Master Project WiSe 2020/21

Detecting Adversarial Techniques and Exploits

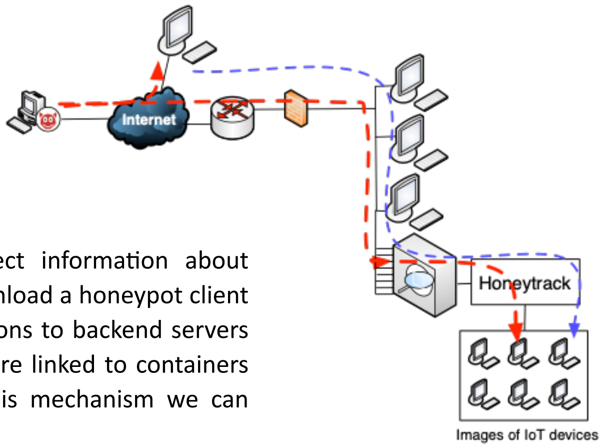
FG Cybersecurity and Enterprise Security, Prof. Dr. Christian Dörr

Background

In order to launch an effective defense, information about malicious actors and how they work is essential.

For example, if I can obtain a real-time view into what vulnerabilities are currently being used on the Internet to exploit hosts, I can check whether I operate any computers in my organization that are potentially affected by such vulnerability. Furthermore, if I can get insights how particular adversaries work — for example they brute force passwords, then install a malicious program on the compromised host and from there scan the local network for additional victims -, I can use this knowledge to select better defenses or develop highly effective signatures to detect a compromise.

The Cybersecurity and Enterprise Security group operates a distributed honeypot infrastructure, aimed to safely collect information about adversarial activities. Companies and private citizens can download a honeypot client that opens a local port, and forwards any incoming connections to backend servers of the HPI. In the backend, the requests of the adversaries are linked to containers running one ore more vulnerable services, and through this mechanism we can observe adversarial scanning and exploitation attempts.



Project Goal

In this project, we are going to leverage the data collected by this infrastructure to obtain in-depth information about adversarial techniques. As we can dynamically connect container images to incoming requests, the goal is to link vulnerable and misconfigured software to a new connection, and from the attack traffic try to establish which exploit is trying to be exploited and also which tools the attacker is using. For this we will collect exploits and proof-of-concepts from sources such as metasploit or pastebin, and develop a framework to automatically derive network-level signatures from the PoC code that can match the used payloads and packets to incoming exploitation attempts.

As the infrastructure presents an actual full-stack system towards the outside, adversaries cannot easily identify it as a decoy and our data shows that nearly all attackers continue after the exploit to take full ownership of the machine. While clustering based on connection meta-data such as outgoing requests gives a first angle to quantify what happens after the compromise, we would like to obtain a deeper, conceptual understanding of what the adversary is doing on the victim system, for example establishing persistence, installing a backdoor or initiating scanning. In the second part of this master project we will thus implement a behavioral machine learning approach to analyze activities on the VM and classify them along common taxonomies such as the MITRE ATT&CK framework for a fully automatic characterization of adversarial techniques.

Contact Information

Prof. Dr. Christian Dörr	christian.doerr@hpi.de	Haus III, G-3.1.09
Harm Griffioen	harm.griffioen@hpi.de	Haus III, G-3.1.12