Comparison of cyber attacks on services in the clearnet and darknet

York Yannikos, Quang Anh Dang, Martin Steinebach

Chapter 3

COMPARISON OF CYBER ATTACKS ON SERVICES IN THE CLEARNET AND DARKNET

York Yannikos, Quang Anh Dang and Martin Steinebach

Abstract Cyber attacks on clearnet services are discussed widely in the research literature. However, a systematic comparison of cyber attacks on clearnet and darknet services has not been performed. This chapter describes an approach for setting up and simultaneously running honeypots with vulnerable services in the clearnet and darknet to collect information about attacks and attacker behavior. Key observations are provided and the similarities and differences regarding attacks and attacker behavior are discussed.

Keywords: Clearnet, darknet, services, honeypots, attacks, attacker behavior

1. Introduction

Almost every publicly-available Internet service is subjected to cyber attacks. Much research has focused on attack frequencies, attack patterns and attacker behavior. Research has also investigated attacks on services in the clearnet by deploying honeypots [13, 25]. However, very little is known about attacks on services in the darknet, including onion services in the Tor network. At this time, there are no insights into how often clearnet services face cyber attacks compared with darknet services, the sophistication of the attacks and the motives of the attackers. Another open question is how automated attacks on clearnet services compare with those on darknet services.

Considerable research has focused on malware collection, traffic analysis and honeypots in the darknet [2, 11, 18, 21, 22]. However, the research essentially considers the darknet as a network telescope – public IP addresses that are not assigned to legitimate hosts [16] – instead

of an overlay network with a strong focus on anonymity as in the case of the Tor network.

This chapter discusses the similarities and differences between attacks observed in the clearnet and darknet using honeypots. The approach towards implementing, deploying and observing multiple honeypots in the two environments during the same time span is also detailed. Diverse attack vectors were considered by covering multiple protocols such as HTTP, SSH and Telnet that are attractive attack targets. The collected data was analyzed with a focus on automated attacks, attack frequencies and attacker behavior.

2. Background

Several definitions have been proposed for the clearnet and darknet. This section clarifies the definitions based on recent work by Kavallieros et al. [14]:

- **World Wide Web:** The World Wide Web, or simply the web, is an important part of the Internet. It houses a massive collection of diverse documents, many of which are linked. The documents are usually accessed via web browsers using the HTTP or HTTPS protocols. Other important components of the Internet, which are not part of the web, include services that support file transfer, email and hostname resolution.

- **Surface Web:** The surface web, also called the visible web, is a subset of the web that holds all the indexed web content on the Internet. The indexed web content is accessed using search engines like Google and Bing. Of course, the surface web only holds a small portion of the total content available on the Internet.

- **Deep Web:** The deep web is a subset of the web holding content that is not indexed by search engines. The reasons for content not being indexed vary. Examples are private web resources with no external links pointing to them and web content accessible only via login credentials or a virtual private network. Because search engines continuously expand and refine their indexing capabilities, the deep web and surface web content are always in flux.

- **Darknet:** The darknet is a subset of the Internet that comprises many separated networks. These decentralized networks function as overlay networks on top of the Internet infrastructure. The networks, which are only accessed via additional software, support anonymous participation and communications. The darknet can

provide content from the deep web and surface web as well as non-web services.

- **Dark Web:** The dark web is essentially the web portion of the darknet. It holds all the web content of the darknet. The content is usually accessed via web browsers using the HTTP or HTTPS protocols.

- **Clearnet:** The clearnet is the counterpart to the darknet. Every service, participant and content in the Internet that is not in the darknet is part of the clearnet. The union of the clearnet and darknet is the Internet.

This research uses the terms clearnet and darknet as defined above. The Tor network in the darknet was chosen as the testing environment due to its popularity:

- **Tor Network:** Tor, which stands for The Onion Router, is one of the overlay networks of the darknet. Tor provides anonymity to its users via onion routing, an ingenious way of routing packets that also leverages several layers of encryption. Every packet in the Tor network is routed via a predefined path with a minimum of three nodes, ensuring that the source node of a packet is never directly connected to its destination node. The routing paths are called Tor circuits.

  Tor supports two important features. It enables its participants to anonymously access clearnet services outside Tor as well as anonymously access host-specific services called onion services in Tor. Onion services typically host websites, but almost any other common Internet service can be hosted (e.g., services that support file transfer, email and chats).

  Since Tor is the most popular darknet network, it is often used as a synonym for the entire darknet.

3. Common Targets and Attacks

This section describes the services that are commonly targeted by cyber attacks. Based on resources such as OWASP Top Ten [20], this section also discusses the types of attacks expected when deploying honeypots hosting commonly-targeted services.

Common targets of cyber attacks are:

- **Web Servers:** As in the clearnet, web content is the most-accessed content in the darknet. Therefore, web servers are attractive tar-

gets for attackers, especially those seeking quick monetary gains (e.g., taking over Tor marketplaces or bitcoin escrow services).

- **Remote Access Services:** Remote access services such as SSH are typically used for system administration and are regularly attacked. In the worst case, a vulnerable SSH server, insecure configuration or weak login credentials for a root account can enable an attacker to take over a system that hosts a variety of services.
- **Email Servers:** Email servers are popular targets for attackers. Successful attacks often provide access to sensitive information contained in email. Additionally, email servers can be leveraged to launch other attacks such as spamming and phishing.
- **Database Servers:** Database servers are targeted for the sensitive information maintained in databases. The servers should not be directly accessible from the Internet, but this is often not the case. Several prominent data leaks are the direct result of targeting insecure, remotely-accessible database servers.

Attackers typically employ reconnaissance techniques such as port scanning and fingerprinting to gather information about potential targets. Tools such as Nmap, ZMap and Masscan support efficient port scanning of large portions of the Internet. Attacks are launched based on the open ports that are discovered and the services they support. Common attacks are:

- **Injection Attacks:** SQL, NoSQL, LDAP and other injection attacks are very common on the Internet [20]. Injection attacks are often successful when untrusted user inputs, such as SQL database queries, are not properly checked and sanitized before they are processed. Services vulnerable to injection attacks can be found and exploited automatically using tools such as `sqlmap` and Metasploit.
- **Brute Force Attacks:** Brute force attacks are commonly used against services that require authentication with login credentials. Examples are websites, SSH and FTP. Brute force attacks can be automated very easily when two requirements are met. First, the server's response to a failed login attempt is distinguishable from a successful attempt (which is almost always the case). Second, the server does not use a mechanism such as reCAPTCHA to defeat automated requests.
- **Cross-Site Scripting Attacks:** Cross-site scripting (XSS) attacks are included in the OSWASP list of common attacks against

web servers. A cross-site scripting attack is an injection attack where untrusted user input is (temporarily or permanently) embedded in a web page. If the input is not properly sanitized, then an attacker could, for example, embed JavaScript code in a web page that is executed when a victim visits the page. The embedded code could steal cookies, hijack a user session and/or redirect the victim to a malicious website.

4. Related Work

The darknet and its offerings have drawn the attention of many researchers in recent years. This section provides an overview of research related to the Tor network.

Catakoglu et al. [6] conducted research similar to that described in this chapter. They deployed a honeypot service to collect and analyze data pertaining to attacks in the darknet. In addition to a web service, they deployed other services such as SSH and IRC in their honeypots. However, they did not compare simultaneously-running services in the darknet and clearnet to gain insights into the similarities and differences in attacker behavior.

Zeid et al. [26] used two honeypots to collect evidence of malicious and illegal activities in the darknet. They installed a chatroom and a vulnerable website as onion services in Tor, and collected information about individuals who searched for child pornography and hacking techniques. They also developed an advanced chatbot for the darknet by training a recurrent neural network with data collected from two large darknet marketplaces [17].

From 2013 through 2015, Branwen [3] scraped a large amount of data from 89 Tor marketplaces and made it publicly available. Several researchers have analyzed the data. For example, Broseus et al. [5] conducted a comprehensive analysis of data about global trafficking in darknet marketplaces, providing valuable insights into the types of goods, originating countries and specializations of the dealers.

Steinebach et al. [23] analyzed the popularity of hidden services and demonstrated that a significant percentage of the accessed darknet services belonged to botnets. Other popular services included drug markets, hacking forums and websites with political content. Flamand and Decary-Hetu [12] obtained insights into the human factors underlying cyber crime by analyzing drug offerings in the Tor network with a focus on the entrepreneurship activities of online drug dealers.

5. Honeypot Deployment

A three-phase process was applied to deploy honeypot services on two virtual machines (VMs), one each in the clearnet and darknet. This section describes the security considerations and the phases of the deployments.

5.1 Security Considerations

Because the goal was to have the honeypot services attacked, security issues pertaining to the virtual machines had to be minimized. These included the complete takeover of the virtual machines, deletion of virtual machine data and using the virtual machines to launch internal and external attacks. Therefore, the virtual machines were placed in an isolated network protected by strict firewall rules. Additionally, each honeypot service was executed in an isolated environment in the virtual machines. Details about the firewall configuration and service isolation are provided below.

5.2 Deployment Process

The deployment of honeypot services involved three phases:

- **Phase 1:** In the first phase, two popular services – a web service and an SSH service – with vulnerabilities were deployed on the clearnet and darknet virtual machines. The two services constituted the foundation of the honeypots. In order to announce the availability of the honeypot services on the Internet, a domain name was registered and an onion address was created for the clearnet and darknet virtual machines, respectively. All activities involving the deployed honeypot services were monitored. After a few days of monitoring, the availability of the services was announced using various platforms and search engines. The services were not announced immediately because the intent was to first check if traffic related to automated scans for web or SSH services was observed. Although it was very unlikely that the onion address of the darknet virtual machine would be found by a visitor purely by guessing, it was expected that the clearnet virtual machine would encounter some web or SSH traffic before the announcement.

- **Phase 2:** In the second phase, all the ports were monitored, and the ports (and HTTP paths) that were most frequently scanned were analyzed. Based on the number of observed port scans, it was decided to deploy additional honeypot services for Telnet, SMTP

and FTP. In the case of the darknet virtual machine, high interest in port 443 for HTTPS was observed. Therefore, an HTTPS proxy with a self-signed certificate was deployed on the darknet virtual machine.

- **Phase 3:** In the third phase, adjustments were made to the honeypot services based on the numbers of HTTP path scans and attacks observed during the second phase. To increase the probability of receiving attack traffic, the honeypot services were announced on additional platforms (e.g., by posting links on social networks). Additional vulnerabilities were also added to the honeypot services to expand the attack surface.

6. Implementation Details

This section describes the virtual machine architectures and the honeypot service implementations.

6.1 Virtual Machine Architectures

Figures 1 and 2 present the architectures of the virtual machines deployed to observe attackers in the clearnet and darknet, respectively. The virtual machines used the Arch Linux operating system and docker to run each honeypot service in an isolated manner. An additional SSH service for administrative purposes and a MariaDB instance for the central storage of logged data were installed on each virtual machine. Also, iptables was installed on each virtual machine to serve as a firewall, specifically to control the amount and type of traffic reaching each honeypot service. The iptables configurations forwarded traffic to the services and simultaneously logged the traffic on all ports. To prevent outgoing attacks, any and all attempts to create new connections from the virtual machines to external systems were blocked.

A Tor proxy was installed on the darknet virtual machine because it required a Tor connection. All the IP addresses that requested honeypot services on the clearnet virtual machine were monitored; however, it was not possible to collect IP addresses when using Tor. Therefore, the Tor proxy was configured to log the IDs of the Tor circuits associated with requests and prevent early closures of the circuits. This enabled the mapping of requests via Tor to the Tor circuits used by individuals.

6.2 Honeypot Services

The honeypot services were deployed in separate docker containers:

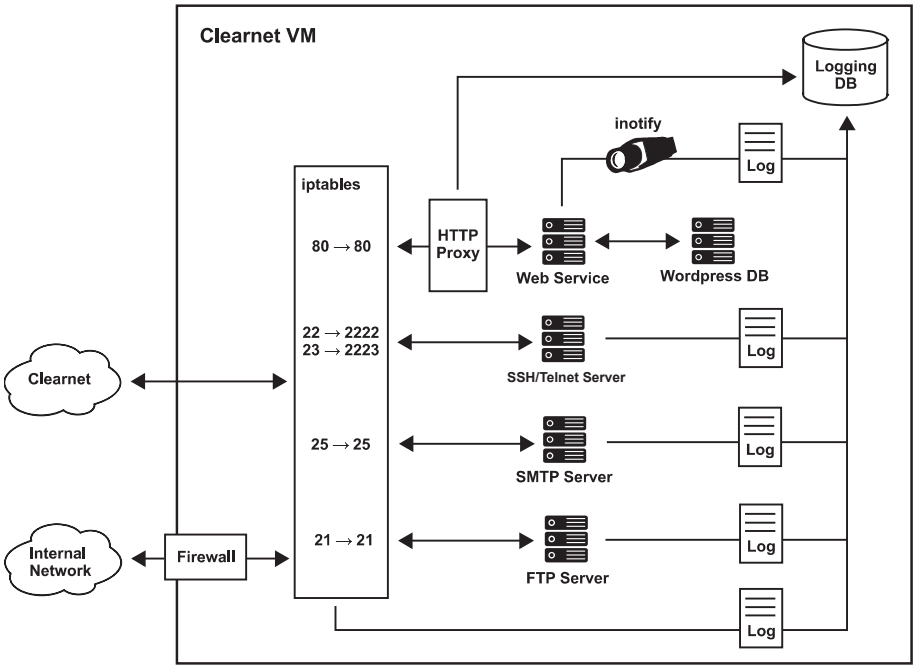

Figure 1. Architecture of the clearnet virtual machine.

- **Web:** Web services were deployed in two docker containers, one with an Apache web server (version 2.4.25 from December 2016) hosting a Wordpress instance and the other with the MySQL server required for Wordpress. Wordpress was used as an image gallery with pictures of cats and dogs. Harmless content was chosen over political, religious and other controversial content because the intent was not to attract attackers with specific agendas. Instead, the goal was to attract attackers based on the technical vulnerabilities of the deployed software.

  A small HTTP proxy was installed in front of the Apache web server to filter certain HTTP requests, perform geolocation lookups of IP addresses (clearnet virtual machine only), and log all HTTP requests (including payloads) and responses. In the case of the darknet virtual machine, an HTTPS proxy using a self-signed certificate was also installed.

  Two versions of honeypot web services were employed. The first version, which employed Wordpress version 4.9 from November 2017, incorporated the following additional vulnerabilities:

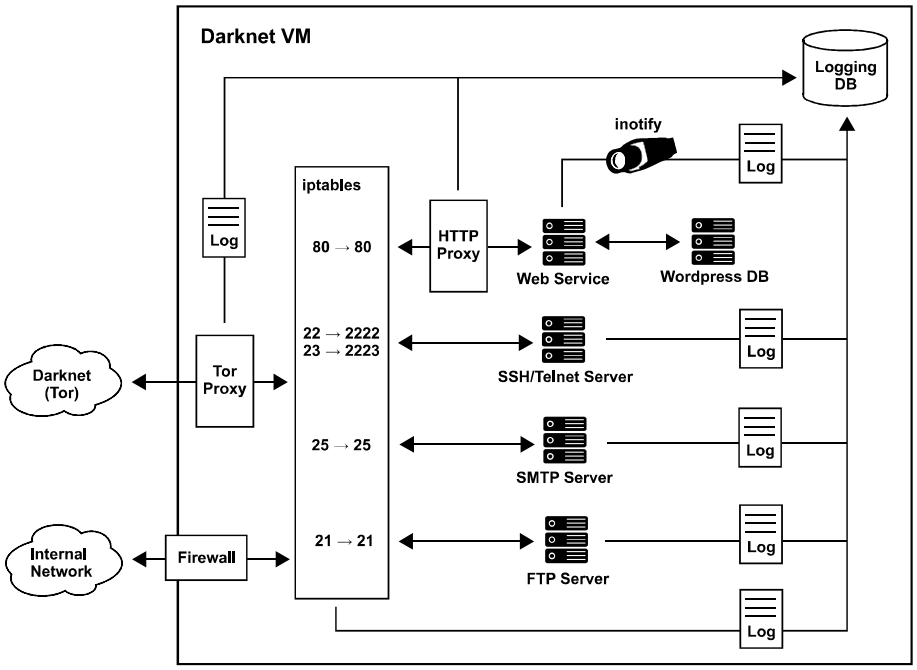

Figure 2. Architecture of the darknet virtual machine.

- Verbose directory listings to disclose files in directories with no index pages.
- Multiple custom login and registration forms with weak credentials to gain database access. The credentials could also be used for authentication by the SSH/Telnet services.
- A built-in SQL injection vulnerability to bypass authentication using the login form.
- A file providing `phpinfo()` and other verbose PHP errors to support information disclosure.
- A customized 404 error page in PHP with a built-in cross-site scripting vulnerability.

A crawler trap was set up to distinguish automated attacks from manual attacks. Specifically, an empty HTML page with a random (i.e., hard to guess) name was inserted in the root directory and a link to it was placed in the index page using a small (10×1 pixel) invisible image. A `robots.txt` file that referred to files suggestive of sensitive content was also created.

The second version of the web honeypot service incorporated the following adjustments and additional vulnerabilities:

- Wordpress version 4.4 from December 2015 was used instead of version 4.9 from November 2017 to attract more attackers.
- A PHP file `upload.php` that allowed arbitrary file uploads was created.
- Web shells `shell.php`, `cmd.php` and `c99.php` that allowed command execution on the web server were installed to indicate that the server had already been compromised.

- **SSH/Telnet:** The SSH/Telnet honeypot services used Cowrie [8], a UNIX emulation with limited functionality that does not provide system access. As a result, an attacker working manually would probably recognize it as a honeypot. Nevertheless, Cowrie was a good choice based on its functionality and the desire to observe automated attacks. Cowrie's SSH service was enabled first and the Telnet service was enabled later.

- **SMTP:** The SMTP honeypot service used Mailoney [1], which imitates the functionality of a real SMTP server but silently logs and discards all sent emails. It was configured to enable email to be sent without authentication, thereby imitating an open mail relay.

- **FTP:** An FTP honeypot [4] written in Python was installed to masquerade as a real FTP server. The FTP service was configured to allow anonymous and authenticated uploads with weak credentials and log all FTP commands.

7. Experiments and Results

This section describes the steps involved in deploying and announcing the honeypot services hosted by the virtual machines in the clearnet and darknet. Data was collected from December 29, 2018 through March 14, 2019. This section also presents the results obtained by analyzing the attacker traffic logged on the virtual machines.

7.1 Service Deployments

Web and SSH honeypot services were deployed on December 29, 2018. Additional honeypot services were deployed and existing honeypot services were adjusted incrementally to attract attackers. Table 1 shows the deployment dates and uptimes of the honeypot services.

Table 1. Chronological list of honeypot service deployments.

Service	Start	End	Uptime
Web honeypot (v1)	29.12.2018	22.02.2019	56d
SSH honeypot	29.12.2018	14.03.2019	76d
iptables	14.01.2019	14.03.2019	60d
Telnet honeypot	20.01.2019	14.03.2019	54d
SMTP honeypot	24.01.2019	14.03.2019	50d
FTP honeypot (port 2121)	24.01.2019	25.02.2019	33d
HTTPS proxy (darknet)	13.02.2019	14.03.2019	26d
Web honeypot (v2)	23.02.2019	14.03.2019	20d
FTP honeypot (port 21)	26.02.2019	14.03.2019	17d

7.2 Announcements

After deploying the virtual machines during the first phase, the honeypot services were announced on the Internet a few days later. The announcements, which included web addresses or onion addresses, were issued as posts or comments on platforms such as Reddit. Indexing requests were also sent to several search engines. Table 2 shows the dates of the announcements issued after the initial deployment.

7.3 Observed Web Requests

It was necessary to distinguish automated web requests from manual requests. The following criteria were used to identify web requests as automated (bot) traffic:

- Crawler trap, `robots.txt` and referenced files were accessed.
- Common IP address ranges for bots from Google, Bing and other search engines were employed (clearnet only).
- The IP address instead of the domain name was targeted (clearnet only).
- Non-existent files or directories not linked to or visible on the website were requested (e.g., using automated web path scanners).
- User agents of common scanners, bots or similar tools were employed instead of typical web browsers.

After 56 days of monitoring the first version of the web honeypot service, no attempts to exploit the custom SQL injection or cross-site scripting vulnerabilities were observed. However, activities by several

Table 2. Chronological list of web and onion address announcements on the Internet.

Date	Event
29.12.2018	Initial deployment of web and SSH honeypot services in the clearnet and darknet.
14.01.2019	Announcement of onion addresses in several subforums of Reddit ("subreddits") and the `4chan` bulletin board, announcement of web addresses on `4chan`, sending of web address indexing requests to Google, sending of onion address indexing requests to several Tor search engines.
17.01.2019	Announcement of web addresses as comments on several YouTube videos and subreddits related to cats.
25.01.2019	Announcement of web addresses on different subreddits.
10.02.2019	Posting of the onion addresses on Hidden Wiki and Onion List.
26.02.2019	Submission of web addresses to different subreddits and other link aggregators, announcement of Tor2Web proxy URLs.

tools that scanned web paths were observed, especially requests for common web shells. Based on these observations, the second version of the web honeypot service with additional vulnerabilities (namely, arbitrary file uploads and web shells) was deployed. The second version was monitored for 20 days.

Clearnet. The web honeypot service on the clearnet virtual machine fielded 32,605 HTTP requests (69.6% GET, 30.1% POST, 0.1% HEAD) in 4,157 sessions (i.e., requests from the same IP address within a short time span) during the 76 days of uptime. A total of 2,931 unique IP addresses originating from 113 countries were logged. Application of the criteria listed above classified 79.7% of the sessions as automated (bot) sessions; 67.3% of the web requests were issued in these sessions. Figure 3 shows the distribution of web requests by country.

The logs recorded 7,918 GET requests (24.3% of the total) that were used for web path scanning (i.e., attempting to find hidden files via brute force). Examples of the scanned paths were:

- Database administration software (e.g., `/phpmyadmin`).
- A file used by the Muhstik botnet [19] (i.e., `muhstik.php`).
- WebDAV directories (e.g., `/webdav`).
- VoIP administration software (e.g., `/voip` and `/cisco`).

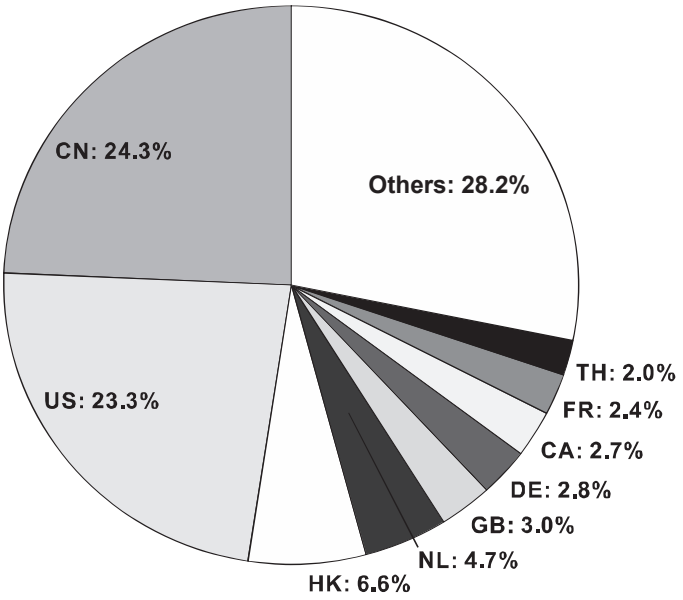

Figure 3. Distribution of 32,605 HTTP requests by country (clearnet).

The logs recorded 6,906 POST requests (21.2% of the total) that were attempting to launch a PHP code injection attack against a non-existent file `/test.php`. The requests originated from 63 IP addresses in 12 countries (10 countries in Asia, the majority from China, Hong Kong and Taiwan). Since all the attacks shared the same pattern, they were likely launched from a botnet in the region.

The logs also recorded 2,873 POST requests (8.8% of the total) involved in brute force attacks against the Wordpress instance. The requests originated from 701 IP addresses in 54 countries. All the requests used the same user agent, an indication that the brute force attacks were launched from a single botnet.

Other attacks included small numbers of attempts to exploit vulnerabilities in D-Link modems [9], ThinkPHP [10] and Avtech IP cameras [7]. Just before the monitoring of the web honeypot service ended, an attacker from an IP address in the United Kingdom accessed the `/shell.php` web shell and manually attempted to gather information. The attacker eventually issued a command to wipe all the system data.

Darknet. The web honeypot service on the darknet virtual machine fielded 112,082 HTTP requests (97.6% GET, 0.1% POST, 2.2% HEAD)

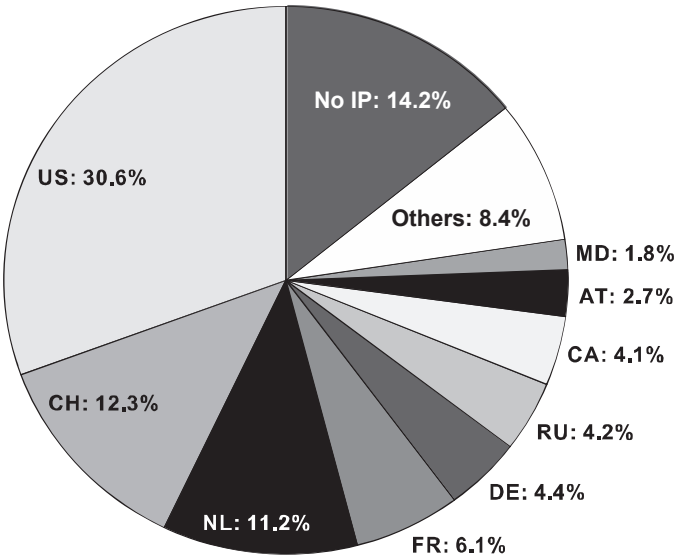

Figure 4. Distribution of 1,728 HTTP requests by country (darknet).

in 7,784 sessions during 76 days of uptime. Of these requests, 24.4% were classified as automated (bot) and 75.6% as manual requests.

The logs recorded 1,728 (1.5%) HTTP requests sent using a Tor2Web proxy. Upon examination, 85.8% of these requests contained HTTP headers that revealed the originating IP address (i.e., headers such as `X-Real-Ip`, `X-Tor2web` and `X-Forwarded-Host`). Most of the requests came from the United States and European countries. Figure 4 shows the distribution of web requests issued using a Tor2Web proxy by country.

The logs recorded 14,879 GET requests (13.3% of the total) that were used for web path scanning. Examples of the scanned paths were:

- Database dump files in various formats (e.g., `mysql.zip`, `dump.sql` and `dump.sql.zip`).
- Private key file of the onion service that can be used to steal the onion domain name (i.e., `/private_key`).
- Files related to cryptocurrencies (e.g., `wallet.zip`, `wallet.dat` and `bitcoin.php`).
- Files and path traversial attempts to gather information about Linux systems (e.g., `/etc`, `/.ssh` and `.bash_history`).
- Web shells (e.g., `/shell.php`).

- WebDAV directories (e.g., `/webdav`).

No attacks using GET requests were logged during the uptime. Several registration and login attempts using the custom forms were attempted, but no attempts to exploit the custom SQL injection or cross-site scripting vulnerabilities were observed.

Upon deploying the second version of the web honeypot server, a manual attack that lasted more than 17 minutes was observed. The attacker located the arbitrary file upload script `upload.php` and used it to upload the `wow.php` web shell. A total of 19 POST requests addressing `wow.php` were observed, each with a different user agent. Finally, the attacker uploaded the `c99.php` shell and modified the permissions of `/var/www/html`, which resulted in the web page becoming inaccessible.

Comparison. The darknet virtual machine had 3.4 times as many HTTP requests as the clearnet virtual machine. However, a much larger proportion of automated traffic/attacks (67.3% of all the HTTP requests) was observed on the clearnet virtual machine compared with the darknet virtual machine (24.4%). The web honeypot service on the darknet virtual machine had more visits likely out of curiosity after the website was announced. In absolute terms, both the web services received similar numbers of automated HTTP requests, about 22K for the clearnet and 27K for the darknet honeypots.

Visiting bots from major search engines such as Google were observed on both web services. The web service on the darknet virtual machine was crawled via a Tor2Web proxy, resulting in both web services being indexed (e.g., by Google). Also, both the web services experienced automated brute force attacks against the installed Wordpress instances, directly targeting specific pages such as `/xmlrpc.php` and `/wp-login.php`. This may indicate that automated vulnerability scanners use search engines to look for vulnerable Wordpress installations in the clearnet as well as the darknet. Path scanners appeared to perform different searches in the clearnet and darknet (e.g., looking for data related to cryptocurrencies), but similar interests were observed with respect to database dumps, administration tools and web shells.

Most attackers addressed the web honeypot service on the clearnet virtual machine using its IP address instead of its domain name. This may be another indication that automated tools were used to scan and attack large IP address ranges. In the case of the clearnet virtual machine, many attacks attempted to target popular software vulnerabilities directly without first checking if the software was actually installed. In the case of the darknet virtual machine, more initial reconnaissance traf-

fic was observed along with interest in darknet-specific data such as the onion service private key.

7.4 Observed SSH and Telnet Access

This section presents details about and comparisons of the SSH and Telnet honeypot service access on the clearnet and darknet virtual machines.

Clearnet. The clearnet virtual machine had incoming traffic to the SSH honeypot just 62 minutes after the initial deployment, even before any announcements were made. During the uptime, a total of 1,194,952 SSH and Telnet sessions (85.6% SSH and 14.4% Telnet) were logged; they originated from 18,345 IP addresses located in 177 countries. About 77.6% of the sessions achieved successful logins by targeting the weak login credentials. About 92% of the login attempts sought to brute force the `root` user password. The most popular password attempted was `root` (18.8%).

The logs recorded about 890K sessions during which attackers attempted to utilize the SSH honeypot as a SOCKS proxy for tunneling their traffic (e.g., to attack other targets). The most popular targets for access using SSH were web (HTTP(S)) and email (SMTP/IMAP/POP3) services. Table 3 lists the most popular targets along with the numbers of SSH sessions during which connections to the targets were attempted using the honeypot as a SOCKS proxy. The majority of the HTTP requests were simple GET requests without payloads or query parameters. These requests were likely issued to check if the SOCKS proxy was working as intended.

Attacker activity on the SSH honeypot was analyzed by examining the commands issued in the SSH sessions. From among the approximately 900K shell commands (10,018 unique commands) issued in 83K sessions, a large number of sessions (33,732 corresponding to 40.5% of all the sessions) had activity patterns associated with the Mirai botnet [15]. The commands fell into six categories:

- **Mirai:** Typical commands known to be issued by the Mirai botnet.
- **Info:** Commands used to gather information (e.g., `ping`, `whoami`, `uptime`, `cat` and `ls`).
- **Harmless:** Commands with no significant effects (e.g., `exit`) and empty commands.
- **Script:** Nested and scripted commands (e.g., multiple actions in a command line).

Table 3. Service connections attempted using the SSH honeypot.

Domain Name/IP Address	Protocol	Sessions
ya.ru	HTTP	301,208
163.172.20.152	HTTP	39,566
bot.whatismyipaddress.com	HTTP	16,643
2a02:6b8:a::a 6574	HTTP	3,249
186.190.212.26	HTTP	2,984
ya.ru	HTTPS	478,923
www.google.com	HTTPS	28,197
www.walmart.com	HTTPS	10,006
www.netflix.com	HTTPS	8,970
www.google.co.uk	HTTPS	7,856
96.114.157.80	SMTP	7,215
68.87.20.5	SMTP	7,201
98.136.101.117	SMTP	7,145
66.218.85.52	SMTP	6,870
67.195.229.59	SMTP	6,870
imap.apple.mail.yahoo.com	IMAP/POP3	11,534
imap.email.comcast.net	IMAP/POP3	4,264
imap.aol.com	IMAP/POP3	4,192
imap.gmx.net	IMAP/POP3	508
imap-mail.outlook.com	IMAP/POP3	482

- **Change:** Commands used to change filesystem permissions, download files, free memory, etc.
- **Other:** Commands not belonging to the five preceding categories.

Table 4. Commands per category issued to the SSH/Telnet honeypot.

Category	Commands	SSH	Telnet	IP Addresses	Countries
Mirai	746,871	20	33,712	2,774	119
Info	140,616	2,104	33,887	4,669	138
Harmless	13,719	26	13,363	71	16
Script	129	67	61	86	25
Change	50	30	1	13	10
Other	32	23	1	6	4

Table 4 shows the numbers of commands in the six categories issued by attackers on the SSH/Telnet honeypot. Associated with each command

category are the numbers of SSH and Telnet sessions and originating IP addresses and countries.

Telnet appears to be a popular service for issuing Mirai botnet and information gathering commands. In 128 sessions (0.2%), attackers used similar numbers of fully-scripted commands on SSH and Telnet (e.g., to download and deploy backdoors or remove traces from log files).

Examination of the files downloaded by the attackers revealed multiple samples that were presumably intended or used to launch large-scale automated attacks. For example, six shell scripts were found to share the same pattern: they first tried to download and then run 10 different executable files of the same program, each compiled for a different processor architecture (e.g., x68, MIPS, PowerPC and ARM).

Darknet. Although the SSH service on the clearnet virtual machine had a lot of activity, the SSH service on the darknet virtual machine received very little attention. In total, 184 sessions (50% SSH and 50% Telnet) were logged and only after the onion address was published for the first time on January 14, 2019. Compared with the clearnet virtual machine, no login attempts were observed on the darknet virtual machine. All 184 sessions appeared to originate from port scanners or banner grabbers.

Comparison. Compared with the darknet virtual machine, almost 6,500 times as many SSH and Telnet sessions were logged on the clearnet virtual machine. While the SSH/Telnet honeypot was subjected to many different attacks such as brute force login attempts, use as a SOCKS proxy and automated botnet attacks, the honeypot on the darknet virtual machine was not attacked at all and received very little reconnaissance traffic.

7.5 Observed SMTP Requests

This section presents details about and comparisons of the SMTP honeypot service access on the clearnet and darknet virtual machines.

Clearnet. During the uptime of the SMTP honeypot, 2,502 sessions were logged on the clearnet virtual machine; they originated from 130 IP addresses in 27 countries. Approximately 90% of the sessions originated from Russia. A closer look of the logged data revealed that they came from the PushDo/Cutwail botnet [24] that attempted to send spam email. However, since the attackers only sent a few packets in most SMTP sessions, only 74 attacker email messages were captured. Of the

74 captured email messages, 46 (62.2%) were empty messages or test messages, the other 28 (37.8%) were phishing email messages.

Darknet. Absolutely no traffic was observed on the SMTP honeypot installed on the darknet virtual machine.

7.6 Observed FTP Requests

This section presents details about and comparisons of the FTP honeypot service access on the clearnet and darknet virtual machines.

Clearnet. The FTP honeypot installed on the clearnet virtual machine had 205 FTP sessions that originated from 131 IP addresses in 24 countries. One hundred of the 205 sessions originated from 76 IP addresses in the United States. Across all sessions, the following activities were observed:

- A total of 86 banner grabbing attempts.
- A total of 16 HTTP GET requests, including four requests issued from IPIP.NET, presumably for research purposes.
- Four attempts to gather information via FTP commands without logging in (e.g., `HELP`, `STAT` and `LIST`).
- A total of 99 login attempts,

Of the 99 login attempts, 41 were successful – 35 as `anonymous` and six as `root`. Of these, 39 attempted to gain information using FTP commands such as `CWD`, `FEAT`, `HELP`, `LIST` and `PWD`. The remaining two authenticated sessions originated from the same IP address in Germany. The attacker logged in as `anonymous`, gathered information about the FTP server and downloaded a public-private key pair that was intentionally stored on the server. The attacker proceeded to create three directories and eventually uploaded a harmless file containing an unrelated OpenVPN traffic dump.

Darknet. Only 35 banner grabbing attempts from 31 unique circuits were logged by the FTP honeypot on the darknet virtual machine. No login attempts or attacks were observed.

Comparison. The clearnet virtual machine had almost six times the amount of FTP session activity seen on the darknet virtual machine. As in the case of the SSH/Telnet honeypot, the FTP honeypot hosted on the darknet did not receive any traffic apart from a few reconnaissance attempts.

7.7 Discussion

Analysis of the collected data revealed that services such as SMTP and FTP are of much less interest to attackers than web, SSH and Telnet services. In the darknet, only web services attracted attackers, which supports the findings of Catakoglu et al. [6].

The majority of attacks on the honeypot services hosted in the clearnet were automated (e.g., from botnets). In contrast, the darknet web service, which was the only darknet service to see any attacks, was targeted mostly via manual means. Additionally, in the darknet, a higher amount of reconnaissance traffic preceded the attack attempts.

Certain similarities were observed between attacks on the clearnet and darknet service deployments. The clearnet and darknet virtual machines both encountered scans for web shells and database-related data as well as brute force attempts that sought to exploit common Wordpress vulnerabilities. No attempts were made to exploit the SQL injection and cross-site scripting vulnerabilities on the clearnet and darknet virtual machines, neither by automated tools nor by manual means.

8. Conclusions

This chapter has described an approach for setting up and simultaneously operating honeypots with vulnerable services to collect information about attackers in the clearnet and darknet. The two systems, which were deployed for 76 days, collected valuable information about attacks conducted by individuals and bots.

Attacks common to the clearnet and darknet deployments focused on Wordpress instances in the web honeypot services. While a large number of automated (mostly botnet) attacks on web services were observed in the clearnet deployment, attacks on web services in the darknet deployment were mostly conducted manually. Interestingly, custom vulnerabilities were neither discovered nor exploited in the clearnet and darknet deployments. In the clearnet, attackers were more interested in web, SSH and Telnet services than SMTP and FTP services. In contrast, only web services attracted attackers in the darknet deployment, which confirms previous research results [6].

Future research will expand the range of the deployed honeypot services by installing additional vulnerable versions of popular content management systems. Additionally, the uptime for data collection will be increased significantly.

Acknowledgement

This research was supported by the German Federal Ministry of Education and Research (BMBF) under the PANDA Project (`panda-projekt.de`).

References

[1] awhitehatter, Mailoney – An SMTP Honeypot, GitHub (`github.com/awhitehatter/mailoney`), 2021.

[2] E. Bou-Harb, M. Debbabi and C. Assi, A time series approach for inferring orchestrated probing campaigns by analyzing darknet traffic, *Proceedings of the Tenth International Conference on Availability, Reliability and Security*, pp. 180–185, 2015.

[3] G. Branwen, Darknet Market Archives (2013–2015) (`www.gwern.net/DNM-archives`), 2019.

[4] A. Bredo, `honeypot-ftp` – FTP Honeypot, GitHub (`github.com/alexbredo/honeypot-ftp`), 2014.

[5] J. Broseus, D. Rhumorbarbe, M. Morelato, L. Staehli and Q. Rossy, A geographical analysis of trafficking in a popular darknet market, *Forensic Science International*, vol. 277, pp. 88–102, 2017.

[6] O. Catakoglu, M. Balduzzi and D. Balzarotti, Attack landscape in the dark side of the web, *Proceedings of the Symposium on Applied Computing*, pp. 1739–1746, 2017.

[7] Cisco Systems, Announcement regarding non-Cisco product security alerts, San Jose, California (`tools.cisco.com/security/center/viewAlert.x?alertId=49625`), 2019.

[8] Cowrie, `cowrie` – Cowrie SSH/Telnet Honeypot, GitHub (`github.com/cowrie/cowrie`), 2021.

[9] Exploit Database, D-Link DSL-2750B – OS Command Injection (`www.exploit-db.com/exploits/44760`), May 25, 2018.

[10] Exploit Database, ThinkPHP 5.X – Remote Command Execution (`www.exploit-db.com/exploits/46150`), January 14, 2019.

[11] C. Fachkha and M. Debbabi, Darknet as a source of cyber intelligence: Survey, taxonomy and characterization, *IEEE Communications Surveys and Tutorials*, vol. 18(2), pp. 1197–1227, 2016.

[12] C. Flamand and D. Decary-Hetu, The open and dark web, in *The Human Factor of Cybercrime*, R. Leukfeldt and T. Holt (Eds.), Routledge, London, United Kingdom, pp. 34–50, 2019.

[13] D. Fraunholz, D. Krohmer, S. Anton and H. Schotten, Investigation of cyber crime conducted by abusing weak or default passwords with a medium interaction honeypot, *Proceedings of the International Conference on Cyber Security and Protection of Digital Services*, 2017.

[14] D. Kavallieros, D. Myttas, E. Kermitsis, E. Lissaris, G. Giataganas and E. Darra, Understanding the dark web, in *Dark Web Investigations*, B. Akhgar, M. Gercke, S. Vrochidis and H. Gibson (Eds.), Springer, Cham, Switzerland, pp. 3–26, 2021.

[15] C. Kolias, G. Kambourakis, A. Stavrou and J. Voas, DDoS in the IoT: Mirai and other botnets, *IEEE Computer*, vol. 50(7), pp. 80–84, 2017.

[16] D. Moore, C. Shannon, G. Voelker and S. Savage, Network Telescopes: Technical Report, Cooperative Association for Internet Data Analysis, San Diego Supercomputer Center, University of California San Diego, La Jolla, California, 2004.

[17] J. Moubarak and C. Bassil, On darknet honeybots, *Proceedings of the Fourth Cyber Security in Networking Conference*, 2020.

[18] K. Nakao, D. Inoue, M. Eto and K. Yoshioka, Practical correlation analysis between scan and malware profiles against zero-day attacks based on darknet monitoring, *IEICE Transactions on Information and Systems*, vol. E92-D(5), pp. 787–798, 2009.

[19] L. O'Donnell, Muhstik botnet exploits highly critical Drupal bug, *Threatpost*, April 23, 2018.

[20] OWASP Foundation, OWASP Top Ten, Bel Air, Maryland (`owasp.org/www-project-top-ten`), 2020.

[21] J. Song, J. Choi and S. Choi, A malware collection and analysis framework based on darknet traffic, in *Neural Information Processing*, T. Huang, Z. Zeng, C. Li and C. Leung (Eds.), Springer, Berlin Heidelberg, Germany, pp. 624–631, 2012.

[22] J. Song, J. Shimamura, M. Eto, D. Inoue and K. Nakao, Correlation analysis between spamming botnets and malware-infected hosts, *Proceedings of the International Symposium on Applications and the Internet*, pp. 372–375, 2011.

[23] M. Steinebach, M. Schafer, A. Karakuz, K. Brandl and Y. Yannikos, Detection and analysis of Tor onion services, *Proceedings of the Fourteenth International Conference on Availability, Reliability and Security*, article no. 66, 2019.

[24] B. Stone-Gross, T. Holz, G. Stringhini and G. Vigna, The underground economy of spam: A botmaster's perspective of coordinating large-scale spam campaigns, *Proceedings of the Fourth USENIX Conference on Large-Scale Exploits and Emergent Threats*, 2011.

[25] A. Vetterl and R. Clayton, Honware: A virtual honeypot framework for capturing CPE and IoT zero days, *Proceedings of the APWG Symposium on Electronic Crime Research*, 2019.

[26] R. Zeid, J. Moubarak and C. Bassil, Investigating the darknet, *Proceedings of the International Wireless Communications and Mobile Computing Conference*, pp. 727–732, 2020.