthermore, our findings revealed an instance of the Quantum Spark Security appliance with even more enhanced features (firewall, VPN, threat prevention, email security, but also Wi-Fi capabilities [44]). Finally, we also observe several instances of Check Point devices that also have running SMTP and BGP services. The identification of these multi-functional Check Point devices underscore the diverse applications of middleboxes. It is generally difficult to tell if all features are actually enabled, or which feature causes the modifications that we detect with *yarrpbox*. Further research is warranted to understand in which cases the interference reported by OONI and Censored Planet is a true censorship or malicious interference, and when it is merely a byproduct of some devices' features. This is also supported by the distribution of Check Point devices across 16 countries, with the majority in Turkey and Isreal, which have quite different political directions. The whole distribution of Check Point devices can be found in Figure 3.

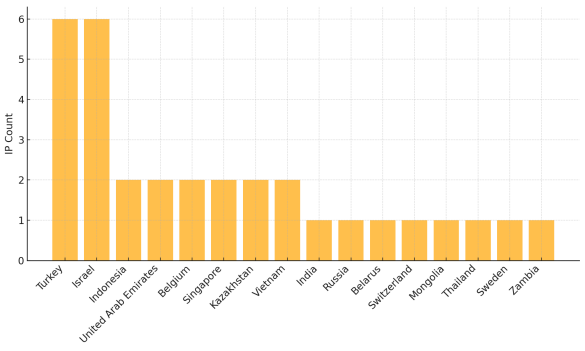

Figure 3. Geographic Distribution of Check Point Devices

4.2.2. Palo Alto. Devices from Palo Alto Networks make up the second-largest share of the identified middleboxes (19.8%). In all these devices, we identify the GlobalProtect Portal in the HTTP response. This is a secure remote access solution primarily designed for enterprise environments. It serves as a VPN or secure gateway solution and is often integrated with Palo Alto's next-generation firewalls (NGFWs) [45]. Investigating the traceroutes where these middleboxes occur, we see that 19 out of 21 are the last hops without our packets reaching the target. Hence, we conclude that these devices are Palo Alto NGFWs that block our connections. As with Check Point, Palo Alto NGFWs are also multi-functional devices that provide advanced capabilities (such as DPI, intrusion prevention, and traffic analytics).

The distribution of Palo Alto devices across different countries is shown in Figure 4. They are deployed in 8 countries, with Singapore and Turkey being the most common (44.4% and 22.2%, respectively).

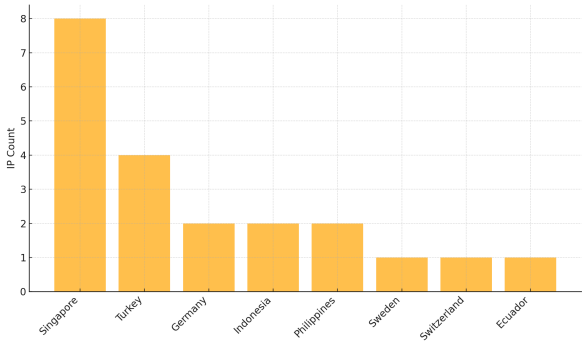

Figure 4. Geographic Distribution of Palo Alto Devices

4.2.3. Cisco. Nearly 18% of the devices middleboxes we identify are from Cisco. Cisco's product range includes many devices with functionality similar to the products from Check Point and Palo Alto. However, the devices we identify have very few ports open, mostly mostly 161 (SNMP). This could be an indication that they are configured for specific functionalities or operated in a more passive capacity (although this leaves the question open why the SNMP port is open to the public Internet). We could not extract detailed information about these devices, except in one instance where ports 22 (SSH) and 500 (IKE) are open, which is a hint that the device may also operate as a VPN endpoint.

With Cisco devices, we also find a peculiar behavior. In most cases (12 out of 19), we see the same IP address repeated at different hops, suggesting an unwanted loop in paths to different targets.

Perhaps more importantly, further scans with *yarrpbox* show that some of the devices do not modify the traffic all the time. Further investigation is needed to understand the reasons behind such behavior. Possible explanations include misconfigurations, routing issues, or the *yarrpbox* methodology leading to artefacts. The distribution of Cisco devices across countries is shown in Figure 5. As with Palo Alto devices, deployment in Singapore is common, but otherwise the distribution resembles neither that of Palo Alto nor that of Check Point devices.

4.2.4. Other Devices. We also identify other device types and vendors.

Firewalls. We find a number of pfSense instances. This is an open-source firewall and router software distribution built on FreeBSD. We observed 4 pfSense devices that function as firewalls. Another vendor we

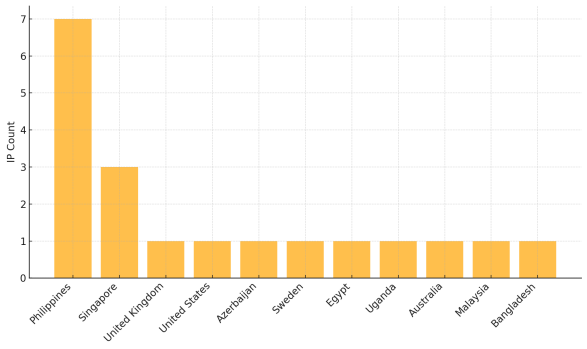

Figure 5. Geographic Distribution of Cisco Devices

find is Sophos, a company whose products also combine threat management and protection, firewalls, and intrusion prevention. We observed 4 Sophos devices with combined firewall and VPN functionalities.

Router-like devices. We also identify several router-like products. Traditional routers should not show up as middlebox, but some products can be configure to manipulate traffic. We discuss some products here. For instance, MikroTik devices are known for their RouterOS software, which also provides bandwidth management, NAT functionality, and even firewall and VPN capabilities. Devices from Ruije Networks devices can offer ACL management and firewall-like inspection functionalities. We observe these devices only in China. Juniper routers also may have additional advanced features such as application-aware routing, intrusion prevention, VPN, and notably content filtering.

General purpose OSes Table 5 shows also instances of general-purpose operating systems, such as Microsoft Windows and Ubuntu, even after we fixed the bug we mentioned earlier in Methodology that sometimes mistook endpoints for middleboxes. Indeed, checking their open ports and services reveals that some Windows systems run additional intermediary services such as SAProuter (a proxy to the well-known SAP product). From port scans of Ubuntu devices, we determine that all 3 of them are configured to perform middlebox-like operations, including routing and reverse proxy services. We assume these devices to be indeed a form of middlebox, albeit an unusual one.

4.3. Vulnerability Analysis

We also find a number of middleboxes to be linked to known vulnerabilities. We observed 13 devices(2 Cisco, 1 Juniper, and 10 unidentified vendor) running NTPv3, which has been superseded by NTPv4 for more than a decade and lacks a number of security features that NTPv4 added to make the protocol more robust. All identified Cisco devices have open SNMP ports, which is unnecessary attack surface unless there is a strong use case for remote management across the Internet.

24 out of 32 Check Point devices were observed to disclose host or even cluster names (which run a corresponding service on port 264). While this information may not constitute a direct security vulnerability, it can facilitate information gathering by potential attackers. One of the Palo Alto devices was running an outdated version of PAN-OS, specifically version 9.0.5, for which several CVEs are known including a recent zero-day vulnerability [46]. We did not attempt to check the patch status of the devices ourselves for ethical reasons.

We observe devices from Ruije Networks and MikroTik to support the PPTP protocol, which is outdated and has known vulnerabilities. Some Mikrotik devices displayed the "MikroTik MRO-QSR 1.0 0.1" banner, indicating the presence of older firmware versions potentially lacking critical security updates.

Lastly, two Microsoft devices ran Windows Server 2008 R2 and Windows Server 2012 R2, both of which have reached their end-of-life and no longer receives security updates.

Although not directly linkable to a weakness, we also frequently observe a use of HTTP headers that would be considered problematic on an interactive website. One example is the use of *unsafe-inline* in *Content-Security-Policy* headers, which can lead to Cross-Site-Scripting attacks (although we did not check if this was exploitable here, again for ethical reasons). Another example is repeated use of *Set-Cookie* headers for the same session token within a single response—although not a weakness itself, this leaves questions open how the session management was implemented.

5. Discussion and Conclusion

Our study represents a first step in investigating the feasibility of attributing vendors to middleboxes, an area of research that has so far been little explored. Starting from a dataset of approximately 2 million IP addresses across 518 interference-related ASNs (as reported by OONI and Censored Planet), we employed `yarrpbox` to identify middleboxes. Out of 1289 potential middlebox IPs, we identified 258 with high confidence. Analysis of their network locations revealed that 74.4% of these middleboxes reside within the same AS as the targets. We examined middlebox deployment at the country level and correlated our findings with the Internet Freedom Index. From this analysis we conclude that the presence of middleboxes in a country does not necessarily correlate with its level of Internet freedom, suggesting that many middleboxes are deployed for benign purposes such as performance optimization.

In the vendor analysis, we successfully identified the manufacturers for 106 middleboxes. The predominant vendors were Check Point (30.2%), Palo Alto (19.8%), and Cisco (17.9%), which was not in line with previous work on studying censorship and middleboxes. Although this difference may be influenced by the data sources and scanning methods we employed, it could also indicate changes in market share. The distribution of vendors also varies by country, suggesting preferences for certain vendors—either due

to functionality or simply market effects. Notably, the top 3 vendors are all headquartered in the USA.

Our analysis of the identified devices indicates a trend towards the deployment of unified middleboxes with advanced functionalities, such as next-generation firewalls (NGFWs), which blurs the distinction among traditional middlebox types and complicates their classification. This observation also appears to align with the broader trend of softwarization of network components, where standalone devices dedicated to specific tasks are increasingly being replaced by more integrated systems capable of performing multiple functions. However, further research is needed to effectively distinguish between software- and hardware-based middleboxes and to investigate the extent of the shift toward softwarization and network function virtualization. Moreover, further research is also warranted to determine when the reported interference is intentional (*e.g.,* due to censorship) and when it is merely a byproduct of some security feature.

The security issues we identified for some devices are associated with outdated software and protocols, and possibly misconfigurations. It seems worthwhile to continue with further investigations here. Since we only studied networks that had been linked to network interference before, a follow-up study should also compare the distribution of middleboxes against a set of networks that have *not* been reported for interference. This would allow to understand the general deployment patterns better. Further work should also consider more intensive port probing on middleboxes for which we had no vendor attribution, insofar as that is ethically defensible.

In summary, our study demonstrates that attributing middleboxes to specific vendors is feasible with reasonable confidence, although several key challenges remain. First, available middlebox detection tools, including Yarrpbox, currently detect modifications at a generic or aggregate level, rather than attributing them explicitly to individual devices or specific functions. This limitation restricts precise functional categorization and vendor attribution. Second, the limited visibility and deliberate obfuscation techniques employed by certain devices further hinder comprehensive middlebox detection. Additionally, our analysis reveals no clear correlation between the prevalence of detected middleboxes in a given country and its Internet Freedom Index ranking. Furthermore, we identify an increasing trend toward deploying unified, multi-functional middlebox devices, complicating straightforward classification. Collectively, these insights emphasize the necessity of developing more targeted, refined detection methodologies and tools to accurately characterize middlebox functionalities, ascertain their intended purposes, and map them reliably to vendors. Overall, we view our findings as indicative that the investigation into middleboxes is far from complete; much remains to be understood about their diverse functionalities, deployment contexts, and broader implications for Internet governance and security.

Acknowledgments

We thank Censys and Shodan for providing us with academic licences. We also thank Censored Planet for giving us access to their data and providing assistance. This work is supported by the research project 'CATRIN' (NWA.1215.18.003) as part of the Dutch Research Council's (NWO) National Research Agenda (NWA) and partly financed by the Province of Gelderland and Centre for Safety & Digitalisation.

References

[1] B. Hesmans, F. Duchene, C. Paasch, G. Detal, and O. Bonaventure, "Are tcp extensions middlebox-proof?," in *Proceedings of the 2013 workshop on Hot topics in middleboxes and network function virtualization*, pp. 37–42, 2013.

[2] A. Medina, M. Allman, and S. Floyd, "Measuring the evolution of transport protocols in the internet," *SIGCOMM Comput. Commun. Rev.*, vol. 35, p. 37–52, Apr. 2005.

[3] K. Edeline and B. Donnet, "A first look at the prevalence and persistence of middleboxes in the wild," in *2017 29th International Teletraffic Congress (ITC 29)*, vol. 1, pp. 161–168, IEEE, 2017.

[4] S. Huang, F. Cuadrado, and S. Uhlig, "Middleboxes in the internet: a http perspective," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 1–9, IEEE, 2017.

[5] T. Kohno, A. Broido, and K. C. Claffy, "Remote physical device fingerprinting," *IEEE Transactions on Dependable and Secure Computing*, vol. 2, no. 2, pp. 93–108, 2005.

[6] L. Izhikevich, R. Teixeira, and Z. Durumeric, "{LZR}: Identifying unexpected internet services," in *30th USENIX Security Symposium (USENIX Security 21)*, pp. 3111–3128, 2021.

[7] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis, "Third time's not a charm: exploiting snmpv3 for router fingerprinting," in *Proceedings of the 21st ACM Internet Measurement Conference*, pp. 150–164, 2021.

[8] R. Sundara Raman, P. Shenoy, K. Kohls, and R. Ensafi, "Censored planet: An internet-wide, longitudinal censorship observatory," in *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, CCS '20, (New York, NY, USA), p. 49–66, Association for Computing Machinery, 2020.

[9] "OONI: Open observatory of network interference," in *2nd USENIX Workshop on Free and Open Communications on the Internet (FOCI 12)*, (Bellevue, WA), USENIX Association, Aug. 2012.

[10] Censys, "Censys scanning and data collection," 2024.

[11] Shodan, "Shodan: The search engine for internet-connected devices," 2024.

[12] B. Carpenter and S. Brim, "Rfc3234: Middleboxes: Taxonomy and issues," 2002.

[13] J. Pahdye and S. Floyd, "On inferring tcp behavior," *SIGCOMM Comput. Commun. Rev.*, vol. 31, p. 287–298, Aug. 2001.

[14] A. Medina, M. Allman, and S. Floyd, "Measuring interactions between transport protocols and middleboxes," in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*, IMC '04, (New York, NY, USA), p. 336–341, Association for Computing Machinery, 2004.

[15] M. Honda, Y. Nishida, C. Raiciu, A. Greenhalgh, M. Handley, and H. Tokuda, "Is it still possible to extend tcp?," in *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, IMC '11, (New York, NY, USA), p. 181–194, Association for Computing Machinery, 2011.

[16] R. Craven, R. Beverly, and M. Allman, "A middlebox-cooperative tcp for a non end-to-end internet," *ACM SIGCOMM Computer Communication Review*, vol. 44, no. 4, pp. 151–162, 2014.

[17] I. R. Learmonth, A. Lutu, G. Fairhurst, D. Ros, and Ö. Alay, "Path transparency measurements from the mobile edge with pathspider," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 1–6, IEEE, 2017.

[18] K. Edeline and B. Donnet, "A bottom-up investigation of the transport-layer ossification," in *2019 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 169–176, IEEE, 2019.

[19] C. Reis, S. D. Gribble, T. Kohno, and N. C. Weaver, "Detecting in-flight page changes with web tripwires.," in *NSDI*, vol. 8, pp. 31–44, 2008.

[20] C. Kreibich, N. Weaver, B. Nechaev, and V. Paxson, "Netalyzr: Illuminating the edge network," in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement*, pp. 246–259, 2010.

[21] R. Sundara Raman, L.-H. Merino, K. Bock, M. Fayed, D. Levin, N. Sullivan, and L. Valenta, "Global, passive detection of connection tampering," in *Proceedings of the ACM SIGCOMM 2023 Conference*, pp. 622–636, 2023.

[22] G. Detal, B. Hesmans, O. Bonaventure, Y. Vanaubel, and B. Donnet, "Revealing middlebox interference with tracebox," in *Proceedings of the 2013 conference on Internet measurement conference*, pp. 1–8, 2013.

[23] V. Thirion, K. Edeline, and B. Donnet, "Tracking middleboxes in the mobile world with traceboxandroid," in *Traffic Monitoring and Analysis: 7th International Workshop, TMA 2015, Barcelona, Spain, April 21-24, 2015. Proceedings 7*, pp. 79–91, Springer, 2015.

[24] R. Zullo, A. Pescapé, K. Edeline, and B. Donnet, "Hic sunt nats: Uncovering address translation with a smart traceroute," in *2017 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 1–6, IEEE, 2017.

[25] R. Zullo, A. Pescapé, K. Edeline, and B. Donnet, "Hic sunt proxies: Unveiling proxy phenomena in mobile networks," in *2019 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 227–232, IEEE, 2019.

[26] F. Hilal and O. Gasser, "Yarrpbox: Detecting middleboxes at internet-scale," *Proc. ACM Netw.*, vol. 1, July 2023.

[27] Y. Vanaubel, J.-J. Pansiot, P. Mérindol, and B. Donnet, "Network fingerprinting: Ttl-based router signatures," in *Proceedings of the 2013 conference on Internet measurement conference*, pp. 369–376, 2013.

[28] T. Albakour, O. Gasser, R. Beverly, and G. Smaragdakis, "Illuminating router vendor diversity within providers and along network paths," in *Proceedings of the 2023 ACM on Internet Measurement Conference*, pp. 89–103, 2023.

[29] R. S. Raman, M. Wang, J. Dalek, J. Mayer, and R. Ensafi, "Network measurement methods for locating and examining censorship devices," in *Proceedings of the 18th International Conference on Emerging Networking EXperiments and Technologies*, CoNEXT '22, (New York, NY, USA), p. 18–34, Association for Computing Machinery, 2022.

[30] Open Observatory of Network Interference, "OONI: Open Observatory of Network Interference," 2025.

[31] Censored Planet, "Censored Planet: Internet Censorship Measurement," 2025.

[32] R. S. Raman, A. Virkud, S. Laplante, V. Fortuna, and R. Ensafi, "Advancing the art of censorship data analysis," *Free and Open Communications on the Internet*, 2023.

[33] Y. Schwartz, Y. Shavitt, and U. Weinsberg, "On the diversity, stability and symmetry of end-to-end internet routes," in *Conference on Computer Communications (Infocom) Workshops*, 2010.

[34] W. de Vries, J. J. Santanna, A. Sperotto, and A. Pras, "How asymmetric is the internet?," in *Int. Conf. on Autonomous Infrastructure, Management, and Security*, June 2015.

[35] L. Bertholdo, S. L. A. Ferreira, J. M. Ceron, Holz, G. L. Z. R., and R. M. van Rijswijk-Deij, "On the asymmetry of internet exchange points-why should ixps and cdns care?," in *Proc. Int. Conf. on Network and Service Management (CNSM)*, October 2022.

[36] A. Snarskii, C. David, C. Jeker, J. Snijders, M. Stucchi, M. Litvak, P. Schoenmaker, and R. Wichertjes, "bgpq4: Bgp filter generator." https://github.com/bgp/bgpq4, 2025.

[37] Z. Durumeric, E. Wustrow, and J. A. Halderman, "Zmap: Fast internet-wide scanning and its security applications," in *Proceedings of the 22nd USENIX Security Symposium (USENIX Security 13)*, pp. 605–620, 2013.

[38] IP2Location, "Db23 ip-country-region-city-latitude-longitude-isp-domain-mobile-usagetype database," 2024.

[39] F. House, "Freedom on the net 2024: The global drive to control big tech," 2024.

[40] Team Cymru, "Ip to asn mapping service," 2025. Accessed: January 19, 2025.

[41] BleepingComputer, "Hackers leak passwords for 500,000 fortinet vpn accounts," September 2021. Accessed: February 24, 2025.

[42] SecurityWeek, "Data from 15,000 fortinet firewalls leaked by hackers," January 2025. Accessed: February 24, 2025.

[43] Check Point Software Technologies, "Connectra datasheet," Retrieved 2025. Accessed: February 3, 2025.

[44] Check Point Software Technologies, "Check point quantum next-generation firewall for small businesses," 2025. Accessed: February 3, 2025.

[45] Palo Alto Networks, *GlobalProtect Administrator's Guide*, 2021. Accessed: February 3, 2025.

[46] Cyber Security News, "Pan-os vulnerability: Web interface authentication," February 2025. Accessed: February 25, 2025.

Appendix

1. Fingerprinting Check Point Devices

By examining Censys and Shodan data, we identified the following fingerprints for Check Point devices.

HTTP Header Fingerprints: Many of the HTTP responses from Check Point devices show "Server: Check Point SVN foundation" in the header indicating the presence of a Check Point gateway or management interface for Check Point firewalls and Unified Threat Management devices. We also observed that some responses contain "Server: CPWS". CPWS is another banner identifying Check Point Web Service which is of the part of the Gaia Portal, Mobile Access/SSL VPN, or the Management Portal. Some devices also show "Location: `https://<gateway>/sslvpn/Login/BrowserSupport`" in the header alongside with CPWS, which makes it a strong indicator of Check Point Mobile Access (SSL VPN) portals.

HTML Body / Error Pages: The combination of "Check Point SVN foundation", "X-UA-Compatible: IE=EmulateIE7", a "Content-Length: 204" is recurring across multiple Check Point error pages.

Certificate / Binary Data Fingerprints: By analyzing raw byte arrays in response data, we identified recurring strings such as "sslca_clear", "sslca_comp", and "sslca_rc4", which frequently appear before or after standard HTTP responses. This pattern suggests the presence of a custom handshake or certificate exchange mechanism in Check Point devices. Additionally, we observed a distinctive pattern, "CN=MyGateway-1,O=MyMgmtServer...", which likely represents auto-generated internal certificates associated with Gaia-OS which is Check Point's unified operating system designed specifically for their security appliances.

Open Port Patterns: Our observations indicate that Check Point devices commonly utilize ports 80, 264, 443, 500, and 18264. These ports are frequently found open concurrently; however, in certain instances, devices exhibit only one or two of these ports being open. Additionally, some devices were found to have supplementary open ports, suggesting variations in configuration or deployment contexts.

Operating System: Most Check Point devices run on Gaia OS, a Linux-based operating system designed for Check Point security devices. The presence of Gaia OS is a strong indicator of Check Point devices, as it is the primary operating system used in Check Point products.

2. Fingerprinting Palo Alto Devices

We deducted the following fingerprints for Palo Alto devices based on our analysis of Censys and Shodan data.

HTTP Header Fingerprints: The majority of Palo Alto devices display the HTTP title: "Global-Protect Portal", which indicates the presence of Palo Alto Networks' next-generation firewalls or Prisma Access, designed to provide secure remote access. Moreover, we observed that Palo Alto devices consistently return the following HTTP response headers: "Strict-Transport-Security: max-age=31536000", "X-Frame-Options: DENY", "X-XSS-Protection: 1; mode=block", "X-Content-Type-Options: nosniff", and "Content-Security-Policy: default-src 'self'; script-src 'self' 'unsafe-inline'; img-src * data:; style-src 'self' 'unsafe-inline'; frame-ancestors 'none';". Furthermore, cookie flags like SameSite=Lax, HttpOnly, and Secure appear consistently, which can be used as a supporting information to identify Palo Alto devices in conjunction with other fingerprints.

Operating System: We observed that Palo Alto devices predominantly run on PAN-OS which is the software that runs all Palo Alto next-generation firewalls. These devices have additional features such as URL filtering, threat prevention, and DNS security.

3. Fingerprinting Cisco Devices

Upon examining Censys and Shodan data, we identified the following fingerprints for Cisco devices.

SNMP Details: All Cisco devices except one instance are identified by their SNMP responses. The SNMP responses contain the SNMP Enterprise OID = 9 which is Cisco's official SNMP enterprise number. Moreover, we observed that Cisco devices have their Engineid Data in the format of "000000090300<MAC Address>".

SSH Banner: Although we observed one instance of it, when SSH is enabled on Cisco devices, the banner contains the string "Cisco" and Software Version as "SSH-2.0-Cisco-1.25". We believe this can be used to identify more Cisco devices as we saw many similar cases checking the community forums of Cisco.

4. Further tables

TABLE 6. Top 10 Countries with the highest number of middleboxes.

Rank	Country	IP Count (%)
1	SG (Singapore)	20 (7.75%)
2	TR (Turkey)	19 (7.36%)
3	US (United States)	19 (7.36%)
4	DE (Germany)	17 (6.59%)
5	GB (United Kingdom)	16 (6.20%)
6	ID (Indonesia)	15 (5.81%)
7	PH (Philippines)	14 (5.43%)
8	BE (Belgium)	10 (3.88%)
9	SE (Sweden)	9 (3.49%)
10	AU (Australia)	9 (3.49%)

TABLE 7. Top 10 ASes with the highest number of middleboxes

ASN	Count	%
TELLCOM-AS(TR)	18	6.9%
IPG-AS-AP(PH)	14	5.4%
HURRICANE(US)	13	5.0%
DTAG(DE)	13	5.0%
SINGNET(SG)	10	3.9%
LIQUID-AS(GB)	10	3.9%
ISEEK-AS-AP(AU)	9	3.5%
AS6453(US)	6	2.3%
TELE2(SE)	6	2.3%
TWELVE99(SE)	5	1.9%

TABLE 8. Observed modifications. See [26] for details.

Type of Interference	#	%
TCP NOP Addition	4,598	22.06%
TCP MP Capable Removal	4,498	21.58%
IP ID/TSval/TSecr/RW or UP	4,282	20.54%
IP ID/TSval/RW or UP	1,843	8.84%
TSecr/RW or UP	1,274	6.11%
TCP Urgent Pointer/Receiver Window	1,272	6.10%
TCP Timestamp Tsecr	1,102	5.29%
TCP Timestamp TSVal	322	1.54%
TCP Timestamp Zeroed/Modified	305	1.46%
TCP Timestamp Removal	303	1.45%
IP Total Length	300	1.44%
TCP Sack Permitted Removal	235	1.13%
TCP Sequence Number	135	0.65%
TCP Data Offset	133	0.64%
TCP Flags	119	0.57%
TCP MSS Removal	119	0.57%
TCP MSS Data	7	0.03%