

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies

FIIT-XXXXXX-82385

Tomáš Belluš

**Bait network based monitoring of
malicious actors**

Master's thesis

Supervisor: Ing. Tibor Csóka, PhD.

2021, January

Slovak University of Technology in Bratislava
Faculty of Informatics and Information Technologies

FIIT-XXXXXX-82385

Tomáš Belluš

**Bait network based monitoring of
malicious actors**

Master's thesis

Study program: Information Security

Field of study: 9.2.4 Computer Engineering

Training workplace: Institute of Computer Engineering and Applied Informatics

Supervisor: Ing. Tibor Csóka, PhD.

Departmental Advisor: Ing. Katarína Jelemenská, PhD.

2021, January

Annotation

Slovak University of Technology Bratislava

FACULTY OF INFORMATICS AND INFORMATION TECHNOLOGIES

Degree Course: Information Security

Author: Tomáš Belluš

Master's Thesis: Bait network based monitoring of malicious actors

Supervisor: Ing. Tibor Csóka, PhD.

Departmental advisor: Ing. Katarína Jelemenská, PhD.

2021, January

Internal network, demilitarized zone (DMZ) or data pipelines have been compromised by a threat actor and the information gathered from this incident is minimal. Knowing the threat actor's agenda (entering, potentially leaving and fulfilling a goal) is more valuable, because it may lead to enforcing the system perimeter or endpoint security. By deliberately baiting access to highly monitored isolated networks, all further activities may be learned and enlighten a security engineer. This thesis, by utilizing state of the art container orchestration mechanism - Kubernetes, designs and implements a monitored isolated environment. Understanding and logging all file system changes, process executions helps to create a timeline of events constructing a possible incident. The resulting implementation is a robust proof of concept virtualized and completely automated immutable infrastructure monitored on the host machine level. Container observation mechanisms are capable, but not accurate enough to yet determine the order of file system events. Further design and implementation is required to identify the *point of enter* and *exit*.

Anotácia

Slovenská technická univerzita v Bratislave

FAKULTA INFORMATIKY A INFORMAČNÝCH TECHNOLOGIÍ

Študijný program: Informačná bezpečnosť

Autor: Tomáš Belluš

Diplomová práca: Sledovanie zlovoľných činiteľov nástražným systémom

Vedúci diplomovej práce: Ing. Tibor Csóka, PhD.

Pedagogický vedúci: Ing. Katarína Jelemenská, PhD.

Január 2021

Interná sieť, demilitarizovaná zóna DMZ alebo zreťazené spracovanie údajov sú kompromitované útočníkom, a získané informácie o tomto incidente sú minimálne. Vedieť útočnickovú agendu (od vstup cez vykonanie agendy až po prípadný výstup) je veľmi cenné, lebo to môže viesť k vylepšeniu informačnej bezpečnosti sieťového okruhu alebo cieľových staníc. Úmyselným lákaním útočníkov na prístup k vysoko sledovaným a izolovaným sieťam, poskytuje možnosť pre bezpečnostného experta sa poučiť zo zlovoľných aktivít. Diplomová práca, využitím vyspelého orchestračného mechanizmu kontajnerov - *Kubernetes*, sa venuje návrhu a implementácii sledovania izolovaného prostredia. Porozumenie a zaznamenávanie všetkým zmenám súborového systému a vykonávania procesov napomáha vytváraniu časovej osi udalostí spájaných do prípadného incidentu. Výsledná implementácia je rozsiahlym dôkazom predstavy ako kompletne virtualizovaná a automatizovaná nemenná infraštruktúra monitorovaná na úrovni hostiteľa. Mechanizmy sledovania kontajnerov sú funkčné, ale nie sú dostatočne presné v udávaní poradia výskytu udalostí o zmene v súborovom systéme. Preto je nevyhnutný ďalší návrh a implementácia aj pre konkrétne identifikovanie *bodov vstupu a výstupu* zo systému.

Tu vložiť zadanie diplomovej práce

Potom, vložiť finálny návrh zadania diplomovej práce

Contents

| | | |
|----------|---|------------|
| 1 | Analysis | xv |
| 1.1 | Malware analysis | xv |
| 1.1.1 | Dynamic and static | xvi |
| 1.1.2 | Mechanisms [and environments and tools] | xvii |
| 1.1.2.1 | Honeypot and honeynet | xvii |
| 1.1.2.2 | Sandbox | xviii |
| 1.2 | Virtualization technology | xviii |
| 1.2.1 | Kernel Virtual Machine | xviii |
| 1.2.2 | Kubernetes | xviii |
| 1.3 | Environment monitoring | xviii |
| 1.3.1 | Existing solutions | xix |
| 2 | Related Work | xxi |
| 2.1 | Malware file analysis solutions | xxi |
| 2.1.1 | Cuckoo sandbox | xxii |
| 2.1.2 | Droidbox | xxii |
| 2.1.3 | Virustotal | xxiii |
| 2.1.4 | Falcon sandbox | xxiii |
| 2.2 | Active analysis | xxiv |
| 2.2.1 | Honeystat | xxiv |

| | | |
|----------|--|---------------|
| 2.2.2 | Honeypoint | xxv |
| 2.2.3 | Cybertrap | xxvi |
| 2.2.4 | A distributed platform of high interaction honeypots and experimental results | xxvi |
| 2.2.4.1 | Experiment | xxvii |
| 2.2.5 | SIPHON | xxviii |
| 3 | Design | xxxi |
| 3.1 | Container monitoring | xxxi |
| 4 | Implementation | xxxiii |

Chapter 1

Analysis

This chapter introduces malware analysis (see section 1.1) and differentiates the techniques (see subsection 1.1.1), outlines and describes mechanisms and environments utilized by cyber security professionals and companies (see subsection 1.1.2), describes the virtualization technology KVM (see subsection 1.2.1) and the orchestration mechanism Kubernetes (see subsection 1.2.2).

1.1 Malware analysis

Monitoring a malicious actor, sample or an entity in any environment is interchangeable from the analyst point of view. Therefore, knowing the malware analysis techniques is crucial for secure monitoring of malicious actors. Malware analysis may be static and dynamic with varying tools and mechanisms [3] [2].

1.1.1 Dynamic and static

Dynamic analysis is a process of actively monitoring, ideally in an isolated environment, the execution of a malware sample. Static analysis, as the name implies, inspects the function calls, readable strings, control and data flow of a malware sample (i.e. binary, program code).

Dynamic analysis may be unsafe and devastating, unless environment or system isolation is applied, but it's more precise and bypasses the reversing of self-modified code. Even though, dynamic analysis does not explore all execution paths, there are techniques resolving this drawback (i.e. virtual machine snapshots [2]). Furthermore, dynamic analysis in a isolated or virtualization environment is exposed to the risk of [isolation] detection [0].

Static analysis is safe [and inspects all execution paths of the binary], but may be difficult to interpret when malicious actors utilize the obfuscation, compression or encryption techniques. Advanced malware sample poses a challenge for a malware analyst due the usage of control-flow flattening [5] and other methods. It suggests that dynamic malware analysis bypasses this troublesome procedure and discloses the sample outcome or agenda.

Manual Egele, et al. in their article [3] highlight the problems of static malware analysis approaches. The authors introduce multiple state-of-the-art techniques and tools dedicated to dynamic malware analysis - Function Call Monitoring, Function Parameter Analysis focusing of the final values, Information Flow Tracking, Instruction Trace and Auto start Extensibility Points (monitor startup programs, cron jobs, etc.).

1.1.2 Mechanisms [and environments and tools]

These mechanisms must provide solution to dynamic malware analysis limitations in order to effectively capture the malicious actor's agenda. It should be a robust system implying the impression of a production environment.

1.1.2.1 Honeypot and honeynet

Honeypot is a bait service, system or a even whole network (honeynet) usually hosted on public server. Its main purpose is to be scanned, attacked or compromised by the malicious actor. Every honeypot provides the desired functionality of the target resource, to mimic the production environment, leaving the malicious actor unaware of the honeypot [4].

Based on the ENISA honeypot study [4], honeypots are classified from the level of interaction view and based on the attacked resource type. The Low Interaction Honeypot (LIH) provides very low availability of the host OS. Most services and application are mocked and simulated in a static environment. Everything accessible is controlled by a decoy application with absolutely minimal in-depth features (e.g. shell, configuration files, other programs etc.). LIHs are more secure for the host, but far less capable or useful for malware/attack detection and inspection.

On the contrary, High Interaction Honeypot (HIH) is a fully responsive system with live applications and services with minimal to none emulated functionalities. It provides the attacker a wide attack surface ranking the HIH far less secure with the whole OS at the malicious actor's disposal. The idea is to make HIHs believable as possible and isolating it from production environment including virtualization [6].

Honeypots are differentiated by the type of attacked resource. The server-side honeypot is the well-known honeypot with running service(s) and monitoring the activity of the server-side connections. The attacked resources are the services listening on the dedicated ports. It's main purpose is to detect and identify botnets and forced authentication/authorization attempts.

The client-side honeypot is deployed as a user application, which utilizes the server's services. The monitored subject is the application (e.g web-browser, document editor). It's main purpose is to detect client-side attacks originating from the application (i.e. web-browser attacks via web pages and plugins).

1.1.2.2 Sandbox

1.2 Virtualization technology

1.2.1 Kernel Virtual Machine

1.2.2 Kubernetes

1.3 Environment monitoring

Knowing what and how to extract and monitor is vital to understanding the threat actor and the agenda. There are various mechanisms and possibilities to effectively observe container file system, networking and process execution. Regarding the expected setup of virtual machines hosting the Kubernetes cluster, the possible monitoring strategies are:

- from the container in form of an agent reporting events

- from the node as agent-less programs "spying" on the containers

Container monitoring from the container itself may appear more straightforward, but it's similar to monitoring of a honeypot, where agents could be discovered or in other way uncovered by the threat actor. Since containers are wrapped environments hosted on the host machine, the file system, networking processes are readable. The apparent advantage is the transparency for the threat actor, since the container is observed from the hosting operating system. For example docker with the *overlayfs* driver has a predefined directory for each container with the root FS.

1.3.1 Existing solutions

Monks ¹ works as a kernel module hijacking system calls on the target system. "Monks is like strace, but tracing all and every single process from any user, at any level" [0]. **execmon** ² is a similar utility intercepting the *execve* syscalls in kernel and notifying the user. It's a kernel module combined with a user-land utility receiving events. For file system (FS) monitoring there is a CLI program **fswatch** ³ acting as the API to the *libfswatch*, which is ultimately an API to *inotify*. It provides real time FS monitoring with the accuracy of one second. Fswatch distinguishes events based on the action (e.g. created, renamed, removed, owner changed).

¹<https://github.com/alexandernst/monks>

²<https://github.com/kfiros/execmon>

³<https://github.com/emcrisostomo/fswatch>

Chapter 2

Related Work

Honeypot, sandbox and deception technology make up the leading techniques in dynamic malware analysis. They differ in the scope of knowledge before the analysis at hand. Some know the filename, malware type, other artifacts and possible the expected outcome (in which case the tool observes the behavior). Other analyze the behavior of all activity - searching for anomalies and malicious behavior indicators. The following sections briefly introduce the malware analysis, deception technology, honeypot and other existing solutions and studies related to the thesis topic.

2.1 Malware file analysis solutions

Malware file analysis is a dynamic procedure when the filename and malware type is known. In comparison to the scope of this thesis, all use similar techniques of malicious activity observation/monitoring, but differ in use case scenarios.

2.1.1 Cuckoo sandbox

A most common sandbox environment for malware analysis by executing a given file in a sandboxed environment with reporting of the outcome. All files affecting mainstream operating systems i.e. Windows, MacOS, Linux and Android are supported. In addition to known artifacts, cuckoo has no interfering processes, so all traces must be followed and provide insight to the behavior. Based on the official cuckoo documentation, the system produces various results (the following artifacts are copied from the documentation site):

- Traces of calls performed by all processes spawned by the malware.
- Files being created, deleted and downloaded by the malware during its execution.
- Memory dumps of the malware processes.
- Network traffic trace in PCAP format.
- Screenshots taken during the execution of the malware.
- Full memory dumps of the machines.

Despite all differences, cuckoo's architecture consists of the management software (host machine) and a number of virtual/physical machines for analysis. It's a tool for different use case, so a comparison is insignificant.

2.1.2 Droidbox

Another open source tool, sadly discontinued several years ago, droidbox utilizes analyzes android applications using Android Virtual Devices (AVD) and the android emulator 4.1.1_rc6, which enables the android activity monitoring. Analyzed applications are sandboxed in the AVDs and afterwards reports the following results (the following artifacts are copied from the documentation site):

- Hashes for the analyzed package
- Incoming/outgoing network data
- File read and write operations
- Started services and loaded classes through DexClassLoader
- Information leaks via the network, file and SMS
- Circumvented permissions
- Cryptographic operations performed using Android API
- Listing broadcast receivers
- Sent SMS and phone calls

Droidbox introduces a simple way of analyzing android applications via an existing API of the emulator.

2.1.3 Virustotal

Similarly to cuckoo, virustotal utilizes both static and dynamic malware analysis. "VirusTotal's aggregated data is the output of many different antivirus engines, website scanners, file and URL analysis tools, and user contributions" [0].

2.1.4 Falcon sandbox

A direct concurrency to virustotal is the **Hybrid Analysis**¹ tool powered by the Falcon sandbox. Again, it's similar to cuckoo, except the anti-evasion feature [0], which allows, even sandbox-aware malware, to be analyzed despite their evasion techniques.

¹hybrid-analysis.com

2.2 Active analysis

This section explores existing honeypot/honeynet technologies and a recently emerged concept - deception technologies. These technologies may be divided into two categories - dynamic [0] and static, where the environment adapts to the scenarios or remains unchanged respectively.

2.2.1 Honeystat

Honeystat ² is a honeypot solution observing the behavior of the Blaster worm and may be used to detect zero day worm threats. The authors assume the infection may be described in a systematic way, so by knowing the worm agenda and steps they model the monitoring procedure. The observation is event-based with memory, disk and network events. Since there are no regular users in the system, the memory events are e.g. interesting violations as buffer overflows and other. Disk events are file system modifications and network events should always be infection related outgoing traffic. Worms require a multi-host network to have spreading possibility, so honeystat is deployed in a multihomed VMWare environment (64 VMs * 32 IP addresses = 2^{11} IP) with minimal honeypots.

The procedure when events are encountered is:

- The honeystat is capturing memory and disk events
- If a network event occurs, the honeypot is reset to stop further spread of the worm to other machines/honeypots.
- Any previous memory/disk event is updated with additional information from the network event.

²<https://people.engr.tamu.edu/guofei/paper/honeystat.pdf>

- Resets ought to be faster in virtual environment. Host VM is not rebooted, only the virtual disk (VD) is kept in a suspended state before it's replaced with a fresh copy of a VD. The reset always completes before a TCP timeout.
- Other steps include an analysis node, which is out of scope of this thesis.

This solution does not introduce any isolation techniques beside utilizing virtualization and the emulation mechanisms are exposing the virtualized environment via e.g. BIOS strings or MAC address. All features and considerations for honeystat are purely for worm infection detection, other infection types could require more observables.

2.2.2 Honeypoint

Service emulation is what Honeypoint ³ utilizes to lure malicious actors and detect their agenda. Production services lie in the same environment as the robust architecture of Honeypoint, which can mimic a complex network environment for deceiving an attacker. The Microsolved CEO Brent Huston claims [0] that having a honeypot is a great deception technology with almost no false positives, since it is expected that no legitimate user interacts with it. It means that any recorded activity should be considered suspicious, if the honeypot targets malicious actors scanning the Internet regardless of possible domain - randomly trying IP addresses and looking for a services ought to have malicious intent. Consists of various components [0] that could be replicated in the Kubernetes architecture design.

³<https://www.microsolved.com/honeypoint>

2.2.3 Cybertrap

A purely documented (commercial) solution Cybertrap [0] operates as a deception technology luring attackers away from production systems. Looking apart from that services in Honeystat are emulated, Cybertrap's deployed services cannot be distinguished by the attacker. Once the malicious actor gets inside such network, all his/her movements are tracked. In addition, the Cybertrap's network is inaccessible by regular users, so any activity within the simulated environment is considered malicious - minimal to none false positives. Cybertrap is close to the idea of the goal of this thesis - sandboxed honeynet.

2.2.4 A distributed platform of high interaction honeypots and experimental results

A case study [0] serving as a proof of concept in live Internet traffic observing malicious actors' trends and agenda. As a monitoring technique they patched the kernel's `tty` and `exec` modules to intercept the keystrokes and system calls respectively. The architecture is 4 machines anywhere in the world working as relays to the authors' local setup of VM honeypots. The traffic incoming to the public interface of the relay is routed to a GRE tunnel connected to the local VM.

In a SSH scenario they created a new syscall and modified the SSH server to use it in order to intercept the login credentials. Logged data is periodically copied from the VM disk to the host disk (such extractions should be undetected by the malicious actors). All login data is stored to the database of this structure:

- data from each ssh login attempt
- data from each successful ssh connection - tty buffer content and tty name

- data of programs executed with parameters and the terminal in which it ran
- session data grouping ssh connections

2.2.4.1 Experiment

- in the period of 30 days, they monitored what are the most common login-password pairs when no accounts are created
 - they found that for most attempts the login and password were the same
- then for almost half a year they monitored the time it took the attacker to successfully login and to login with commands entered
 - in some cases the attacker managed to get root via system vulnerable exploit
- they encountered attackers changing passwords of other accounts on the system
- they sorted their findings by country (mostly China, USA, Germany, UK, Russia, Romania, Japan, Brazil, France, South Korea and Netherlands)
- analyzed the intrusions and commands
 - mostly they tried to download programs from the same country the source IP originated from
- general trends of attacker behaviors:
 - check if i am alone on the system
 - system recon - OS name and version, processor characteristics
 - changes the password of current user
 - install an IP scan program and scans the IP range to recon for potential lateral movement
 - IRC client setup for receiving instructions
 - privilege escalation attempt

- general trends of attacker behaviors (with root):
 - change the root password
 - setup backdoor - open another port
 - checkout info about legitimate users of the computer via custom installed software
 - one attacker replaced the ssh client binary

2.2.5 SIPHON

A case study [0] on Scalable High-Interaction Physical Honeypots (SIPHON), similar to the study before, serving as a proof of concept in live Internet traffic observing the IOT related malicious intents. Leveraging Shodan to appear visible and legitimate in the eyes of malicious actors, the honeypots were based on real devices. The architecture is divided into physical IOT devices, wormholes exposed to the Internet forwarding to the IOT devices via the proxy forwarder. Technically, devices are separated using VLANs 802.1Q and the wormhole to forwarder connection is via reverse ssh tunnels. As compromise countermeasures, the Suricata IPS and IDS features are enabled in the local network and periodic resets of IOT devices.

They observed the influence of device listing in Shodan. The number of scans/connection attempts on the device has tripled between ‘one week before listing’ and ‘one week after listing’. It proves that being visible by Shodan increases the possibility of attack reconnaissance on device at hand. Although, after two weeks after listing in Shodan, the connection attempts have decreased, which is good knowledge before implementation.

Chapter 3

Design

3.1 Container monitoring

Monitoring a container means to effectively observe container file system, networking and process execution. Inspired by some related solutions, this section describes the monitoring mechanisms utilized in this thesis.

- Points of enter, such as honeypots that lure the threat actors to the environment. Could be local to the environment or remote anywhere in the Internet.
- 3 Ubuntu server nodes are the base to Kubernetes cluster holding and orchestrating the whole environment.
- There are to be multiple environments.
- Any environment is automated and deployed to the cluster via Ansible playbooks.
- The core monitoring tools and programs are deployed on the hosting nodes

Chapter 4

Implementation

- Manage base setup of Kubernetes nodes via Vagrant.
- Deployment of those Vagrant based virtual machines (including networking setup) and Kubernetes (via kubespray ¹) is done by Ansible.
- Kubernetes environments are maintained ² separately.
- Sneakpeek ³ properly parses the docker container information. It groups nodes with user defined containers created by the Kubernetes orchestration from deployments and pods. Ultimately used for getting the container root FS directory.
- Fswatch ⁴ is a fork with simple improvements and features. It's ran on every acquired container root directory.
- All is automated with Ansible.

¹<https://kubespray.io/>

²<https://github.com/tomas321/k8s-environment-scenarios>

³<https://github.com/tomas321/sneakpeek>

⁴<https://github.com/tomas321/fswatch>

Literature

- [0] Aditya Anand. “Malware Analysis 101 - Sandboxing. Cons of using a VM”. Sept. 29, 2019. URL: <https://medium.com/bugbountywriteup/malware-analysis-101-sandboxing-746a06432334> (visited on 03/27/2020).
- [2] Ujaliben Kalpesh Bavishi et al. “Malware Analysis”. In: *ijarcsse* (2017). URL: <https://ijarcsse.com/index.php/ijarcsse/article/view/507> (visited on 03/26/2020).
- [0] “Defense in Depth: Detonation Technologies”. Mar. 12, 2018. URL: <https://inquest.net/blog/2018/03/12/defense-in-depth-detonation-technologies> (visited on 10/11/2020).
- [3] Manuel Egele et al. “A Survey on Automated Dynamic Malware-Analysis Techniques and Tools”. In: *ACM Computing Surveys. Article 6* 44.2 (Feb. 8, 2012). Article 6, pp. 5–21. DOI: <https://dl.acm.org/doi/10.1145/2089125.2089126>. URL: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.640.6356&rep=rep1&type=pdf> (visited on 03/26/2020).
- [0] “ENDPOINT DECEPTION”. URL: <https://cybertrap.com/en/solutions/#endpointdeception> (visited on 10/05/2020).

- [4] Tomasz Grudziecki et al. “Proactive Detection of Security Incidents - Honeypots”. enisa study. Nov. 2012. URL: <https://www.enisa.europa.eu/publications/proactive-detection-of-security-incidents-II-honeypots>.
- [0] Juan Guarnizo et al. “SIPHON: Towards Scalable High-Interaction Physical Honeypots”. In: (Jan. 12, 2017). URL: <https://arxiv.org/pdf/1701.02446.pdf> (visited on 09/25/2020).
- [0] “How it works. Many contributors”. Mar. 23, 2020. URL: <https://support.virustotal.com/hc/en-us/articles/115002126889-How-it-works> (visited on 10/11/2020).
- [5] Vladislav Hřčka. “Stantinko’s new cryptominer features unique obfuscation techniques”. Mar. 19, 2020. URL: <https://www.welivesecurity.com/2020/03/19/stantinko-new-cryptominer-unique-obfuscation-techniques/> (visited on 03/23/2020).
- [0] Brent Huston. *State Of Security Episode 15*. Apr. 5, 2019. URL: https://www.podbean.com/media/share/pb-cgwhv-ad161e?utm_campaign=w_share_ep&utm_medium=dlink&utm_source=w_share (visited on 10/05/2020).
- [0] Brent Huston. “What is this HoneyPoint Thing Anyway?” Dec. 6, 2012. URL: <https://stateofsecurity.com/what-is-this-honeypoint-thing-anyway/> (visited on 10/05/2020).
- [0] Yifat Mor. “Using Dynamic Honeypot Cyber Security: What Do I Need to Know?” Oct. 10, 2018. URL: <https://www.guardicore.com/2018/10/dynamic-honeypot-cyber-security/> (visited on 10/11/2020).
- [0] Alexander Nestorov. “What is Monks”. Feb. 24, 2015. URL: <https://github.com/alexandernst/monks> (visited on 12/04/2020).
- [0] Ivan Studnia et al. “A distributed platform of high interaction honeypots and experimental results (extended version)”. In: (June 10, 2012). DOI: <https://hal.archives-ouvertes.fr/hal-00706333>. URL: <https://www.>

`researchgate.net/publication/262277761_A_distributed_platform_of_high_interaction_honeypots_and_experimental_results` (visited on 09/26/2020).

- [6] Lenny Zeltser. “5 Steps to Building a Malware Analysis Toolkit Using Free Tools”. SANS Institute instructor. Mar. 2015. URL: <https://zeltser.com/build-malware-analysis-toolkit/#allocate-virtual-systems>.