Poster: The State of Malware Loaders

Cristian Munteanu
Max Planck Institute for Informatics
Germany, Saarbrücken

Georgios Smaragdakis
Delft University of Technology
Max Planck Institute for Informatics
Netherlands, Delft

Anja Feldmann
Max Planck Institute for Informatics
Germany, Saarbrücken

ABSTRACT

Malware is recognized as one of the most severe cybersecurity threats today. Although malware attacks are as old as the Internet, our understanding of which part of the Internet infrastructure is used to distribute malware software is still rather limited.

In this work, we analyze more than 3 million sessions established with honeypots deployed in 55 countries that are associated with the download and execution of malware binaries. We identify two main tactics to load malware to infected machines: injection of malware by hosts initiating the connection (clients) and downloading malware from third parties (loaders). The latter tactic contributes to more than 80% of this class of sessions but involves a smaller number of cloud and content delivery servers with very different profiles than that of the clients. Our analysis also shows that it is not uncommon for different malware families to rely on the same hosting infrastructures for downloading malware. Further investigation into the code executed to download and activate malware shows that criminals tend to hide their traces by deleting their history and modifying logs and files on the compromised machines.

CCS CONCEPTS

• **Security and privacy → Network security**.

ACM Reference Format:
Cristian Munteanu, Georgios Smaragdakis, and Anja Feldmann. 2024. Poster: The State of Malware Loaders. In *Proceedings of the 2024 ACM Internet Measurement Conference (IMC '24), November 4–6, 2024, Madrid, Spain.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3646547.3689659

1 INTRODUCTION

Malware is malicious software that refers to any intrusive software developed by cybercriminals. There are many facets of malware software. It can be used to steal personal or enterprise data and credentials, that takes the form Trojan [9] or self-propagate to infect and paralyze computing systems, in this case it is a worm, or be used to weaponize an infected host to launch distributed denial of service attacks, e.g., Mirai [2].

For more than two decades, honeypots have been very successful to shed light on security risks in networks and Internet-facing systems, e.g., [8, 13], identify new variants of cyber-threats, e.g., [7, 10], and investigate new attack techniques and tactics, e.g., [2, 5]. Honeypots emulate an operational computing system so intruders unveil their behavior and tactics and log all the interactions in a database.

IMC '24, November 4–6, 2024, Madrid, Spain

ACM ISBN 979-8-4007-0592-2/24/11.
https://doi.org/10.1145/3646547.3689659

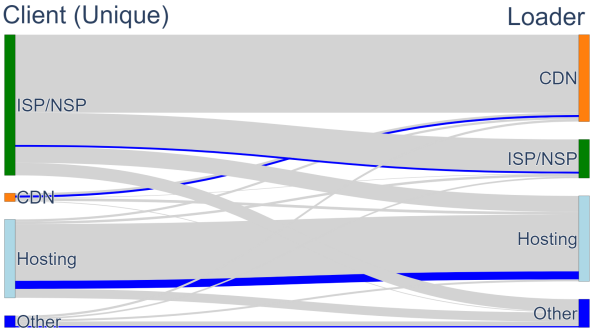

Figure 1: Sankey plots: # of client IPs by AS type vs. # loader IP by AS type. Blue lines correspond to sessions where the client IP and the loader IP are the same.

In a recent paper, Munteanu et al. [6] got access to a honeyfarm (a large collection of honeypots) operated by Global Cyber Alliance (GCA) [4] and analyzed honeypot logs for the first fifteen months of its operation. The authors conclude that the honeypots in the honeyfarm have a complementary view of scanning and intrusion activity. Indeed, even the honeypots with the largest number of observed hashes contribute less than 10% of the overall hashes in the honeyfarm.

In this work, we get access to the same GCA honeyfarm, but we focus on the profile, demographics, and tactics of the clients that download and execute malware as observed by the honeyfarm for over twenty-eight months. This type of malicious activity is the most severe type of attack. Indeed, the clients do not just scan for potentially open vulnerable ports but take actions to compromise the hosts by installing and executing malicious code to get access to the host or create harm, involving in many cases third-party hosts as we show in this study.

2 ANALYSIS

Once a client is connecting to a honeypot there are two possible ways to load the malware: either the client loads the malicious file from its own location or it is using a different server. In the later case, we call the server that loads the file a *loader*, while in the former we say that the *client* is also the *loader*.

To better understand the infrastructure the attackers use, we categorize each client and loader IP according to the AS that announces it. For this purpose, we use a service for looking up historic announcement information for IPv4 [11]. This service returns a historical perspective for each IP and timestamp, including the announcement time period, AS number, and AS organization details.

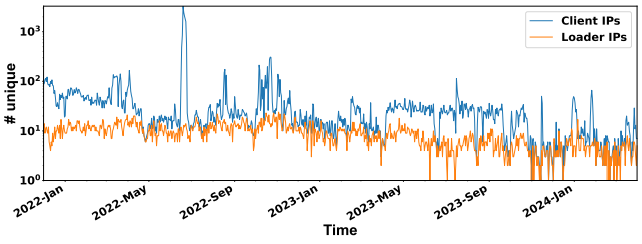

Figure 2: Number of unique loaders IPs and client IPs across time.

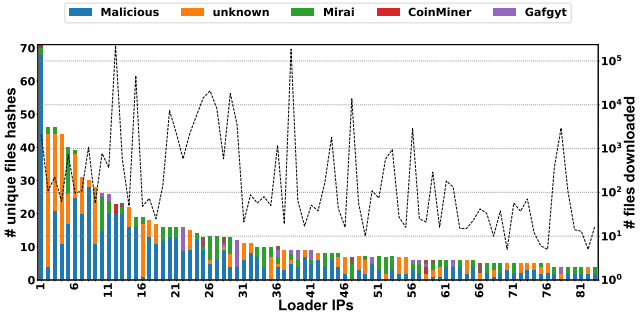

Figure 3: Loader IPs hosting files related to different attack types.

To further categorize the returned ASes, we use bgp.tools [3] as well as PeeringDB. Hereby, we distinguish the following types:

CDNs : Content Delivery Networks.

Hoster : Hosting providers (including web-hosting, VPN, etc.).

ISPs/NSPs : Internet Service providers.

Others : Other types of networks (including governmental, academic, personal networks).

We then group the client IPs as well as the loaders by the AS (network) type and use the AS type to generate a sankey diagram of the sessions, see Figure 1. On the left side we show the client IPs of the session and on the right side the loader IPs. The size of each flow corresponds to the number of unique IP pair. We color those flows blue where the client IP equals the loader IP and those where they are unequal gray.

To better understand how the loader activity changes across time Figure 2 plots the number of unique loader IPs per day across time using a log y-scale. We notice that on average the number of *loaders* (10 unique loader IPs/day) is much smaller than the number of *clients* (more than 50 unique client IPs/day) connecting to the honeypot.

We classify the downloaded malware in different attack categories using public tools such VirusTotal [12] and abuse.ch [1]. We identify 83 loader IPs that are used to retrieve files from different categories to the honeypots. Figure3 shows a barplot with the 83 loader IPs on the x-axis and the number of unique hashes that are loaded from them to the honeyfarm on the y-axis (sorted by # of unique hashes). The colors of the bars indicate the attack category of the file. Using a secondary y-axis (log scale) we also add the number of session per loader IP.

```
uname -a;id;
cat /etc/shadow /etc/passwd; lscpu;
echo 'daemon ALL=(ALL) NOPASSWD: ALL' >> /etc/sudoers;
chsh -s /bin/sh daemon;
echo Password123 |passwd daemon --stdin;
chattr -ia /root/.ssh/*;
wget http://<X.X.X.X>/ns1.jpg -O ~/.ssh/authorized_keys;
chmod 600 ~/.ssh/authorized_keys;
wget -qO - http://<X.X.X.X>/ns2.jpg|perl;
wget http://<X.X.X.X>/ns3.jpg -O /tmp/x;
chmod +x /tmp/x;
/tmp/x; mv /tmp/x /tmp/o;
/tmp/o; rm -f /tmp/o;
```

Figure 4: Attack with false file extension. We replace the actual IP with "<X.X.X.X>".

It is not uncommon that the attackers hide behind harmless and false file extension suffixes, i.e., the name is *file.jpg* even though the content of the file are shell commands and should be considered a *sh* file, see Figure 4 for an example. We also find that most attackers remove files. Overall, this is true for 81% of the sessions. Often attackers do not only delete files, but also clear the history. It is fairly rare and unexpected that attackers only clear the history but do not remove files, yet we see 0.6% of such sessions.

Acknowledgments

This work was supported in part by the European Commission under the Horizon Europe Programme as part of the project Safe-Horizon (Grant Agreement no. 101168562). The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed therein lies entirely with the authors.

REFERENCES

[1] abuse.ch. 2024. abuse.ch: Fighting Malware and Botnets. https://abuse.ch/.
[2] Manos Antonakakis, Tim April, Michael Bailey, Matt Bernhard, Elie Bursztein, Jaime Cochran, Zakir Durumeric, J. Alex Halderman, Luca Invernizzi, Michalis Kallitsis, Deepak Kumar, Chaz Lever, Zane Ma, Joshua Mason, Damian Menscher, Chad Seaman, Nick Sullivan, Kurt Thomas, and Yi Zhou. 2017. Understanding the Mirai Botnet. In *USENIX Security Symposium*.
[3] Basil Fillan, Ben Cartwright-Cox, Cynthia Revström, Elimalko Saado, eraters, Igloo22225, Jeroen Massar, Job Snijders, Molly Miller, Puck Meerburg, Roelf Wichertjes, Tim Stallard, Tommy Bowditch. 2023. BGP.tools. https://bgp.tools/.
[4] Global Cyber Alliance. 2023. GCA AIDE – Automated IoT Defense Ecosystem. https://www.globalcyberalliance.org/.
[5] Harm Griffioen and Christian Doerr. 2020. Examining Mirai's Battle over the Internet of Things. In *ACM CCS*.
[6] Cristian Munteanu, Said Jawad Saidi, Oliver Gasser, Georgios Smaragdakis, and Anja Feldmann. 2023. Fifteen Months in the Life of a Honeyfarm. In *ACM IMC*.
[7] Marcin Nawrocki, John Kristoff, Raphael Hiesgen, Chris Kanich, Thomas C. Schmidt, and Matthias Wählisch. 2023. SoK: A Data-driven View on Methods to Detect Reflective Amplification DDoS Attacks Using Honeypots. In *IEEE Euro S&P*.
[8] Niels Provos. 2004. A Virtual Honeypot Framework. In *USENIX Security Symposium*.
[9] Mohammad Rezaeirad, Brown Farinholt, Hitesh Dharmdasani, Paul Pearce, Kirill Levchenko, and Damon McCoy. 2018. Schrödinger's RAT: Profiling the Stakeholders in the Remote Access Trojan Ecosystem. In *USENIX Security Symposium*.
[10] Christian Rossow. 2014. Amplification Hell: Revisiting Network Protocols for DDoS Abuse. *NDSS* (2014).
[11] Florian Streibelt, Martina Lindorfer, Seda Gürses, Carlos H Gañán, and Tobias Fiebig. 2023. Back-to-the-Future Whois: An IP Address Attribution Service for Working with Historic Datasets. In *International Conference on Passive and Active Network Measurement*. Springer, 209–226.
[12] VirusTotal. 2024. VirusTotal. https://www.virustotal.com/.
[13] Vinod Yegneswaran, Paul Barford, and Vern Paxson. 2005. Using honeynets for Internet situational awareness. In *Proceedings of the Fourth Workshop on Hot Topics in Networks (HotNets IV)*.