

Faculty of Mathematics and Computer Science

Heidelberg University

Master thesis

in Computer Science

submitted by

Stefan Machmeier, 3604391

born in Heidelberg

2022

Honeypot Implementation in a Cloud Environment

This Master thesis has been carried out by Stefan Machmeier

at the

Engineering Mathematics and Computing Lab

under the supervision of

Prof. Dr. Vincent Heuveline

Abstract

In this age of digitalization, Internet services face more attacks than ever. An attacker's objective is to exploit systems and use them for malicious purposes. Such efforts are rising as vulnerable systems can be discovered and compromised through Internet-wide scanning. One known methodology besides traditional security leverages is to learn from those who attack it. A honeypot helps to collect information about an attacker by pretending to be a vulnerable target. Thus, how honeypots can contribute to a more secure infrastructure makes an interesting topic of research. This thesis will present a honeypot solution to investigate malicious activities in heiCLOUD and show that attacks have increased significantly. To detect attackers in restricted network zones at Heidelberg University, a new concept to discover leaks in the firewall will be created. Furthermore, to consider an attacker's point of view, a method for detecting honeypots at the transport level will be introduced. Lastly, a customized OpenSSH server that works as an intermediary instance will be presented to mitigate these efforts.

Zusammenfassung

Heutzutage sind Dienste, die über das Internet zugänglich sind, mehr Angriffen ausgesetzt als je zuvor. Das Ziel von Angreifern ist es, Systeme auszunutzen und sie für ihre eigenen böartigen Zwecke zu verwenden. Derartige Bemühungen nehmen zu, da verwundbare Systeme durch internetweites Scannen entdeckt und kompromittiert werden können. Neben den traditionellen Sicherheitsmaßnahmen besteht auch die Möglichkeit, von den Angreifern zu lernen. Ein Honeypot hilft dabei, Informationen über Angreifer zu sammeln, indem er vorgibt, ein verwundbares Ziel zu sein. Daher ist es eine interessante Forschungsfrage, wie Honeypots zu einer sichereren Infrastruktur beitragen können. In dieser Arbeit wird eine Honeypot-Lösung zur Untersuchung von Böswillige-Aktivitäten in der heiCLOUD vorgestellt und gezeigt, dass die Angriffe aus dem Internet erheblich zugenommen haben. Des Weiteren wird versucht, Angreifer in eingeschränkten Netzwerkzonen der Universität Heidelberg zu entdecken. Die Ergebnisse zeigen, dass die Firewall Lücken aufweist und Angreifer in der Lage waren gewisse Bereiche zu scannen. Zusätzlich wird die Sichtweise eines Angreifers eingenommen und eine Methode zur Erkennung von Honeypots auf Transportebene vorgestellt. Abschließend wird ein angepasster OpenSSH-Server vorgestellt, der als Zwischeninstanz fungiert, um diese Bemühungen zu verhindern.

Erklärung

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Sowohl inhaltlich als auch wörtlich entnommene Inhalte wurden als solche kenntlich gemacht.

Die Arbeit ist in gleicher oder vergleichbarer Form noch bei keiner anderen Prüfungsbehörde eingereicht worden.

Heidelberg, den 01.02.2022

Stefan Machmeier

Acknowledgements

The research included in this thesis could not have been performed if not for many individuals' assistance, patience, and support.

First and foremost, I am deeply grateful to my supervisor, Prof. Dr. Vincent Heuveline for his valuable and constructive input. Without his guidance and mentorship, I would not have been able to finish this thesis. Moreover, I grew as a researcher, and I am immensely grateful for the opportunity to continue my research as a future Ph.D. candidate under his supervision.

I want to extend my gratitude towards Stefan Steiger and Olaf Pichler from the Computing Centre at Heidelberg University. Thank you for offering insightful comments and brilliant suggestions when the task got challenging. They always had time and provided me with ample support no matter what happened.

I am indebted to Joachim Peek for generously agreeing to examine my results and providing valuable inputs. His timely advice and scientific knowledge helped me understand essential parts of the topic assisted me to a great extent in accomplishing this task.

Lastly, I could not have completed this thesis without the support of my girlfriend, Carmen. Thank you for being so patient in providing emotional support and stimulating discussions during my research.

Contents

Acronyms	VII
List of Figures	X
List of Tables	XI
Listings	XII
1 Introduction	1
2 Background	3
2.1 Virtualization	3
2.2 Cloud Computing	4
2.2.1 Definition of Cloud Computing	4
2.2.2 Service models	5
2.2.3 Deployment models	6
2.3 Honeypots	7
2.3.1 Definition of Honeypots	7
2.3.2 Level of Interaction	10
2.3.3 Security concepts	10
2.3.4 Value of Honeypots	11
2.3.5 Honeynets	13
2.3.6 Legal Issues	14
3 Analyze Honeypot Attacks in the Cloud	15
3.1 Introduction	15
3.2 Methodology	17
3.2.1 heiCLOUD	17
3.2.2 T-Pot	19
3.3 Results	25
3.4 Discussion	32
4 Catching Attackers in Restricted Network Zones	36
4.1 University Network	36
4.2 Honeypot-like Connection Detection Tool	38
4.3 Results	42

4.4	Discussion	47
5	Mitigate Fingerprint Activities of Honeypots	50
5.1	OpenSSH	50
5.2	Preliminary Work	52
5.3	Experiment 1: Reproduce Vetterl et al.'s findings	55
5.4	Attempt to Disguise Cowrie	58
5.5	Experiment 2: Avoid fingerprinting of Cowrie	60
5.6	Discussion	61
6	Conclusion	68
	Bibliography	70

Acronyms

ACL	Access Control List
ADB	Android Debug Bridge
ADC	Application Delivery Controller
API	Application Programming Interface
AS	Autonomous System
ASA	Adaptive Security Appliance
ASN	Autonomous System Number
AWS	Amazon Web Services
BeIWÜ	Baden-Württembergs extended LAN
BPP	Binary Packet Protocol
CERT	Computer Emergency Response Team
CGI	Common Gateway Interface
CHARGEN	Character Generator Protocol
CPU	central processing unit
CVE	Common Vulnerabilities and Exposures
DaaS	Data-as-a-Service
DDoS	Distributed Denial of Service
DICOM	Digital Imaging and Communications in Medicine
DMZ	demilitarized zone
DNS	Domain Name System
DTK	Deception Toolkit
EU	European Union
FHIR	Fast Healthcare Interoperability Resources

FTP File Transport Protocol
GB gigabyte
GCP Google Cloud Platform
GPU graphics processing unit
HaaS Hardware-as-a-Service
HTTP Hypertext Transfer Protocol
IaaS Infrastructure-as-a-Service
ICMP Internet Control Message Protocol
ICS Industrial Control System
IDENT Identification Protocol
IDS intrusion detection system
IMAP Internet Message Access Protocol
IOCTA Internet Organised Crime Threat Assessment
IP Internet Protocol
IPD intrusion prevention system
IPP Internet Printing Protocol
MAC Message Authentication Code
MITM man-in-the-middle
NAT network address translation
NIST National Institute of Standards and Technology
NLA Network Level Authentication
NSM network security monitoring
NTP Network Time Protocol
OS operating system
PaaS Platform-as-a-Service
POP Post Office Protocol
RAM random-access memory
RDBMS relational database management system
RDP Remote Desktop Protocol

SaaS Software-as-a-Service
SCADA Supervisory Control and Data Acquisition
SIP Session Initiation Protocol
SMB Server Message Block
SMTP Simple Mail Transfer Protocol
SNMP Simple Network Management Protocol
SOHO small office/home office
SSDP Simple Service Discovery Protocol
SSH Secure Shell
TCP Transmission Control Protocol
TLS Transport Layer Security
UDP User Datagram Protocol
USB Universal Serial Bus
vCPU virtual central processing unit
VLAN Virtual Local Area Network
VM virtual machine
VMM virtual machine monitors
VNC Virtual Network Computing
VoIP Voice over Internet Protocol
VPN virtual private network
XML Extensible Markup Language

List of Figures

2.1	Abstract visualization of service models	5
2.2	Example of honeypots in a simplified network	9
2.3	Example of honeynets in a simplified network	13
3.1	Concept to collect honeypot attacks	18
3.2	T-Pot architecture	20
3.3	Distribution of honeypot attacks	25
3.4	Attack distribution of T-Pot	26
3.5	Attack histogram of T-Pot	27
3.6	Suricata results of T-Pot	28
3.7	RDPY results of T-Pot	29
3.8	Honeytrap results of T-Pot	30
3.9	Cowrie results of T-Pot	30
3.10	Cowrie top 10 credentials on T-Pot	31
4.1	Draft of the University network	37
4.2	Concept to detect connection attempts	40
4.3	Visualization of the MADCAT packet flow	41
4.4	Protocol distribution of MADCAT	42
4.5	Attack distribution of MADCAT	43
4.6	Suricata results of T-Pot	44
4.7	Attack port histogram of T-Pot	45
4.8	Attack distribution of T-Pot	47
4.9	Attack port histogram of T-Pot	48
5.1	OpenSSH architecture	51
5.2	Process to obtain the cosine similarity coefficient	53
5.3	Architecture of OpenSSH and Cowrie	54
5.4	Architecture of OpenSSH and Cowrie	59
5.5	OpenSSH sample session flow diagram	62

List of Tables

2.1	Distinction between security concepts	11
3.1	Overview of attacks on cloud providers	16
3.2	Overview of honeypots of T-Pot	24
3.3	Overview of attacks on heiCLOUD, AWS, GCP, and Azure	33
4.1	Overview of firewall stages	38
5.1	Overview of the cosine similarity of OpenSSH, Cowrie, and Twisted	55

Listings

3.1	Cowrie attack to gather various information about the system	34
3.2	Cowrie attack to exploit the host machine as a crypto miner	35
4.1	MADCAT connection attempt to exploit SIP connection	45
4.2	MADCAT connection attempt to exploit SMB connection	46
5.1	Example OpenSSH connection with probed SSH packet	53
5.2	OpenSSH connection attempt with probed message	56
5.3	Cowrie connection attempt with probed message	57
5.4	TwistedConch packet length validation	58
5.5	Cowrie version string validation	63
5.6	Tweaked OpenSSH authentication	64
5.7	Tweaked OpenSSH channel	65
5.8	Tweaked OpenSSH server loop	66
5.9	Cowrie log information	67

Chapter 1

Introduction

Recently, Europol¹ raised awareness of new cyber threats related to the ongoing pandemic. As stated in their yearly Internet Organised Crime Threat Assessment (IOCTA) report, scanning of corporate infrastructures has been skyrocketing within the last 12 months by ransomware groups, respectively increasing malware usage. Attackers use scans to find potential vulnerabilities in remote desktop sharing software, or virtual private networks (VPNs) in order to deploy malware and blackmail companies [27]. The rapid increase dates back to the pandemic and the shift to home office, forcing companies to adapt their infrastructures quickly. Such changes come with the downside of adding new threats to an organization. The latest incident at the SRH University Heidelberg points out the obstacles institutions face when ransomware groups have access and exploit various parts of the infrastructure with malware. An unknown group infected systems with malware and distributed internal data in the darknet. Such incidents emphasize the rise of malicious activities.

Especially in cloud computing, controlling access to services is becoming a stricter challenge due to access to large data sets and computing resources. Besides traditional security measures such as firewalls or intrusion detection systems, one known methodology to strengthen infrastructures is learning from those who attack them. Honeypots are a proper instrument to gather information about attackers. It is “a security resource whose value lies in being probed, attacked, or compromised” [65]. Collecting attacks can reveal shell-code exploitation or bot activity. In retrospect, this would help to harden infrastructures before proper damage occurs. For a cloud provider, it is crucial to know whether and how attacks on its service can be prevented. Considering the Global Security Report by Trustwave, the number of attacks doubled in 2019 and increased by 20% in 2020 [1], respectively putting cloud providers to the third most targeted environments for cyberattacks, behind corporate and internal networks.

The Heidelberg University offers its own cloud service, called heiCLOUD. It enables users to maintain and control computational resources easily. Thus, it is interesting

¹An agency that fights against terrorism, cybercrime, and other threats [28]

to elaborate on the value of honeypots for this cloud solution. This thesis tries to answer the general research question of whether honeypots can contribute to a more secure infrastructure in a cloud environment. This includes deploying a honeypot solution in heiCLOUD and presenting the results. Prior to that, an insight into a recent study investigating honeypots for the cloud providers AWS, GCP, and Microsoft Azure is given. These findings help to validate the results in heiCLOUD. In addition, the university network will be investigated to find potential leaks in the stateless firewall. Therefore, a concept is created using the BSI's honeypot-like detection tool MADACT and deployed on desktop computers inside the university building. Furthermore, to consider an attacker's point of view, this thesis introduces a recent work to detect honeypots on the transport level. Lastly, a solution to mitigate these efforts will be presented.

This thesis includes six chapters. After the introduction, chapter 2 outlines the background knowledge that is needed to comprehend the upcoming experiments. It gives the reader a profound understanding of cloud computing, honeypots, and virtualization. Chapter 3, *Analyze Honeypot Attacks in the Cloud*, presents the status quo of malicious activities in heiCLOUD. In the beginning, it shows the results that Kelly et al. [40] claim for AWS, GCP, and Microsoft Azure. Next, it gives an insight into the T-Pot solution used to collect the data and shows the results after collecting them for three weeks. Furthermore, chapter 4, *Catching Attackers in Restricted Network Zones*, investigates the university network in which the new concept is deployed for three weeks. It shows that the concept was able to adapt the firewall, thus, improving the network security at the university. Chapter 5, *Mitigate Fingerprint Activities of Honeypots*, presents two experiments. First, it describes the preliminary work to detect honeypots and finishes with an experiment to prove this assumption. Next, it drafts the counterpart of mitigating this activity, also closing up with an experiment. Lastly, chapter 6 completes this thesis with a conclusion that summarizes the results and describes future work in this regard.

Chapter 2

Background

A honeypot is a security resource whose value lies in being probed, attacked, or compromised.

Lance Spitzner

Using honeypots in a cloud environment merges two varying principles. This chapter introduces the fundamental knowledge needed to comprehend the upcoming experiments. If the reader has a profound understanding of cloud computing, honeypots, and virtualization, he can skip this chapter.

2.1 Virtualization

Virtualization, often referred to as virtual machines (VMs), is defined by Kreuter [43] as “an abstraction layer or environment between hardware component and the end-user”. A VM runs on top of the operating system’s (OS’s) core components. Through an abstraction layer, the virtual machine is connected with the real machine by hypervisors or virtual machine monitors (VMM). Hypervisors can use real machine hardware components but also support virtual machine’s operating systems and configurations. Both are similar to emulators, which are defined by Lichstein [45] as a “process whereby one computer is set up to permit the execution of programs written for another computer”. This allows managing multiple VMs with real machine resources. There are three different types of virtualization, (i) software virtual machines, (ii) hardware virtual machines, and (iii) virtual OS/containers. Software virtual machines manage interactions between the host and guest operating systems [21]. Hardware virtual machines offer direct and fast access to the underlying resources [21]. It uses hypervisors, modified code, or Application Programming Interfaces (APIs). Lastly, virtual OS/container partitions the host operating system into containers or zones [21].

2.2 Cloud Computing

Cloud Computing has become a buzzword these days. It has been used by various large companies such as Google and Amazon. However, the term “cloud computing” dates back to late 1996, when a small group of technology executives of Compaq Computer framed new business ideas around the Internet [56]. Starting from 2007, cloud computing evolved into a serious competitor and outnumbered the keywords’ “virtualization”, and “grid computing” as reported by Google trends [73]. Shortly, various cloud providers become publicly available, each with its strengths and weaknesses. For example IBM’s Cloud¹, Amazon Web Services², and Google Cloud³. So, why are clouds so attractive in practice?

- It offers major advantages in terms of cost and reliability. When demand is needed, consumers do not have to invest in hardware when launching new services. Pay-as-you-go allows flexibility.
- Consumers can easily scale with demand. When more computational resources are required due to more requests, scaling up instances in conjunction with a suited price model is straightforward.
- Geographically distributed capabilities supply the need for worldwide scattered services.

2.2.1 Definition of Cloud Computing

According to the definition by Brian Hayes, cloud computing is “a shift in the geography of computation” [33]. Thus, the computational workload is moved away from local instances towards services and data centers that provide the user’s needs [3].

The National Institute of Standards and Technology (NIST) defines cloud computing as “a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction” [47]. NIST not only reflects the geographical shift of resources such as data centers but also mentions on-demand usage that contributes to flexible resource management. Moreover, NIST composes the term into five essential characteristics, three service models (see subsection 2.2.2), and four deployment models (see subsection 2.2.3) [47].

On-demand-self-service refers to the unilateral provision computing capabilities. Consumers can acquire server time and network storage on demand without human interaction.

¹<https://www.ibm.com/cloud>

²<https://aws.amazon.com/>

³<https://cloud.google.com/>

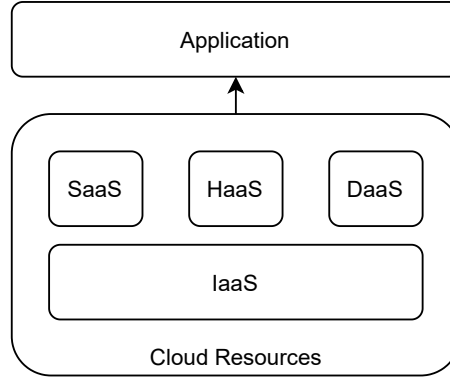


Figure 2.1: Abstract visualization of service models. The container “cloud resources” represents the depth of functionalities. Therefore, Infrastructure-as-a-Service (IaaS) offers the most functionalities, whereas the others have a user-friendly abstraction.

Broad network access characterizes the access of capabilities of the network through standard protocols such as Hypertext Transfer Protocol (HTTP). Heterogeneous thin and thick client platforms should be supported.

Resource pooling allows the provider’s computing resources to be pooled across several consumers. Different physical and virtual resources are assigned on-demand with a multi-tenant model. Other aspects such as location are independent and cannot be controlled on a low-level by consumers. Moreover, high-level access to specify continent, state, or datacenter can be available.

Rapid elasticity offers consumers to extend and release capabilities quickly. Further automation to quickly increase resources when demand surges can be supported at any time, regardless of limit or quantity.

Measured service handles resources in an automated and optimized manner. It uses additional metering capabilities to trace storage, processing, bandwidth, and active user accounts. This helps to monitor and control resource usage. Thus, contributing to transparency between provider and consumer.

2.2.2 Service models

Service models are categorized by NIST into three basic models based on usage and abstraction level. Figure 2.1 shows the connection between each model whereas cloud resource are defined in subsection 2.2.3. Infrastructure-as-a-Service (IaaS) builds with a vast range of functionalities the foundation of service models. Each model on top represents a user-friendly abstraction with derated capabilities.

Software-as-a-Service (SaaS) is a high-level abstraction to consumers. Controlling the underlying infrastructure is not supported. Providers often use a multi-tenancy

system architecture to organize each consumer’s application in a different environment. It helps to employ scaling with respect to speed, security availability, disaster recovery, and maintenance [47]. The main objective of SaaS is to host a consumer’s software or application that can be accessed over the Internet using either a thin or rich client [23]. Users can apply custom configuration settings [47].

Platform-as-a-Service (PaaS) pivots on the full “Software Lifecycle” of an application whereas SaaS distinct on hosting complete applications. PaaS offers ongoing development and includes programming environment, tools, configuration management, and other services. In addition, the underlying infrastructure is not managed by the consumer [47].

Infrastructure-as-a-Service (IaaS) offers a low-level abstraction to consumers with the ability to run arbitrary software regardless of the operating system or application. In contrast to SaaS, IT infrastructure capabilities (such as storage and networks) can be used. It strongly depends on virtualization due to the integration or decomposition of physical resources [47].

Data-as-a-Service (DaaS) serves as a virtualized data storage service on demand. Motivations behind such services could be upfront costs of on-premise enterprise database systems [23]. Mostly they require “dedicated server, software license, post-delivery services, and in-house IT maintenance” [23] whereas DaaS costs solely what consumers need. When dealing with a tremendous amount of data, file systems and relational database management systems (RDBMSs) often lack performance. DaaS outruns such weak links by employing a table-style abstraction that can be scaled [23].

Hardware-as-a-Service (HaaS) offers IT hardware or datacenters to buy as a pay-as-you-go subscription service. The term dates back to 2006 when hardware virtualization became more powerful. It is flexible, scalable, and manageable [73].

2.2.3 Deployment models

Deployment models are categorized by NIST into four basic models. Each differs in data privacy, location, and manageability [47].

With private clouds, users have the highest control regarding data privacy and utilization. Such clouds are mostly deployed within a single organization, managed by in-house teams or third-party suppliers. In addition, it can be on- or off-premise. Within private clouds, consumers have full control of their data. Especially for European data privacy laws, it is not negligible when data is stored abroad, and thus, under the law of foreign countries. However, its popularity has not been diminished due to the immense cost of switching to public clouds [23, 47].

Community clouds can be seen as a conglomerate of multiple organizations that merge their infrastructure with respect to a commonly defined policy, terms, and conditions beforehand [47].

Public clouds represent the most used deployment models. Contrary to private ones, public clouds are fully owned by service providers such as businesses, academics, or government organizations. Consumers do not know where their data is distributed. In addition, contracts underlie custom policies [47].

A hybrid cloud mixes two or more cloud infrastructures, such as private and public clouds. However, each entity keeps its core element. Hybrid clouds define “standardized or proprietary technology to enable data and application portability”[47].

2.3 Honeypots

The term “honeypot” has been established for more than two decades. 1997 was the first time that a free honeypot solution became public. Deception Toolkit (DTK), developed by Fred Cohen, released the first honeypot solution. However, the earliest drafts of honeypots are from 1990/91 and built the foundation for Fred Cohen’s DTK. Clifford Stoll’s book “The Cuckoo’s Egg”[66], and Bill Cheswick’s whitepaper “An Evening With Berferd”[7] describe concepts that are considered nowadays as honeypots [65]. A honeypot itself is a security instrument that collects information on buzzing attacks. It disguises itself as a system or application with weak links, so it gets exploited and gathers knowledge about the adversary. In 2002, a Solaris honeypot helped to detect an unknown `dtspcd` exploit. Interestingly, a year before in 2001, the Coordination Center of CERT⁴ shared their concerns regarding the `dtspcd`. Communities were aware that the service could be exploited to get access and remotely compromise any Unix system. However, such an exploit was not known during this time, and experts did not expect any in the near future. Luckily, early instances based on honeypot technologies could detect new exploits and avoid further incidents. Such events emphasize the importance of honeypots.

2.3.1 Definition of Honeypots

Many definitions for honeypots circulate through the web that causing confusion and misunderstandings. In general, the objective of a honeypot is to gather information about attacks or attack patterns [51]. Thus, contributing as an additional source of security measure. See subsection 2.3.3 for a detailed view regarding honeypots in the security concept. As Spitzner [65] has listed, the most misleading definitions: a honeypot is a tool for deception, it is a weapon to lure adversaries or a part of

⁴Computer Emergency Response Team is an expert group that handles computer security incidents[31]

an intrusion detection system. In order to get a basic understanding, this section wants to exhibit some key definitions. Spitzner [65] defines honeypots as a “security resource whose value lies in being probed, attacked, or compromised”. Independent of its source (e.g., server, application, or router), he expects the instance to be probed, attacked, and eventually exploited. If a honeypot does not match this behavior, it will not provide any value. It is essential to mention that honeypots do not have any production value; thus, any communication that is acquired is suspicious by nature [65]. In addition, Spitzner [65] points out that honeypots are not bound to solve a single problem; hence, they function as a generic perimeter and fit into different situations. Such functions are attack detection, capturing automated attacks, or alert/warning generators. Figure 2.2 shows an example of how honeypots could be used in an IT infrastructure.

In general, he differentiates two types of honeypots (i) production honeypots (ii) research honeypots. This categorization has its origin from Mark Rosch, a developer of Snort, during his work at GTE Internetworking [72].

Production honeypots are the most common type of honeypots that people would think of. The objective is to protect production environments and mitigate the risk of attacks. Usually, production honeypots are easy to deploy within an organization. Mostly, low-interaction honeypots are chosen due to a significant risk reduction, so adversaries cannot exploit honeypots to attack other systems [65]. The downside of a low-interaction honeypot is a lack of information, which means only standard information like the origin of attacks or what exploits have been used can be collected [50]. On the contrary, insides about the communication of attackers or deployment of such attacks are unlikely to obtain, whereas research honeypots fulfill this objective [65].

Research honeypots are used to learn more in detail about attacks. The objective is to collect information about clandestine organizations, new tools for attacks, or the origin of attacks [65, 50]. Research honeypots are unlikely suitable for production environments due to a higher risk increase. Facing an increase in deployment complexity and maintenance does not attract production usage either [65].

It is worth mentioning that there is no exact line between research or production honeypots. A possible use case is a honeypot that functions as a production or a research honeypot. Due to the dynamic range in which they are applicable, it is difficult to distinguish them.

In addition, Provos [55] adds a differentiation for the virtual honeypot framework and splits it into the following types:

- Physical honeypots are “real machines on the network with its own Internet Protocol (IP) address” [55]
- Virtual honeypots are “simulated by another machine that responds to network traffic sent to the virtual honeypot” [55]

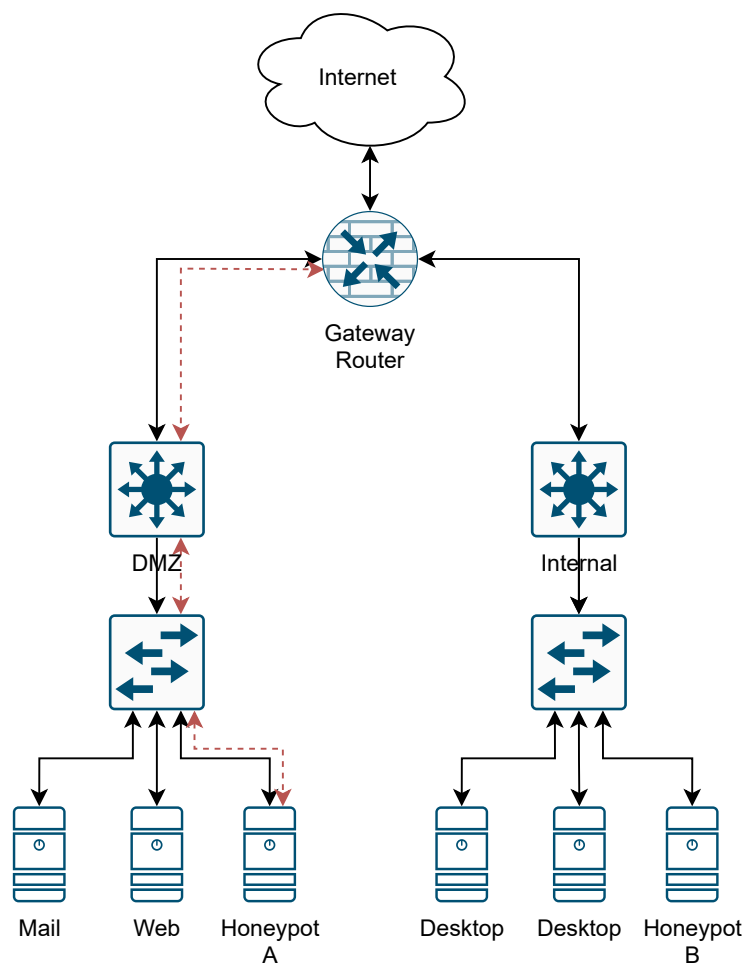


Figure 2.2: Example of honeypots in a simplified network (derived from [65]). Each of the demilitarized zones (DMZs) and internal networks are separated by a router and a Layer-3 switch. In each network a honeypot is available (honeypot A, B). The red path symbolizes the path of an attacker coming from the gateway router.

2.3.2 Level of Interaction

When building and deploying a honeypot, the depth of information has to be defined beforehand. Should it gather unauthorized activities, such as an NMAP scan? Do you want to learn about buzzing tools and tactics? Each depth brings a different level of interaction because some information depends on more actions of adversaries. Therefore, honeypots differ in their level of interaction.

Low-interaction honeypots provide the lowest level of interaction between an attacker and a system. Only a small set of services like Secure Shell (SSH), Telnet, or File Transport Protocol (FTP) are supported, contributing to the deployment time. In terms of risk, a low-interaction honeypot does not give access to the underlying OS which makes it safe to use in a production environment [65]. For example, using an SSH honeypot with emulated services allows attackers to log in and execute commands by brute force or guesswork. The adversary will never gain more access because it is not a real OS. However, safety comes with the downside of less information. The collection is limited for the statistical purpose such as (i) time and data of attack (ii) source IP address and source port of the attack (iii) destination IP address and destination port of the attack [65, 50]. The transactional information can not be collected [65].

A medium-interaction honeypot offers more sophisticated services with a higher level of interaction. It is capable of responding to specific activities. For example, a Microsoft IIS Web server honeypot could respond in a way that a worm is expecting. The worm would get emulated answers and could be able to interact with it in more detail. In this way, more severe information about the attack can be gathered, including privilege assessment, toolkit capture, and command execution Spitzner [65]. In comparison, medium-interaction honeypots allocate more time to install and configure [65, 50]. Also, more security checks have to be performed due to a higher interaction level than low-interaction honeypots [65].

High-interaction honeypots represent a real OS to provide a full set of interactions to attackers [65]. They are so powerful because other production servers do not differ much from high-interaction honeypots. They represent real systems in a controlled environment [65, 50]. The amount of information is tremendous. It helps to learn about (i) new tools (ii) finding new bugs in the OS (iii) the black hat community [65]. However, the risk of such a honeypot is extremely high. It needs severe deployment and maintenance processes; thus, it is time-consuming.

2.3.3 Security concepts

Security concepts are classified by Schneier [63] in prevention, detection, and reaction. Prevention includes any process that (i) discourages intruders and (ii) hardens systems to avoid any breaches. Detection scrutinizes the identification of attacks

Table 2.1: Distinction between security concepts based on areas of operations (derived from [51]).

Objective	Prevention	Detection	Reaction
Honeypot	+	++	+++
Firewall	+++	++	+
Intrusion Detection Sys.	+	+++	+
Intrusion Prevention Sys.	++	+++	++
Anti-Virus	++	++	++
Log-Monitoring	+	++	+
Cybersecurity Standard	+++	+	+

that threatens the systems' (i) confidentiality (ii) integrity and (iii) availability. Reaction treats the active part of the security concept. When attacks are detected, it conducts reactive measures to remove the threat. Each part is designed to be sophisticated so that all of them contribute to a secure environment [51].

Honeypots contribute to the security concept like firewalls, or intrusion detection systems (IDSs). However, honeypots add only a small value towards prevention because security breaches cannot be identified. Moreover, attackers would avoid wasting time on honeypots and go straight for production systems instead.

Detection is one of the strengths of honeypots. Attacks often vanish in the sheer quantity of production activities. If any connection is established to a honeypot, it is suspicious by nature. In conjunction with an alerting tool, attacks can be detected.

Honeypots strongly supply reaction tools due to their clear data. It is difficult to find attacks for further data analysis in production environments. Often data submerge with other activities, which complicates the process of reaction [51]. Nawrocki et al. [51] distinguish honeypots from other objectives such as firewall or log-monitoring.

2.3.4 Value of Honeypots

To assess the value of honeypots, this section looks at their advantages and disadvantages [50, 39, 65].

Advantages

- *Data Value*: Collected data is often immaculate and does not contain noise from other activities. Thus, reducing the total data size and speeding up the analysis.

- *Resources*: Firewalls and IDS are often overwhelmed by the gigabits of traffic, thus, dropping network packets for analysis. This results in far less effective detection of malicious network activities. However, honeypots are independent of resources because they only capture their activities. Due to resource limitations, expensive hardware is not needed.
- *Simplicity*: A honeypot does not require complex algorithms or databases. If a honeypot is too complex, it will lead to misconfigurations, breakdowns, and failures. The challenging research honeypots might come with an inevitable increase in complexity in maintenance.
- *Return on Investment*: Capturing attacks immediately informs users that suspicious activities occur on the infrastructure. This helps to demonstrate their value and contributes to new investments in other security measurements.

In addition, Nawrocki et al. [51] listed four more advantages of honeypots:

- *Independent of Workload*: Honeypots only process traffic directed to them.
- *Zero-Day-Exploit Detection*: It helps to detect unknown strategies and zero-day-exploits.
- *Flexibility*: Well-adjusted honeypots for various specific tasks are available.
- *Reduced False Positives and Negatives*: Any traffic or connection to a honeypot is suspicious. Client-honeypots verify such attacks based on system state changes. This results in either false positive or false negatives.

Disadvantages

- *Narrow Field of View*: Only direct attacks on honeypots can be investigated, whereas attacks on the production system are not detected.
- *Fingerprinting*: A honeypot often has a certain fingerprint that attackers can identify. Especially commercial ones can be detected by their responses or behaviors.
- *Risk to the Environment*: Using honeypots in an environment always increases risk. However, it depends on the level of interaction.

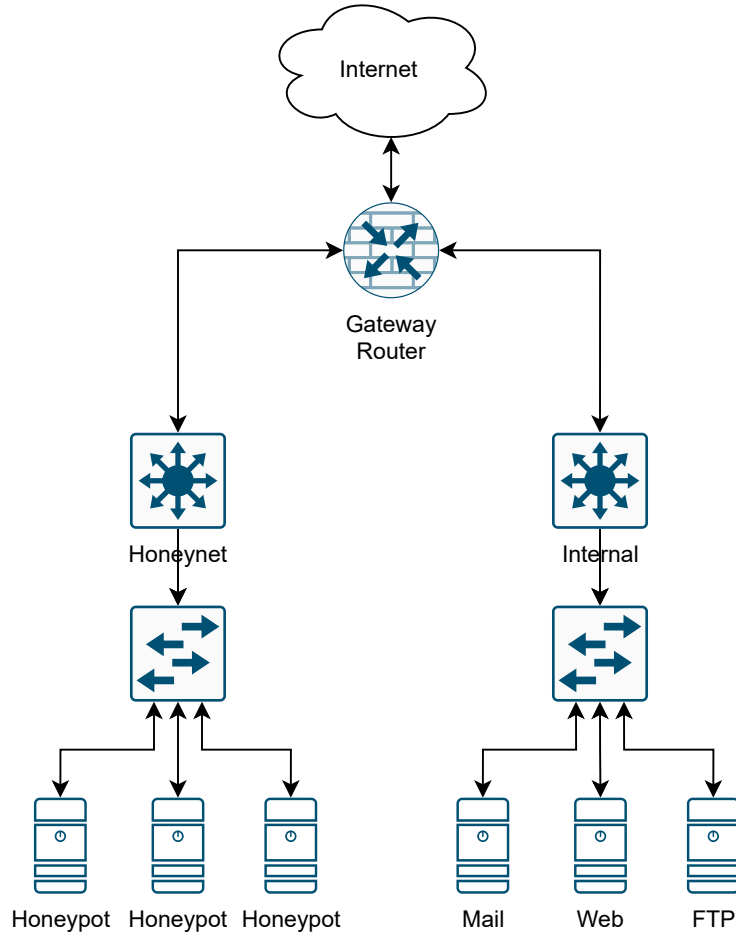


Figure 2.3: Example of honeynets in a simplified network (derived from [65]). This network presents the honeynet consisting of several other honeypots on the left. On the right, the network presents a common subnet consisting of mail, web, and FTP server.

2.3.5 Honeynets

Instead of having single honeypots that can be attacked, a honeynet offers a complete network of standard production systems such as you would find in an organization [64]. Those systems are high-interaction honeypots, thus, allowing them to fully interact with the OS and applications. The key idea is that an adversary can probe, attack, and exploit these systems so that the maintainer can derive interaction within this network [65, 64]. It should be mentioned that a honeynet has to be protected by firewalls. For example, Figure 2.3 represents such a honeynet within an organization.

Compared to a traditional honeypot, the most significant value of honeynets is the usage of proper production systems. Black hats often do not know that they attack

a honeynet, thus, adding value to prevention. However, the downsides are the high complexity and maintenance needed to keep a honeynet running [65].

2.3.6 Legal Issues

Considering questions related to legal issues of honeypots can easily exceed this thesis. In this regard, this section restricts the study to the country the author resides in. Thus, only the European Union (EU) regulations, EU directives, and international agreements are considered. Honeypots collect (i) content data that is used for communication, and (ii) transactional data that is used to establish the connection. Sokol et al. [64] studied the legal conditions for data collection and data retention. They have concluded that administrators of honeypots have a legal ground of legitimate interest to store and process personal data, such as IP addresses. Moreover, for production honeypots, the legitimate interest is to secure services. Regarding the length of data retention, the principle of data minimization has to be considered, which means there is no clear answer. Any published data of research honeypots needs to be anonymized.

Chapter 3

Analyze Honeypot Attacks in the Cloud

Attacks from the Internet often originate from bots. A bot, short for “robot”, is an automated process that interacts with different network services. Despite good intentions, bots can be used for malicious purposes. Mostly, bots try to self-propagate malware across the Internet and try to capture hosts that merge into a botnet [29]. Recently, Universities in Germany received more cyberattacks than ever, respectively increasing their costs for damage repairs. Honeypots are a good solution to catch attackers and learn from their exploits. However, it is not clear whether honeypots are an appropriate countermeasure to prevent such damage in the age of bots. Following the rise of cyberattacks, this chapter introduces a method to collect and analyze cyberattacks in a cloud environment. It further proposes an answer if honeypots are helpful to detect bot activities.

3.1 Introduction

As previously mentioned in section 2.2, using cloud resources is becoming the go-to option for new services and applications. Kelly et al. [40] thoroughly investigated honeypots on Azure, Amazon Web Services (AWS), and Google Cloud Platform (GCP). Consequently, this chapter presents their results briefly to compare them with the ones heiCLOUD achieves. The results are collected by T-Pot version 20.06.0 for three weeks. In addition, Kelly et al. [40] considered different server geographical locations. They have collected data from East US, West Europe, and Southeast Asia. Table 3.1 shows the results presented by Kelly et al. [40]. Dionaea (a honeypot to capture malicious payload), Cowrie (SSH and Telnet honeypot), and Conpot (industrial honeypot for ICS and SCADA) are the most attacked honeypots in comparison to the others. Regarding AWS, Dionaea accounts for 91% of the total attacks, Glutton and Cowrie are minor with 5%, and 2%. Interestingly, Cowrie reported several attacks related to the COVID-19 pandemic to enable social

engineering methods. In contrast to AWS, Cowrie logged the majority of attacks with 51% on GCP. Besides several automated attacks trying to log in with default credentials, adversaries tried to gather information about the GPU architecture, scheduled tasks, and privilege escalation. Microsoft Azure reflects nearly the same results as the other two cloud providers beforehand.

Table 3.1: Overview of attacks on cloud providers. For a better overview, only the three most attacked honeypots are listed. The remaining honeypots are listed in the column named "others".

PROVIDER	HONEYPOT				IN TOTAL
	Dionaea	Cowrie	Glutton	others	
Amazon Web Services	228,075	4,503	11,878	3,688	248,144
Google Cloud Platform	162,570	297,818	84,375	36,403	581,116
Microsoft Azure	308,102	9,012	17,256	6,365	340,735

The overall results show an average ratio of 55,000 attacks per day, summing up to roughly 1.17 million in total. Similar results for different regions could have been reproduced. Their results clearly show the Europe, US, and Asia disparity. An important question that Kelly et al. [40] answered is if attackers target services on cloud providers based on the cloud providers' market share. The study could not confirm this assumption because Google Cloud received most of the attacks with the smallest market share. In total, most of the attacks are originated from Vietnam, Russia, the United States, and China. Due to technologies such as VPN or Tor, the geolocation only indicates the last node so that location data might be distorted. Across all providers, roughly 80% of the source IP addresses had a bad reputation (identified by Suricata) and could have been filtered by the organization. The operating devices used for attacking the services are mostly Windows 7 or 8 and different Linux kernels and distributions. Windows devices target vulnerabilities in remote desktop sharing software. Such vulnerabilities are (i) CVE-2006-2369[14] (RealVNC) in the US region, (ii) CVE-2001-0540[11] (Remote Desktop Protocol (RDP)) in EU and Asia regions, (iii) CVE-2012-0152[15] (RDP) in the Asia region, and (iv) CVE-2005-4050[13] (Voice over Internet Protocol (VoIP)) in EU region. In addition, attackers were also capable of disguising any fingerprinting activity of P0f.

This chapter compares the findings Kelly et al. [40] claimed in the paper "A Comparative Analysis of Honeypots on Different Cloud Platforms" with ours using the Heidelberg University's cloud solution. First, a short introduction of heiCLOUD is given, followed by a closer lookup of the T-Pot used to acquire data. Lastly, it presents the results and does a thorough comparison closing up with a discussion based on a technical report of Cambridge University.

3.2 Methodology

The foremost goal is to track as many attacks as possible. Figure 3.1 sketches the concept to achieve this goal to gather various attacks from the Internet. Honeypots should be deployed on a single instance, and their data or log files are stored in a database. The attacks are analyzed with the help of data visualization tools. For security reasons, honeypots should run in a virtualized environment to avoid harming the host system. The host machine runs on a Debian distribution. The instance runs on heiCLOUD, a cloud service provided by Heidelberg University. It is capable of 16 GB of RAM, 8 vCPUs, and volatile memory of 30 GB. In addition, it mounts a 125 GB permanent volume to store the data securely. In the very early stage of this chapter, different approaches to achieve this goal have been compared. For example, native implementation approaches, additional frameworks, and ready-to-use solutions have been evaluated. However, the T-Pot, developed by Telekom, offers a profoundly ready-to-use solution with significant advantages. It combines several honeypots with various analytic tools to trace the newest attacks. Furthermore, it helps to compare the findings with the ones Kelly et al. [40] claim.

Running the instance and exposing it to the Internet needs some adjustments beforehand. Therefore, a virtual network with subnet `192.168.145.0/24` has been created wherein the IP address `192.168.145.4` is assigned to the instance. The instance is accessible from the outside with a floating IP address `129.206.5.74`. Access rules are similar to a stateless firewall, and thus, do not block any attacks. Ports `1–64000` are exposed and can be attacked by anyone. Ports higher than `64000` are only accessible through the university network `129.206.0.0/16` or eduroam `147.142.0.0/16` and should provide a basic authentication with username and password.

3.2.1 heiCLOUD

University Computing Center Heidelberg offers a “IaaS specially tailored for higher education and research institutions”[69] called heiCLOUD. It supplies multiple departments at Heidelberg University with storage, virtual machines, or network components. In addition, heiCLOUD is a DFN¹ member and offers others to use their services. As stated on their information website[68], it (i) is capable of freely manageable IT resources, (ii) beholds a stable and fast connection, (iii) ensures high availability and scalability, (iv) has freely selectable VM operating systems, and (v) has a transparent payment model [68]. Users can easily create their network areas and manage their space individually based on the open-source application OpenStack. Unlike well-known cloud providers, heiCLOUD servers are located within Germany,

¹German National Research and Education Network is the communications network for Science and research in Germany

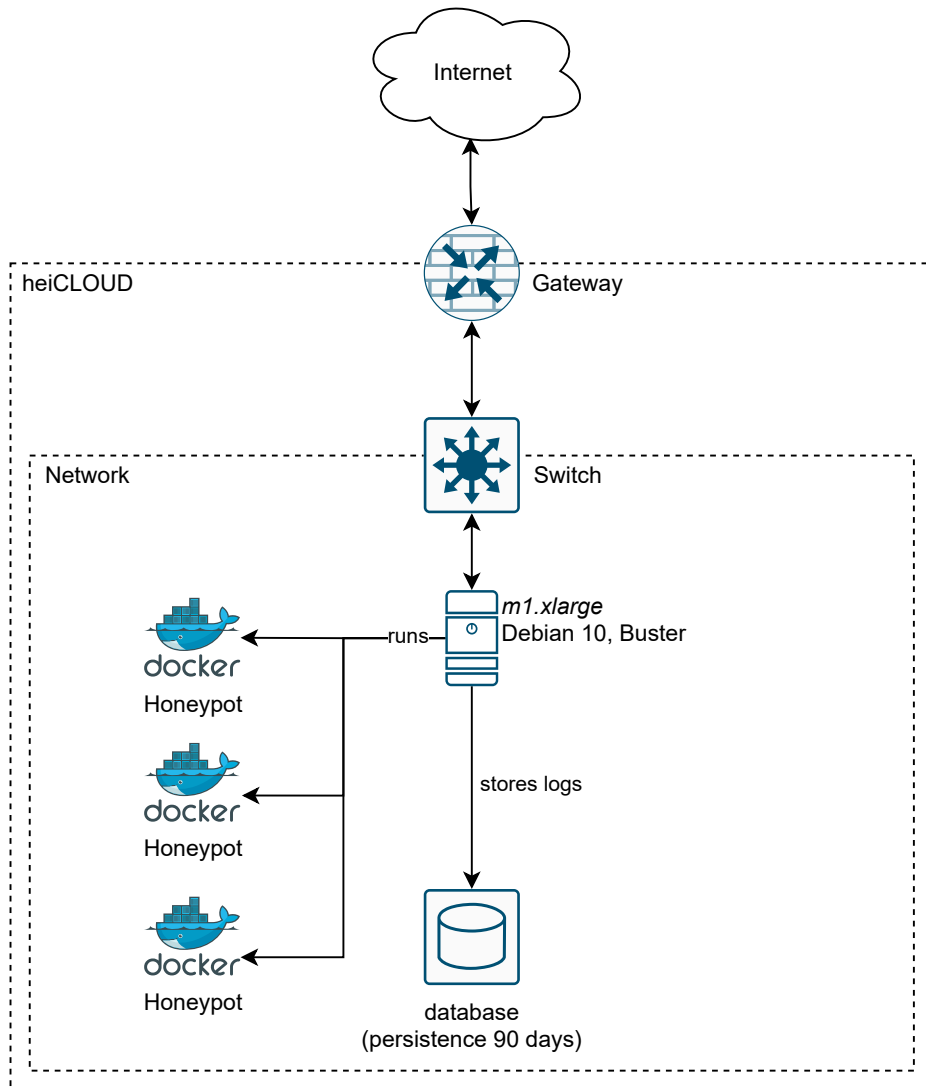


Figure 3.1: Concept to collect honeypot attacks. The instance size is referred to the available resources of OpenStack. The network is an encapsulated subnet with a switch for incoming and outgoing connections. The database is independent of the instance and could run on a separate host.

thus, abide by the European data privacy law. HeiCLOUD has never considered implementing honeypots for additional cybersecurity measurements.

3.2.2 T-Pot

To be able to compare the results with Kelly et al. [40], the same approach to capture recent cyberattacks is used. The T-Pot solution, a mixture of Telekom and Honeypot, stands out with its sheer quantity of various honeypots. It requires at least 8 GB of RAM and a minimum of 128 GB of hard drive storage. Based on a Debian 10 Buster distribution, it relies on Docker to run their services [25]. T-Pot has to be deployed in a reachable network where intruders are expected. Either Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) traffic are forwarded without filtering to the network interface, or it runs behind a firewall with forwarding rules. Specified ports for attackers are 1-64000; higher ports are reserved for trusted IPs; thus, a reverse proxy asks for basic authentication. All daemons and tools run on the same network interface, but some are encapsulated in their own Docker network. The lightweight virtualization technology Docker uses containers to run on the host system [9]. Unlike virtual machines, Docker reduces overhead with the downside of a greater attack surface. To mitigate attacks, Docker wraps containers in an isolated environment. This is achieved by restricting the kernel namespace and control groups (`cgroups`) [9]. Figure 3.2 visualizes the technical concept of T-Pot. Each service has dedicated ports or port ranges that are exposed. Attackers can communicate either with TCP or UDP. All honeypots and tools create log files used to get any knowledge about attackers. In order to view and trace current attacks, T-Pot uses the ELK stack. ELK is the acronym of Elasticsearch, Logstash, and Kibana [26]. The search engine Elasticsearch is based on the Lucene library. It is multitenant-capable and offers full-text search via HTTP. Logstash is used to feed Elasticsearch. In general, it offers an open server-side data processing pipeline that helps to send data from multiple sources to an Elasticsearch node. Kibana is the primary data visualization tool. It enables users to create plots and dashboards, crawl Elasticsearch, and trace the system's health. All logs of the honeypots and tools are forwarded to the search engine Elasticsearch by Logstash. The ELK stack is not directly exposed to the Internet; thus, authentication is unnecessary. Users can monitor all log files with Kibana by pre-defined dashboards or custom search queries. In addition, T-Pot features different services types, namely (i) standard, (ii) sensor, (iii) industrial, (iv) collector, (v) next generation, and (vi) medical. Each service type has a different set of honeypots and tools tailored to its core idea. T-Pot feeds their data to an external Telekom service; however, this data submission can be turned off. The latest version, 20.06.0, has been used in this chapter. Newer versions might be available by the end of this study and could differ from this.

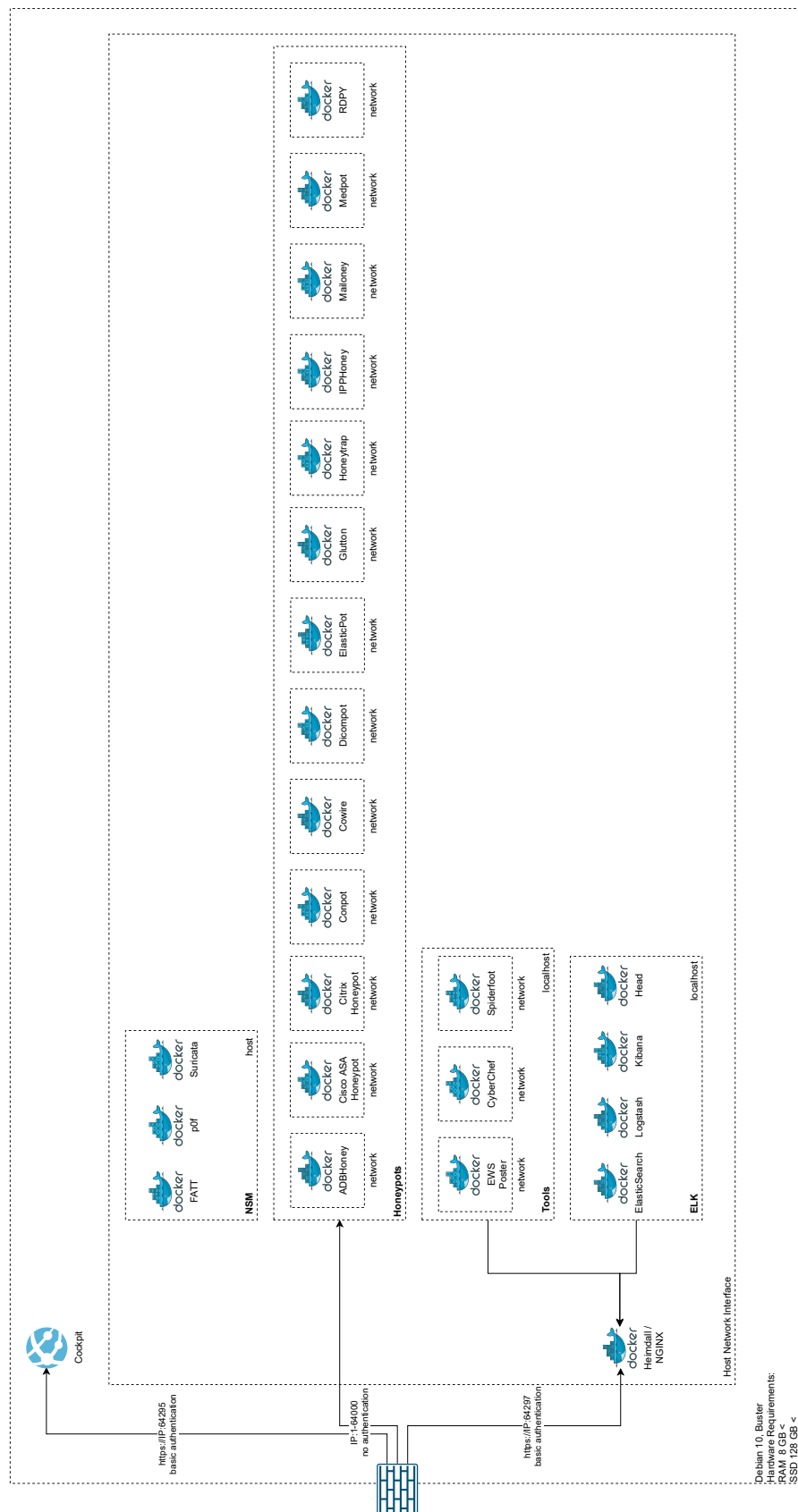


Figure 3.2: T-Pot architecture derived from [52]. Honeybots are encapsulated in their network. NSM runs on the host network, and thus, receives every packet. ELK and tools run on localhost and are accessible through NGINX. The Cockpit application is a web-based graphical interface for servers.

Honeypots

T-Pot consists of 20 honeypots. Albeit the sheer quantity of it, a short explanation is given. In addition, Table 3.2 gives a quick overview of all available honeypots in conjunction with (i) the port they are running on, (ii) their interaction level, and (iii) a short description.

ADBHoney [8] is a low-interaction Android Debug Bridge (ADB) honeypot over TCP/IP. The importance of it lies in the ADB protocol that is used for debugging and pushing content to an Android device. However, unlike a Universal Serial Bus (USB) connection, it does not support any kind of ample mechanisms of authentication and protection. By exposing the ADB service over any port, an adversary could connect and exploit it. ADBHoney is designed to catch malware that has been pushed onto devices.

Cisco Adaptive Security Appliance (ASA) [57] is a low-interaction honeypot that detects CVE-2018-0101[16]. It is a vulnerability that could allow an unauthenticated, remote attacker to cause a reload of the affected system and remotely execute code. This can be achieved by flooding a webvpn-configured interface with crafted Extensible Markup Language (XML) packets. Consequently, the attacker obtains full control by executing arbitrary code.

Citrix Application Delivery Controller (ADC) honeypot [34] detects and logs CVE-2019-19781[18] scans and exploitation attempts. This vulnerability allows adversaries to perform directory traversal attacks. Files are accessible by path strings to denote the file or directory. In addition, some file systems include special characters to traverse the hierarchy easily. Attackers take advantage of it by combining special characters to get access to restricted areas. [30]

Conpot [58] is a low-interaction industrial honeypot for Industrial Control System (ICS), and Supervisory Control and Data Acquisition (SCADA). It provides a variety of different standard industrial control protocols. An adversary should be tricked by the complex infrastructure and lured into attacks. In addition, a custom human-machine interface can be connected to increase the attack surface. By randomly delaying the response time, Conpot tries to emulate a real machine handling a certain amount of load.

Cowrie [53] is a medium- to high-interaction SSH and Telnet honeypot. It offers to log brute-force attacks and shell interactions with attackers. In medium-interaction mode, Cowrie emulates a Unix shell in Python, whereas in high-interaction mode, it proxies all commands to another system.

DDoSPot [22] is a low-interaction honeypot to log and detect UDP-based Distributed Denial of Service (DDoS) attacks. It is a platform used to support various plugins for different honeypot services and servers. Currently, it supports Domain

Name System (DNS), Network Time Protocol (NTP), Simple Service Discovery Protocol (SSDP), Character Generator Protocol (CHARGEN), and random/mock UDP server.

Dicompot [41] is a low-interaction honeypot for the Digital Imaging and Communications in Medicine (DICOM) protocol. As with other honeypots before, it mocks a DICOM server in Go to collect logs and detect attacks.

Dionaea [24] is a medium-interaction honeypot that tries to capture malware copies by exposing services. It supports various protocols such as FTP, Server Message Block (SMB), and HTTP. Several modules can be integrated to work with Dionaea for further malware results, such as VirusTotal.

Elasticpot [4] is a low-interaction honeypot for Elasticsearch, a search engine based on the Lucene library.

Glutton [59] is a generic low-interaction honeypot that works as a man-in-the-middle (MITM) for SSH and TCP. However, lacking documentation does not provide a deeper insight into this honeypot.

Heralding [71] is a credential catching honeypot for protocols like FTP, Telnet, SSH, HTTP, or Internet Message Access Protocol (IMAP).

HoneyPy [32] is a low to medium-interaction honeypot that supports several protocols such as UDP or TCP. New protocols can be added by writing a custom plugin for them. HoneyPy gives the freedom of quickly deploying and extending honeypots.

HoneySAP [32] is a low-interaction honeypot tailored for SAP services.

Honeytrap [75] is a low-interaction honeypot network security tool. As stated by Werner [75], Honeytrap is vulnerable to buffer overflow attacks.

IPPHoney [5] is a low-interaction Internet Printing Protocol (IPP) honeypot.

Mailoney [46] is a low-interaction Simple Mail Transfer Protocol (SMTP) honeypot written in Python.

MEDpot [62] is a low-interaction honeypot focused on Fast Healthcare Interoperability Resources (FHIR). It is a standard description data format to transfer and exchange medical health records.

RDPY [54] is a low-interaction honeypot of the Microsoft RDP written in Python. It features client and server-side, and it is based on the event-driven network engine Twisted. It supports authentication over Transport Layer Security (TLS) and Network Level Authentication (NLA).

SNARE and TANNER [60, 61] is a honeypot project. SNARE is an abbreviation for Super Next-generation Advanced Reactive honEypot. It is a successor of

Glastopf, a web application sensor. In addition, it supports the feature of converting existing web pages into attack surfaces. TANNER [61] can be seen as SNARES' brain. Whenever a request has been sent to SNARE, TANNER decides how the response should be.

Tools

T-Pot integrates tools to screen network traffic and block DoS attacks.

FATT [37] is used to extract metadata and fingerprints such as JA3 [2] and HASSH [38] from captured packets. JA3 is a method for “creating SSL/TLS client fingerprints” whereas HASSH is a network fingerprinting standard that is used to identify specific client and server SSH implementations. In addition, it features live network traffic. As noted by the author, FATT is based on a python wrapper for tshark, namely pyshark, and thus has performance downturns. T-Pot applies FATT on every request made on the host network.

Spiderfoot [48] is an open-source intelligence automation tool that helps to screen targets to get information about what is exposed over the Internet. It can target different entities such as IP address, domain, hostname, or network subnet. In addition, it features more than 200 modules that can be integrated as an extension. T-Pot uses it to scan defensively and thus not include any other module.

Suricata [67] is “a high performance IDS, intrusion prevention system (IPD) and network security monitoring (NSM) engine”. T-Pot lets Suricata analyze and assess any request made on the host network.

P0f [77] is a fingerprinting tool that uses passive traffic fingerprinting mechanisms to check TCP/IP communications. T-Pot lets P0f passively check any request made on the host network.

Endlesssh [74] is an SSH server that sends an endless, random SSH banner. The key idea is to lock up SSH clients that try to connect to the SSH server. It lowers the transaction speed by intentionally inserting delays. Due to the established connection before the cryptographic exchange, this module does not require any cryptographic libraries.

HellPot [35] is an “endless honeypot”. If someone connects to this honeypot, it results in a memory overflow. Its key idea is to send an endless data stream to the attacker until its memory or storage runs out.

Table 3.2: Overview of all available honeypots of T-Pot with interaction level, port, and a short description. Ports are marked with either TCP or UDP; if a port misses any definition, both TCP and UDP are allowed.

HONEYPOTS		Port	Interaction-level	Description
ADBHoney [8] Cisco ASA [57]		5555/TCP	low	ADB protocol honeypot
		5000/UDP, 8443/TCP	low	honeypot for CVE-2018-0101[16] de-tection
Citrix honeypot [34]		443/TCP	low	detects and logs CVE-2019-19781[18] scans and exploitation attempts
Conpot [58]		80, 102, 161, 502, 623, 1025, 2404, 10001, 44818, 47808, 50100	low	industrial honeypot for ICS and SCADA
Cowrie [53]		2222, 23	high	SSH and Telnet honeypot
DDoSPot [22]		1112/TCP	low	log and detect UDP-based DDoS attacks
Dicompot [41]		1112/TCP	medium	honeypot for the DICOM protocol
Dionaea [24]		21, 42, 69/UDP, 8081, 135, 443, 445, 1433, 1723, 1883, 1900/UDP, 3306, 5060/UDP, 5061/UDP	low	capture malware copies
Elasticpot [4]		9200	low	honeypot for Elasticsearch
Glutton [59]		NFQ	medium	MitM proxy for SSH and TCP
Heralding [71]		21, 22, 23, 25, 80, 110, 143, 443, 993, 995, 1080, 5432, 5900	low	credential catching honeypot
HoneyPy [32]		7, 8, 2048, 2323, 2324, 4096, 9200	low	extendable honeypot
HoneySAP [32]		3299/TCP	low	honeypot for SAP services
Honeytrap [75]		NFQ	medium	captures attacks via unknown protocols
IPPHoney [5]		631	low	IPP honeypot
Mailoney [46]		25	low	SMTP honeypot
MEDpot [62]		2575	low	FHIR honeypot
RDPY [54]		3389	low	Microsoft RDP honeypot
SNARE/TANNER [60]		80	low	web application honeypot

3.3 Results

The T-Pot has been deployed for three weeks (from 26th of September to 16th of October) and collected in total 607,747 attacks. Overall, RDPY (46.08%), Honeytrap (33.23%), and Cowrie (12.42%) received most of the attacks with a total amount of 540,398 attacks. Figure 3.3 shows the distribution of honeypot attacks. The total numbers are based on Table 3.3.

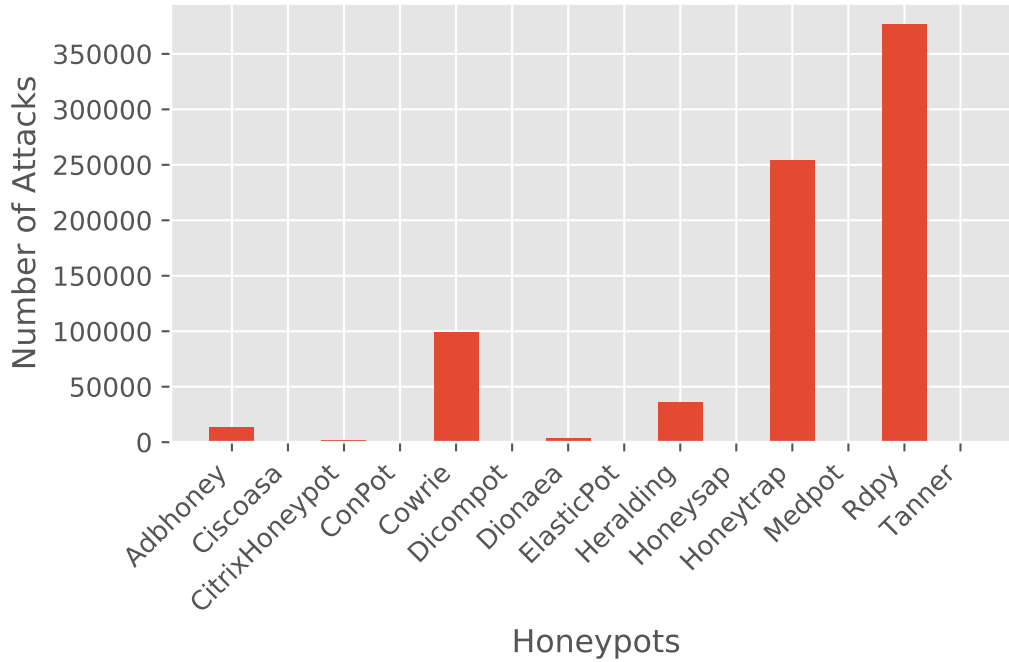


Figure 3.3: Distribution of honeypot attacks. Timestamp; 26th of September to 16th of October. A description of each honeypot can be found in section 3.2.2.

What is striking is the large disparity between the previously mentioned attacks on AWS, GCP, and Azure. Especially with the honeypot Dionaea, it is unclear why only 2,368 attacks have been performed. 96% of IP addresses connected to Dionaea are known attackers, and 70% were acquired on port 81, unofficially known for Tor routing. Neither any malware nor suspicious payload could be identified. An assumption is that the packets run through a static filter. Heidelberg has a centralized stateless firewall, indicating that specific ports or protocols are excluded. A NMAP TCP SYN scan (`nmap -sS -A 129.206.5.74`) has been performed to prove this assumption that ports are excluded. The result clearly shows that port 139 for SMB is filtered, although the access security explicitly allows it. The stateless firewall runs in front of heiCLOUD and filters many ports, including 113. Based on this, it can be assumed that most of the attacks on Dionaea are carried out via

SMB, which would explain the total number of attacks. The administrator of the university firewall had been consulted to exclude the T-Pot instance to validate if the actual number is even higher without any packet filter in front of it. Respectively, no stateless packet filter has been applied to the T-Pot for three weeks (2nd of December until 23rd of December). It could identify a drastic increase in Dionaea attacks with a total number of 213,053. Overall, 93% of all attacks are on the SMB protocol followed by many database protocols such as MongoDB and MSSQL. This confirms the assumption that a higher total number of attacks would be the result without the packet filter in front of the instance.

Comparing the number with Kelly et al. [40] it shows that Dionaea attacks surpass every other cloud provider. However, Dionaea attacks will not be included in later results because usually, a server is not allowed to be excluded from the university firewall. Only for this research purpose to assess the effect of the packet filter has an exclusion been granted.

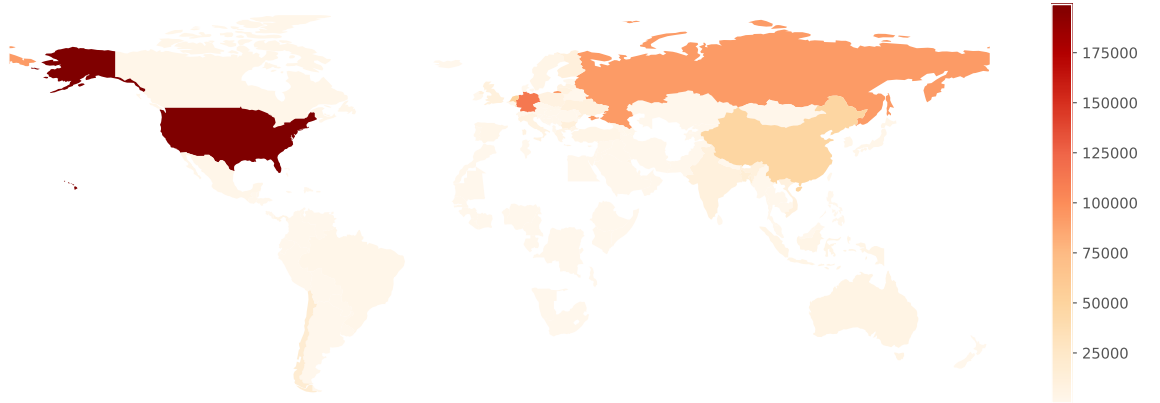


Figure 3.4: Attack distribution of T-Pot. The USA, Russia, China, and Germany are the most attacking countries. Timestamp; 26th of September to 16th of October.

Logstash uses GeoLite2 to resolve the source IP address with information such as location, Autonomous System Number (ASN), continent code, country name, and Autonomous System (AS) organization. Figure 3.4 indicates the geographical location of connections acquired to any honeypot. Most attacks are originated from the United States, Germany, Russia, and China. Large security scans of DFN or Baden-Württembergs extended LAN (BelWÜ) pushes Germany to second place; therefore, Germany can be considered negligible. On the contrary, the geographical location of an IP address merely indicates the true origin. Due to technologies like VPN or Tor, the last known node of an IP address could be spoofed, and thus as stated by Kelly et al. [40], would remain insufficient to use. Hence, no one should rely on geographical information.

Attacks are not equally distributed among all honeypots, and thus, different protocols and applications receive more attention than others. Figure 3.5 shows the time-

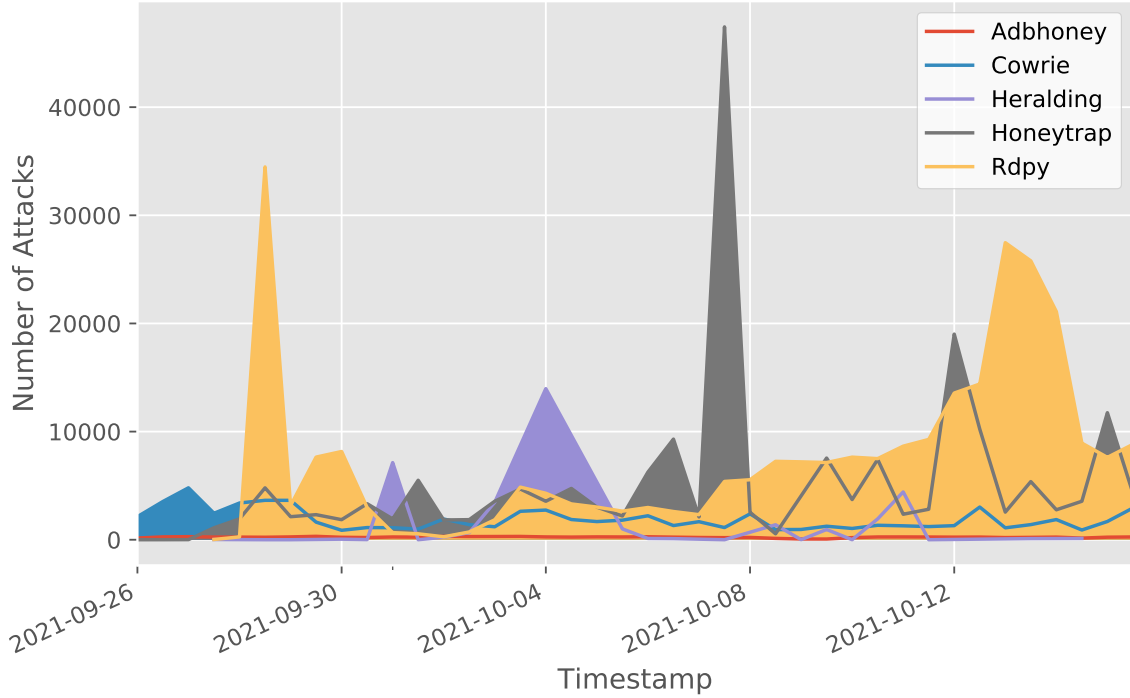


Figure 3.5: Attack histogram of T-Pot. Only the five most attacked honeypots are considered. Timestamp; 26th of September to 16th of October. A description of each honeypot can be found in section 3.2.2.

line of attacks that are executed on our instance separated by honeypots. RDPY, Honeytrap, and Cowrie are the most attacked honeypots. The high peak of Honeytrap in the middle indicates a full NMAP scan from Germany that has been done to get an insight of the packet filtering at the Heidelberg University. It identifies a bias towards remote desktop protocol attacks, shell-code exploitations, and commands to retrieve information about the CPU, scheduled tasks (`cat /proc/cpuinfo`, or `crontab`), or privilege escalation.

Suricata registered several alerts and CVEs. The vast majority of alerts are RDP-related policies, Virtual Network Computing (VNC) authentication failures, and NMAP scans. Most used vulnerabilities are (i) CVE-2001-0540[11] which is a memory leak in terminal servers in Windows NT and Windows 2000 causing a denial of service (memory exhaustion) by malformed RDP requests, (ii) CVE-2006-2369[14] which is a RealVNC vulnerability allowing hackers to bypass authentication, and (iii) CVE-2012-0152[15] which enables attackers for RDP in Microsoft Windows Server 2008 R2 and R2 SP1 and Windows 7 Gold and SP1 to cause a denial of service by sending a series of crafted packets. As derived from Figure 3.6, the T-Pot has not received many attacks in the first week. Starting from the 28th of September, the number of alerts is skyrocketing. This would indicate that bots crawl IP address ranges to find new machines and probe them. Interestingly, zero-day exploits like the Apache vulnerability [20] that came with version 2.49.0 got registered in CVE on

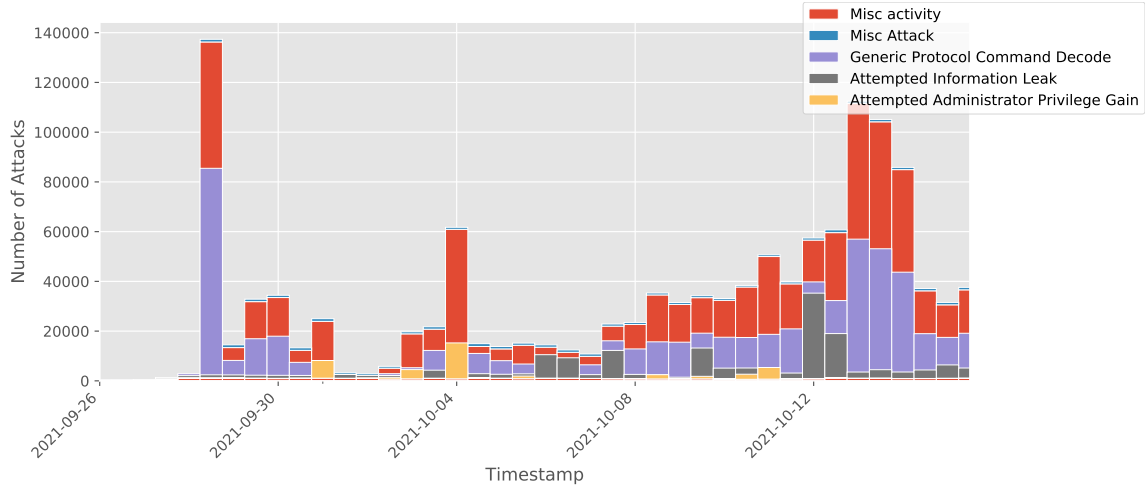


Figure 3.6: Suricata results of T-Pot. Displays the five most listed alert categories. Timestamp; 26th of September to 16th of October.

the 6th of October and immediately recognized by Suricata on the 15th of October. Attackers could perform a remote code execution using path traversal attacks when the Common Gateway Interface (CGI) scripts of Apache are enabled. The logs could trace back similar attacks like `/cgi-bin/.\\%2e/\\%2e/\\%2e/bin/sh` until the 7th of October, leaving an even smaller time frame to adapt to new exposures. This shows how fast bots adapt to new vulnerabilities to compromise more systems.

The results from RDPY in Figure 3.7 backups the assumption that attacks originate from bots. It shows that only a small margin represents unique source IP addresses. The rest of the attacks result in either a bad reputation, bot, crawler, or known attacker. Figure 3.6 shows the distribution of alert categories that Suricata identified. Respectively, misc activities sum up to roughly 1.5 million entries, RDP related alerts account for two-thirds of it. Several RDP attacks from 2021 back to 2001 had been executed on the T-Pot. Respectively, CVE-2012-0152 and CVE-2001-0540 coincide with the ones Kelly et al. [40] claim.

For NFQ related attacks, Honeytrap could identify three major services that are not provided by default. Honeytrap functions as a honeypot to provide a service on ports that are not specified by default. NFQ intercepts incoming TCP connections during the TCP handshake, and Honeytrap provides a service for it. Most of these interceptions are made on (i) port 5038, which is used by a machine learning database called MLDB, (ii) port 5905, which an Intel Online Connect Access uses on Windows machines, and (iii) port 7070 which is used by Apple’s QuickTime streaming server (RTSP). Nearly all ports attacks focused on RDP connection attempts (`Cookie: msthash=Administr`). However, 94% of all connected IP addresses on Honeytrap are resolved as known attackers.

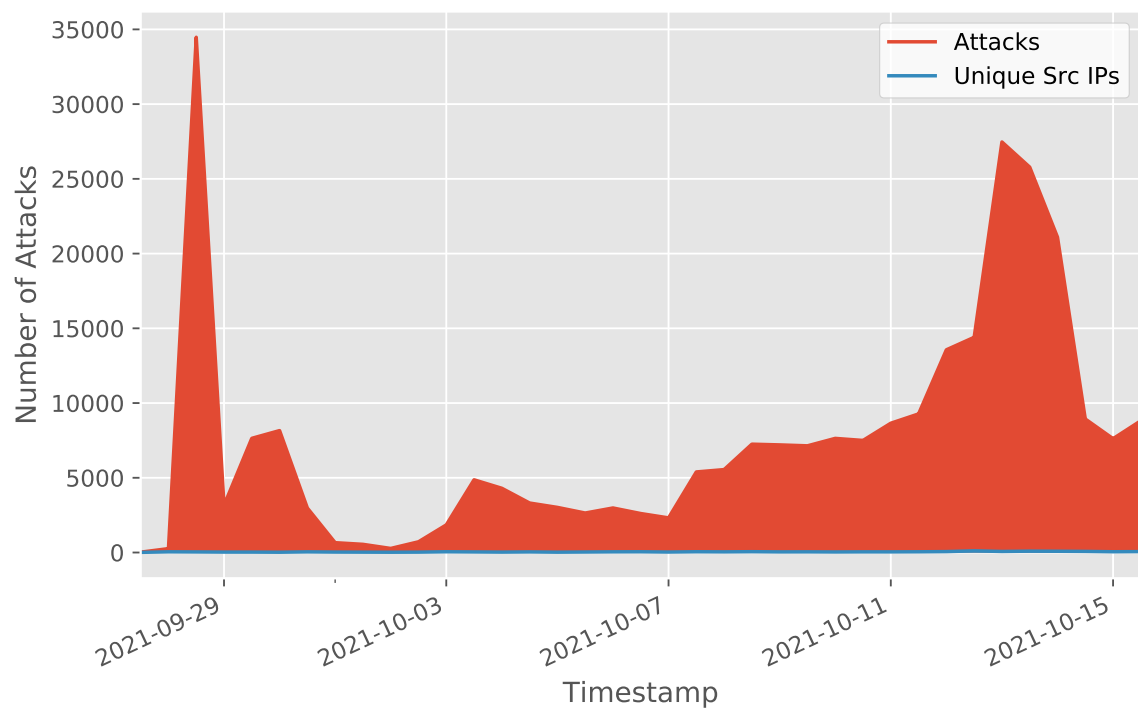


Figure 3.7: RDPY attacks are separated into attacks and unique source IP addresses. Timestamp; 26th of September to 16th of October. A description of the honeypot can be found in section 3.2.2.

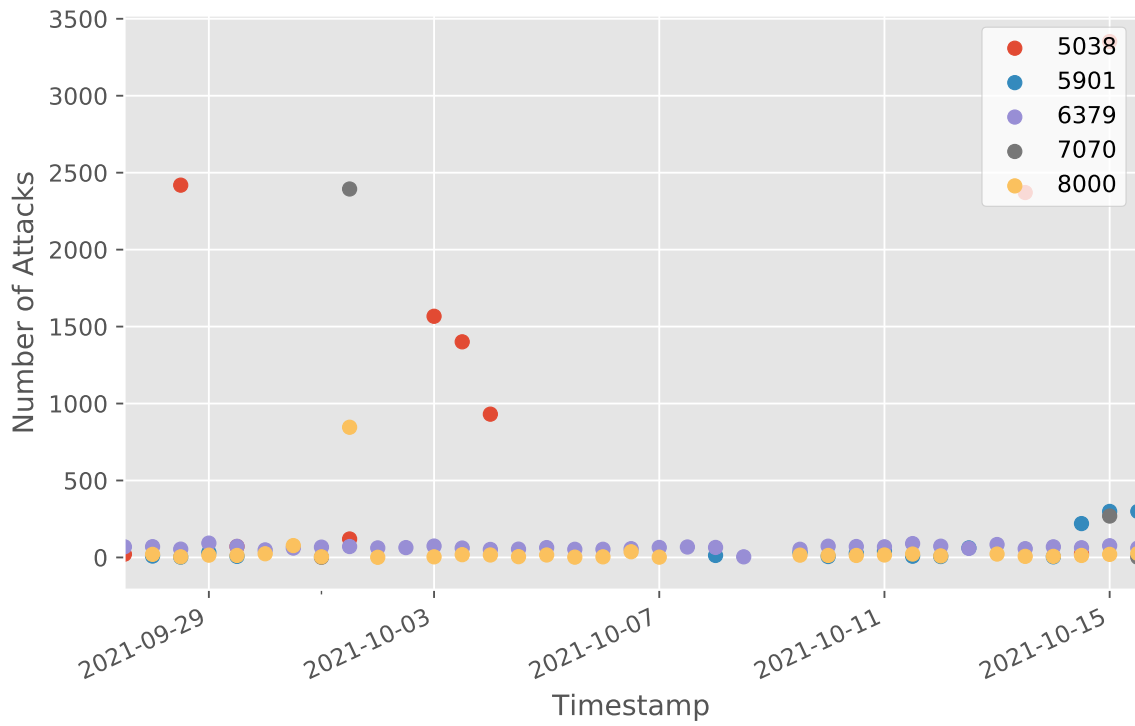


Figure 3.8: Honeytrap results of T-Pot. Timestamp; 26th of September to 16th of October. A description of the honeypot can be found in section 3.2.2.

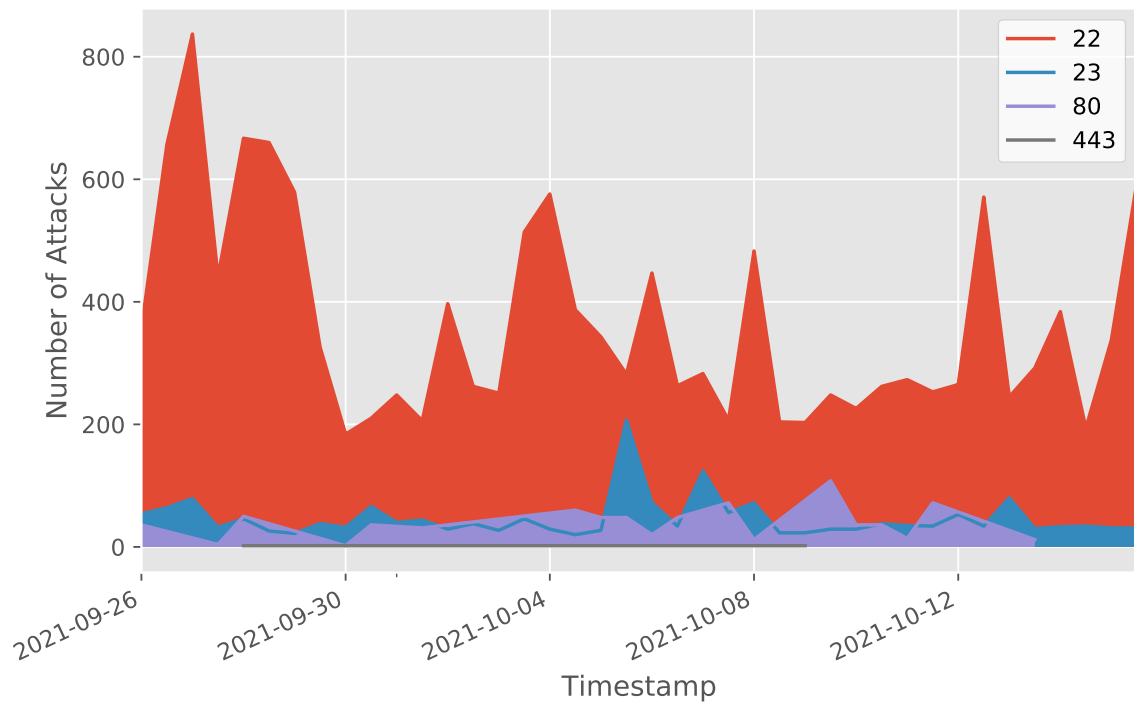
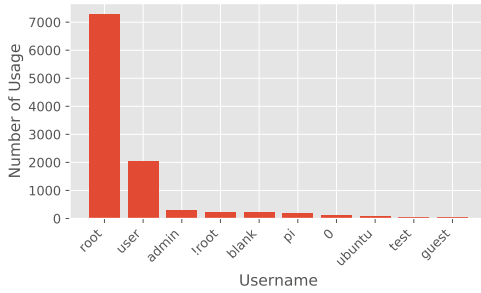
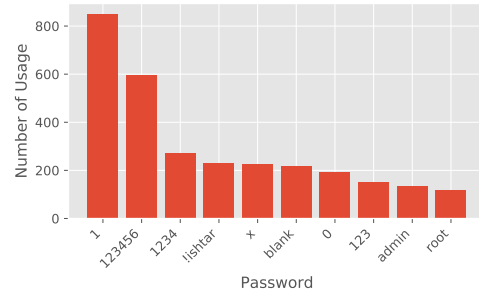


Figure 3.9: Cowrie results of T-Pot. Timestamp; 26th of September to 16th of October. A description of the honeypot can be found in section 3.2.2.

The third most compromised honeypot is Cowrie, with a strong bias towards SSH and FTP. Figure 3.9 shows all attacks executed on Cowrie separated by their port. Respectively, SSH port 22 is the most considered port, resulting in high use for privilege escalation. Besides using default credentials to log in (`username: root`, `password: root`, see Figure 3.10 for top 10 credentials), adversaries used various commands to retrieve any information about the host system (`nproc;uname -a`, `cat /proc/cpuinfo`). A unique information gathering attack could be identified that has been widely used on the T-Pot. Listing 3.1 shows all shell commands that are executed. Attackers try to gain knowledge about running processes on the system (`/bin/busybox`). Interestingly, crypto mining attacks are getting more attractive to criminals. For example, XMRig has been the most downloaded malware for cryptocurrency mining. Some adversaries even executed complex tailored shell commands to exploit the host machine as a crypto miner (Listing 3.2). It is not surprising that such attacks gain attraction concerning the current time. Attackers could exploit machines for crypto mining in order to earn more money. This looks more appealing than acquiring mining machines and hijacking electricity from surrounding apartments.



(a) Cowrie username credentials



(b) Cowrie password credentials

Figure 3.10: Cowrie top 10 credentials used on T-Pot. Timestamp; 26th of September to 16th of October. A description of the honeypot can be found in section 3.2.2.

P0f identified different Windows versions and Linux distributions in conjunction with various SSH clients to compromise the T-Pot. Like Kelly et al. [40] presented, Windows 7 or 8 and Windows NT Kernel are the most used OS with 81%. Unfortunately, disguising OS fingerprinting activities account for 84% of all fingerprints. Lastly, the results are cleaned up and all IPs from DFN and BelWÜ are excluded. Both scan frequently and check if any vulnerability exists. This distorts the findings, and thus, they have been filtered based on their subnet addresses. However, the results show no notable changes. The total number of attacks was hardly influenced by it. This indicates that these scans do not greatly interfere with the findings.

On average, heiCLOUD has received 55.83% more than Azure, GCP, and AWS. Attacks on Cowrie, RDPY, and Honeytrap are the most compromised honeypots. In contrast to Kelly et al. [40], Dionaea and Glutton used to be the most considered honeypots for adversaries. It can be assumed that attacks by bots had increased significantly since last year when Kelly et al. [40] did their research. Respectively, one unresolved question is if other cloud providers filter their network traffic. It would explain the major difference between Heidelberg University and the big tech companies. The cause for such an increase remains doubtful. One explanation could root back to the Corona pandemic and the skyrocketing increase in home office activities. Related to that is a higher usage in screen sharing software. Considering the BSI², report for cybersecurity 2021 [6], they revealed an increase of attack surfaces during the pandemic. Respectively, the IT infrastructure could not keep up with this fast change and widen the company’s attack surface. Their conclusion overlays our assumption that attackers took advantage and increased their activities. This phenomenon shows that nearly all attacks originate from bots that scan through IP address ranges. In total, 73% of all IP addresses are unresolved. The known attacker reputation represents the largest part of resolved IP addresses with 23%. Fortunately, such reputations could technically be filtered by an organization’s firewall and would lower the chance of an exploit. Interestingly, after three weeks, the number of attacks originating from China decreased to almost zero percent. This might indicate that the honeypot has been exposed, and further attacks represent a risk of revealing their compromises. However, this assumption cannot be confirmed due to the lax geographical reliability of IP addresses.

Our results emphasize the importance of honeypots. It gives a proper security measure of an IT infrastructure and helps to identify potential leaks or vulnerabilities. Moreover, it shows that T-Pot helps detect recent bot activities and gives an outlook on the newest trends of attacks.

3.4 Discussion

One downside of T-Pot is the static hostname representation of Cowrie. It always returns `#1 SMP Debian 3.2.68-1+deb7u1 (uname -a)` as hostname information, leaving a tiny footprint when bots crawl through the web. A random choice of hostname information could harden Cowrie from being exposed. Next, if attackers scan open ports on T-Pot, it might be suspicious when many ports with services are open. From a technical perspective, bots could check this state if it is uncommon and thus, exclude T-Pot from being probed. However, T-Pot includes reasonable preventions like a random hostname and scheduled tasks. Another major drawback

²The Federal Office for Information Security is responsible for managing communication security for the German government. Each year they publish a report for recent cybersecurity threats.

Table 3.3: Overview of attacks on heiCLOUD, AWS, GCP, and Azure. Only the top 10 most attacked honeypots are considered. “-” entails that a honeypot is not part of the top 10. The red and green arrows indicate whether heiCLOUD received more or fewer attacks than the other cloud providers on average .

HONEYPOTS	BASIS		COMPARISON					
	HEICLOUD		AWS		GC		AZURE	
	Number	Pct.	Number	Pct.	Number	Pct.	Number	Pct.
ADBHoney	9,302	1.65% ↑	413	0.17%	2,497	0.43%	442	0.13%
Cisco ASA	674	0.11% ↑	260	0.10%	750	0.13%	134	0.04%
Citrix honeypot	1,121	0.18%	-	-	-	-	-	-
Conpot	615	0.10%	-	-	-	-	-	-
Cowrie	75,511	11.97% ↓	4,503	1.81%	297,818	51.25%	9,012	2.64%
DDoSPot	0	0%	-	-	-	-	-	-
Dicompot	22	0%	-	-	-	-	-	-
Dionaea	2,368	0.40% ↓	228,075	91.91%	162,570	27.98%	308,102	90.42%
Elasticpot	385	0.06%	-	-	-	-	-	-
Glutton	0	0%	11,878	4.79%	84,375	15.52%	17,256	5.06%
Heralding	35,680	4.34% ↑	1,885	0.76%	12,255	2.11%	3,370	0.99%
HoneyPy	0	0%	172	0.07%	2,149	0.37%	497	0.15%
HoneySAP	15	0%	-	-	-	-	-	-
Honeytrap	201,949	32.01%	-	-	-	-	-	-
IPPHoney	0	0%	-	-	-	-	-	-
Mailoney	0	0%	720	0.29%	9,419	1.62%	146	0.04%
MEDpot	2	0%	-	-	-	-	-	-
RDPY	280,040	49.15% ↑	100	0.04%	7,916	1.36%	1,463	0.43%
SNARE/TANNER	63	0.02% ↓	138	0.06%	1,367	0.24%	313	0.09%
IN TOTAL	607,747	100%	248,144	100%	581,116	100%	340,735	100%

is the latest endeavor to detect honeypots on the transport level. As recently investigated by Vetterl [72], detecting honeypots is becoming easier due to a fatal flaw in the underlying protocol implementation. Vetterl [72] states that attackers always try to prevent their methods, exploits, and tools from being divulged. Therefore, detecting honeypots before attacking them strongly motivates black hats. Chapter 5 will present a way to avoid such fingerprint activities with the honeypot Cowrie.

Listing 3.1: Cowrie attack to gather various information about the system.

```
1 enable
2 system
3 shell
4 sh
5 cat /proc/mounts; /bin/busybox $PROCESS_NAME
6 cd /dev/shm; cat .s || cp /bin/echo .s; /bin/busybox
  ↪ $PROCESS_NAME
7 tftp; wget; /bin/busybox $PROCESS_NAME
8 dd bs=52 count=1 if=.s || cat .s || while read i; do echo
  ↪ $i; done < .s
9 while read i
10 /bin/busybox $PROCESS_NAME
11 rm .s; exit
```


Listing 3.2: Cowrie attack to exploit the host machine as a crypto miner.

```
1 mkdir -p /home/osmc/.ssh/
2 echo ssh-rsa $RSA_KEY >> /home/osmc/.ssh//
  ↪ authorized_keys
3 echo '<cmd7uname>'; uname -a
4 echo '</cmd7uname><cmd7uptime>'; uptime
5 echo '</cmd7uptime><cmd7w>'; w
6 echo '</cmd7w><cmd7who>'; who
7 echo '</cmd7who><cmd7last>'; last
8 echo '</cmd7last><cmd7lastlog>'; lastlog
9 echo '</cmd7lastlog><cmd7authkey>'; cat /home/osmc/.ssh//
  ↪ authorized_keys
10 echo '</cmd7authkey><cmd7lshome>'; ls -la /home
11 echo '</cmd7lshome><cmd7passwd>'; cat /etc/passwd
12 echo '</cmd7passwd><cmd7shadow>'; sudo -n cat /etc/shadow
13 echo '</cmd7shadow><cmd7psfaux>'; ps -faux
14 echo '</cmd7psfaux><cmd7netstat>'; netstat -npta
15 echo '</cmd7netstat><cmd7arpan>'; /usr/sbin/arp -an
16 echo '</cmd7arpan><cmd7ifconfig>'
17 /usr/sbin/ifconfig
18 echo '</cmd7ifconfig><cmd7localconf>'; cat /home/ethos/
  ↪ local.conf
19 echo '</cmd7localconf><cmd7remoteconf>'
20 cat /home/ethos/remote.conf
21 echo '</cmd7remoteconf><cmd7rclocal>'
22 cat /etc/rc.local
23 echo '</cmd7rclocal><cmd7claymorestub>'; cat /home/ethos/
  ↪ claymore.stub.conf
24 cat /hive-config/rig.conf; cat /hive-config/wallet.conf
25 cat /hive-config/vnc-password.txt
26 echo '</cmd7claymorestub><cmd7claymorezstub>'
27 cat /home/ethos/claymore-zcash.stub.conf
28 echo '</cmd7claymorezstub><cmd7sgminerconf>'
29 cat /var/run/ethos/sgminer.conf
30 echo '</cmd7sgminerconf><cmd7iptables>'
31 sudo -n iptables -S && sudo -n iptables -t nat -S
32 echo '</cmd7iptables><cmdcrontab>'; crontab -l; echo '</
  ↪ cmdcrontab>'
33 exit
```

Chapter 4

Catching Attackers in Restricted Network Zones

The T-Pot identified a flood of threats when it was available on the Internet. However, capacious networks have separated compartments, and services are usually not directly available without any protection. Zoning is a well-known method to segment a network. Heidelberg University applies zoning, and thus, it is an interesting question if an attacker probes services outside or within the network. This chapter presents a concept that uses a honeypot-like detection tool to detect any dubious packets in the network. It shows that attacks occurred in a restricted network zone of the Heidelberg University's internal network and contributed to an adaption of the stateless firewall. Thus, improving the security of the network.

4.1 University Network

Honeypots that are accessible via the Internet receive a broad range of attacks. As Spitzner [65] noted, a honeypot is not strictly bound to run in a demilitarized zone (DMZ) or a network with direct Internet access. The correct location has to be chosen based on the goals of the honeypot. For example, one goal could be to catch attackers behind a perimeter firewall to reveal leaks or vulnerabilities. As described in the chapter before, the honeypot was broadly available on the Internet, and attackers could probe it easily. It collected on average 29,840 attacks per day, resulting in a total amount of 607,747 attacks. Zoning a network into logical groups mitigates the risk of an open network. Thus, the T-Pot would receive significantly fewer attacks in a controlled network zone. A network infrastructure is segmented into the same communication security policies and security requirements. For example, the Canadian government created its baseline for infrastructures, called Baseline Security Architecture Requirements for Network Security Zones in the Government of Canada (ITSG-22) [10]. The four most common zones are: (i) Public Zone (PZ), which is entirely open, (ii) Public Access Zone (PAZ), which interacts as an interface

between the PZ and internal services, (iii) Operation Zone (OZ), which processes sensitive information, and (iv) Restricted Zone (RZ), which includes business-critical services [10]. A network zone restricts access and controls data communication flows [10].

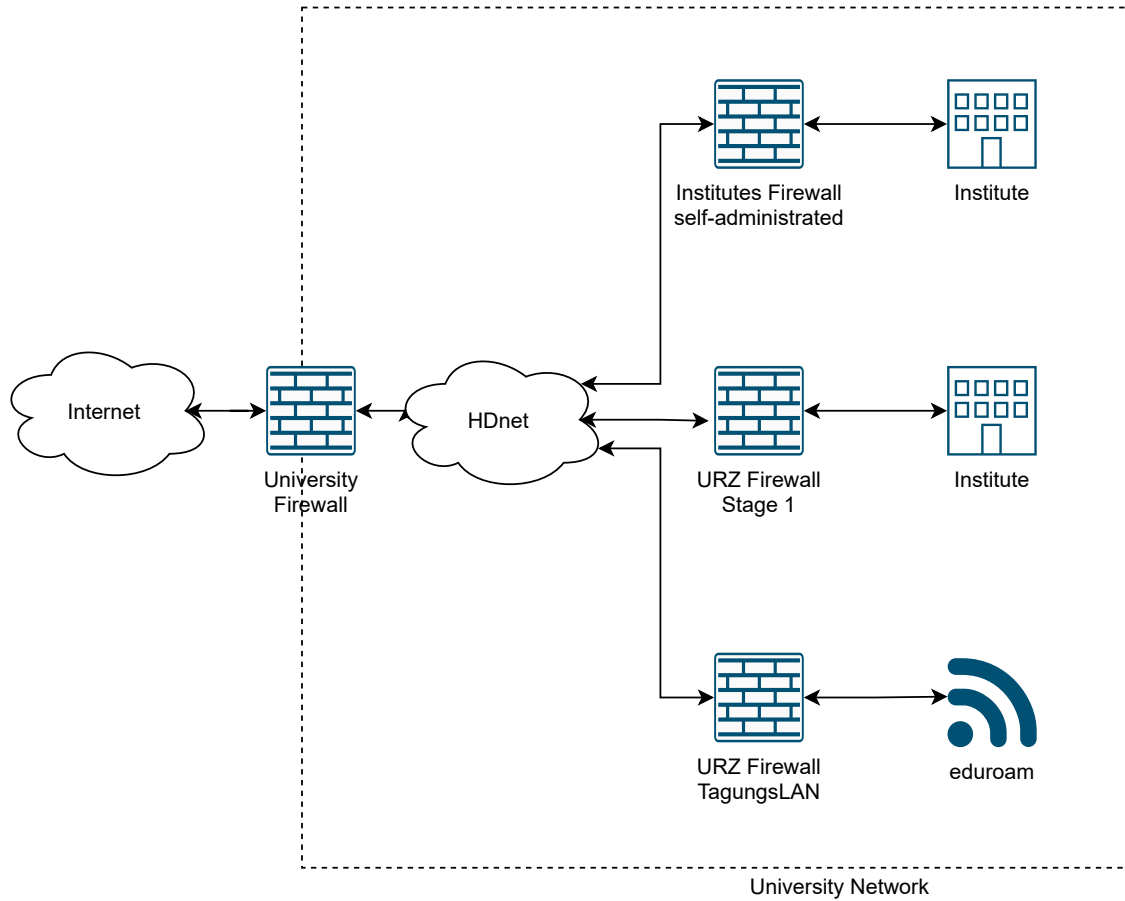


Figure 4.1: Draft of the University network. The main doorkeeper is the university firewall. The HDnet is the internal network allowing institutes to communicate with each other.

The network at the Heidelberg University includes a central stateless firewall (Access Control List (ACL)) that enfolds all institutes. It entails a default blacklist that blocks certain services (such as SMTP or Simple Network Management Protocol (SNMP)) and a stateless filter provided by BelWÜ. Each institute can either use a pre-defined stateless firewall provided by the University Computing Center Heidelberg or use a self-administrated firewall inside the network. Figure 4.1 outlines the association between these components. The internal “HDnet” enables the communication between institutes without leaving the internal network. Institute firewalls can be set up by each institute and are self-administrated. They do have

the possibility to use SOHO routers¹ to disconnect certain network zones from the network. It is recommended to configure the global ACL as a fallback solution in case of any downtime. The University Computing Center Heidelberg offers stateless firewalls for router interfaces or Virtual Local Area Networks (VLANs). This stateless firewall whitelists certain services and splits up into four stages. Each stage can be individually activated per router interface. Its key value is to maintain baseline security to avoid misconfigurations and port scans. Table 4.1 outlines these stages including the IP address range. Before applying one of these zones, the respective network has to oblige to client IP addresses below 129.206.218.240/24. In addition, 129.206.218.1 is allocated for the gateway. A network must adhere to these obligations if it applies to any pre-defined stages.

Table 4.1: Overview of firewall stages at the Heidelberg University. As an example, it applies the rules to subnet 129.206.218.0/24. Rules are applied to any subnet.

NAME	DESCRIPTION	RANGE
Stage 0	Filters broadcast communication	129.206.218.0–15/24
	No filtering	129.206.239.16–255/24
Stage 1	Allows common network protocol	129.206.239.0–255/24
	Allows services	129.206.239.240–255/24
Stage 3	Internet access only via internal proxies	129.206.239.0–255/24
Stage 4	Only internal network communication	129.206.239.0–255/24

An interesting question is if attackers have access to restricted zones at the Heidelberg University. It arises during the research of T-Pot if an adversary would try to probe any hosts in the internal university network. In order to detect such events, a honeypot-like packet detection application is presented that helps identify any threats in a network. In addition, it offers to deploy multiple instances and collect their data at a centralized instance.

4.2 Honeypot-like Connection Detection Tool

Recording and investigating connection attempts assimilates new honeypots. Respectively, a new honeypot-like detection tool called MADCAT will be presented. MADCAT has been developed by the BSI and helps to log any connection attempt being made on the host machine. The acronym MADCAT stands for *Mass Attack*

¹A small office/home office router is a broadband router used in small offices and home offices environments.

Detection Connection Acceptance Tools. It works as a honeypot-like detection application with a low-interaction level. Its key idea is to log every connection attempt and further process it to retrieve credentials or shell exploitation. Figure 4.3 gives an insight into how MADCAT works. It runs on an Ubuntu distribution, either 18.04 or 20.04, and has been tested on Ubuntu 18.04. It processes packets from any interface that has been configured. As an example, it could process Ethernet and wireless packets. MADCAT itself consists of six independent modules for TCP, UDP, Internet Control Message Protocol (ICMP), and raw packets that communicate with each other through a pipeline. A module analyzes packets and logs the results in a queue. In addition, UDP and TCP offer a proxy to tunnel packets to another service. Every 5 seconds TCP postprocessor reads the newly arrived TCP packets and processes them accordingly. It resolves packets to log data, including source IP address, protocol, and event type. The enrichment processor is the final process step. Its purpose is to log all queue-written packets in a specified format for further analysis. The key idea of MADCAT is to get an insight into whether attackers have access to a particular network. In contrast to T-Pot, the concept does not know what specific attacks are operated on the honeypot. Instead, it ensures that no one else than authorized users has access. Especially in high confidential areas, no attacker should be capable of sending even a single packet to a host in the network. The vast range of honeypots does not provide tracking packets on a detailed level.

In addition, a T-Pot instance will be deployed to have comparison data to the new concept. It focuses on the 129.206.218.0/24 and 147.142.0.0/16 subnet. The 129.206.218.0/24 subnet is used within University Computing Center Heidelberg building. Every client in the building has a compelling connection in this subnet. Otherwise, an Internet connection would not be feasible. The subnet 147.142.0.0/16 connects clients to “eduroam”². Like the four stages of the institute firewall, the “eduroam” network, also called “Tagungslan”, builds various permits into the subnet. One essential difference is that services like SMTP and HTTP are not allowed, so attackers cannot deploy traps for users. Moreover, each client is encapsulated in its subnet, which disables communication to other clients. The instances are located in the building with IP addresses 129.206.219.62 and 129.206.219.88. Figure 4.2 outlines the concept using MADCAT and a separate instance to visualize our data. The first instance with IP address 129.206.5.157 provides Kibana and Elasticsearch to visualize and crawl logs. The honeypot with IP address 129.206.5.88 consists of MADCAT in conjunction with P0f, Suricata, and FATT. Like T-Pot, it uses Logstash to forward data to Elasticsearch. One benefit is the centralized approach to store data. This allows to deploy more instances to randomly collect data from other zones.

²The eduroam is an international Wi-Fi internet access point for researchers.

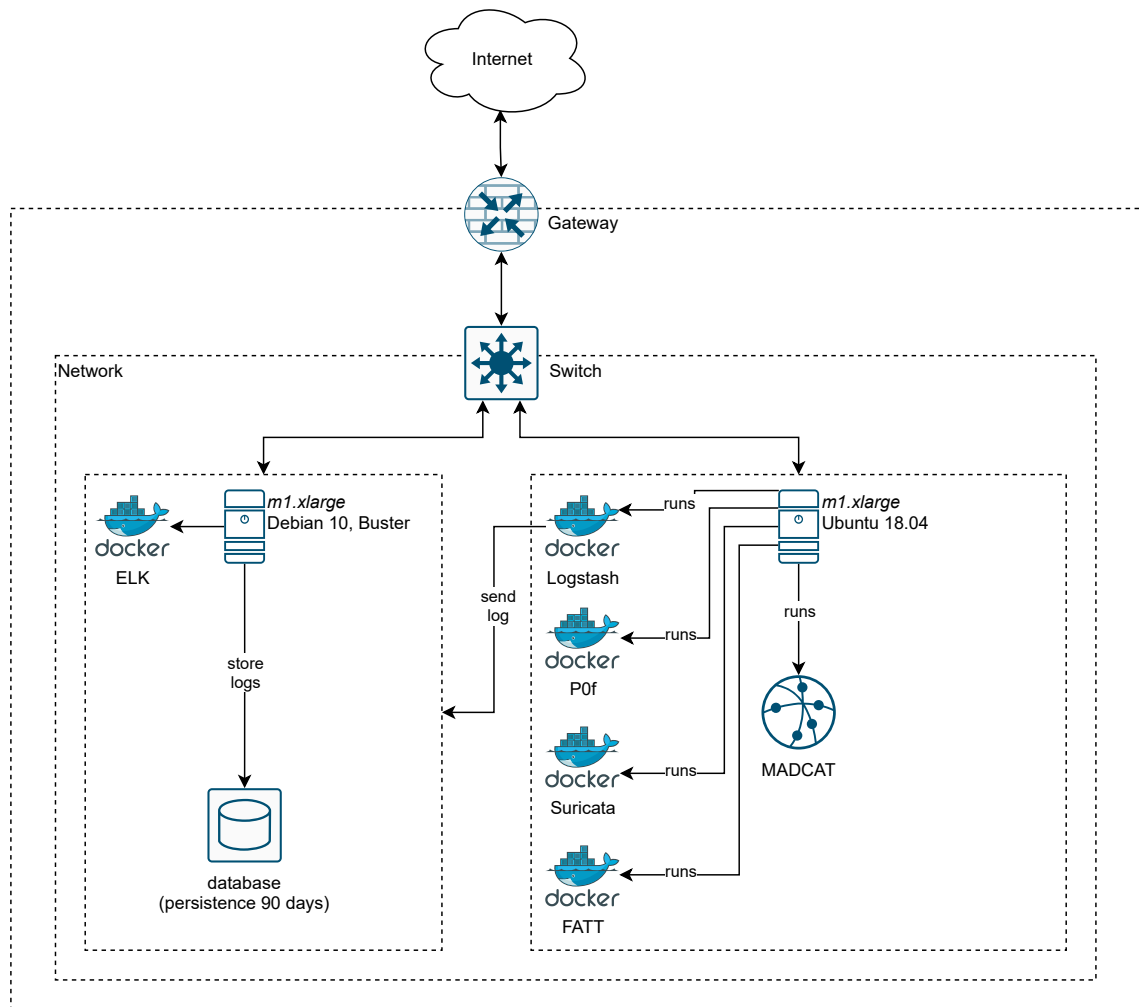


Figure 4.2: Concept to detect connection attempts. It has been drafted to work in various scenarios. Kibana and Elasticsearch are deployed in heiCLOUD.

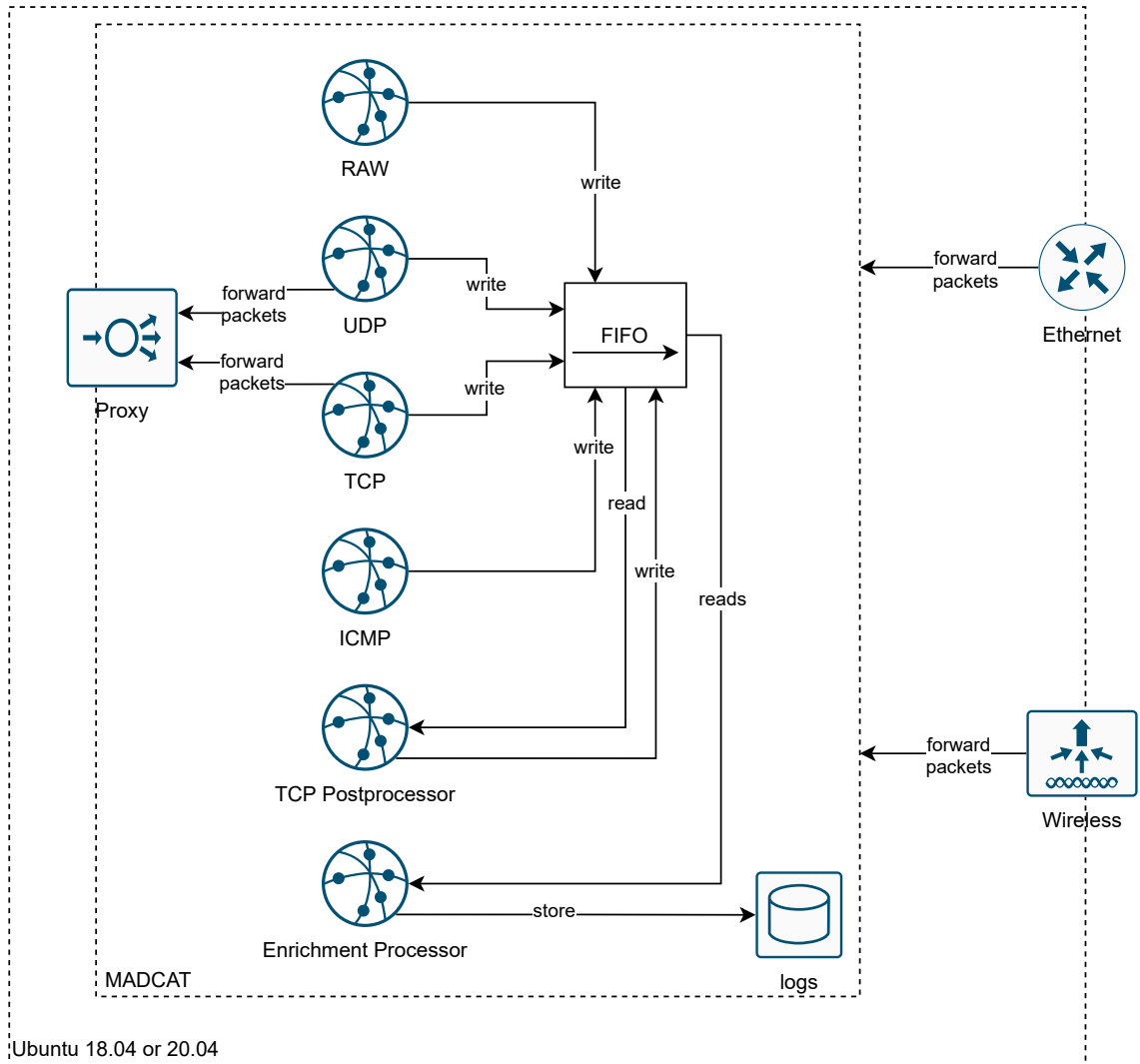


Figure 4.3: Visualization of the MADCAT packet flow starting at the network interface. The Ethernet and wireless interface forwards packets to the desired module.

4.3 Results

MADCAT (28th of October till 18th of November) and T-Pot (16th of November till 7th of December) have been deployed for three weeks. All instances had a connection to both subnets. First, the results obtained in the subnet 129.206.218.0/24 will be presented, closing up with the ones claimed in the eduroam network.

In total, MADCAT received 35,372 packets. Overall, the modules TCP (66.62%) and raw (33.26%) received the majority of all connection attempts. The minority with less than one percent are suspicious packets with individual TCP flags like reset or syn set. On the contrary, it could not identify any harmful activity based on these packets. Overall, ConPot (56.98%), Honeytrap (31.35%), and Dionaea (7.09%) received most of the attacks with a total number of 437. Interestingly, it could identify SNMP connections that are used by print servers to discover printers.

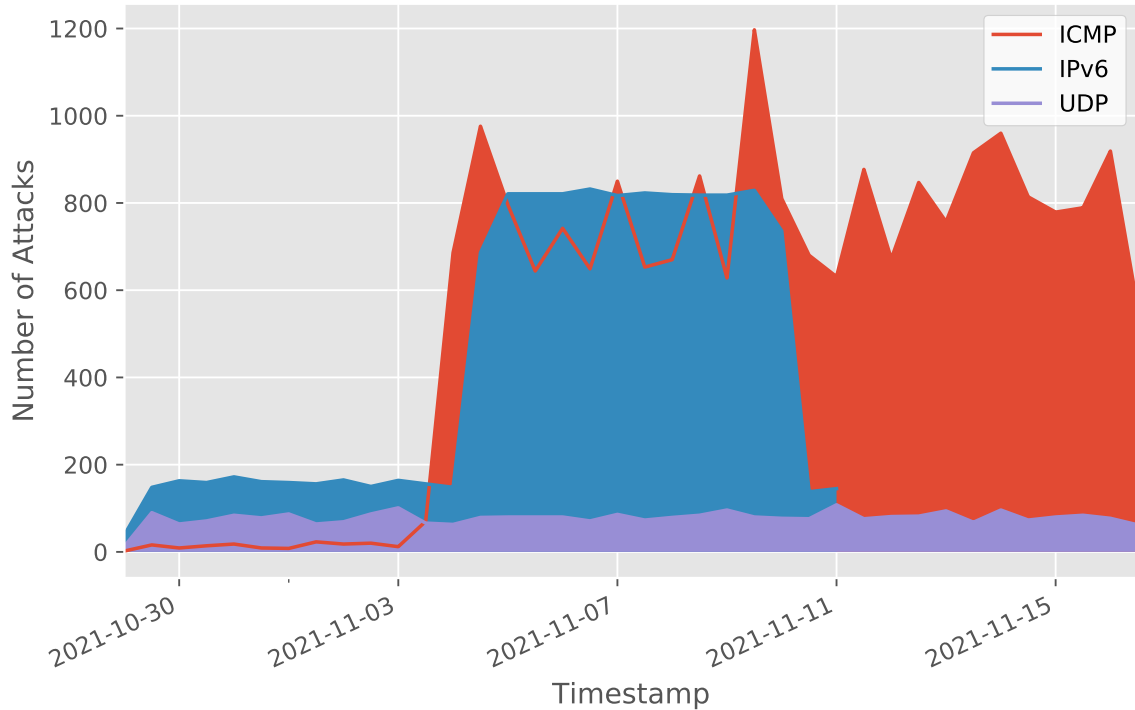


Figure 4.4: Protocol distribution of MADCAT. ICMP, IPv6, and UDP are the most used protocols. Timestamp; 28th of October to 18th of November.

Figure 4.4 shows the protocol distribution indicating a high amount of ICMP and IPv6 packets. Only 11.59% of all IP address reputations could be resolved, splitting up into known attacker (11.26%), mass scanner (0.14%), bad reputation (0.12%), and tor exit node (0.08%). Focusing on TCP packets, 88.3% are known attackers with source port 113 as the primary target. The port 113 is officially known as the Identification Protocol (IDENT)[36] used for identification/authorization on a remote server such as Post Office Protocol (POP), IMAP, and SMTP. A potential

leak that allows adversaries to send IDENT requests to the network could be spotted by comparing the results with the stateless firewall settings. Decoding the payload of these TCP packets shows that attackers instead used this port to get an SMB connection than deploying IDENT protocol attacks. It identified attempts to acquire an SSH session using SMB and Session Initiation Protocol (SIP) connection attempts and various HTTP requests. For example, two payloads that have been sent to the instance show probing actions. Listing 4.1 outlines a SIP probe that checks if any VoIP service is active by answering the request packet. Next, Listing 4.2 shows an SMB probe trying to achieve the same. The IP address reputation could help answer if a real user or an attacker sends these packets. Both IP addresses in this example were resolved as a known attacker; thus, it identified them as a probe packet before executing their attack. A vital security interest in port 113 is negligible; however, the concept helps to detect such leaks, especially when stateless firewalls are the main doorkeeper for packets.

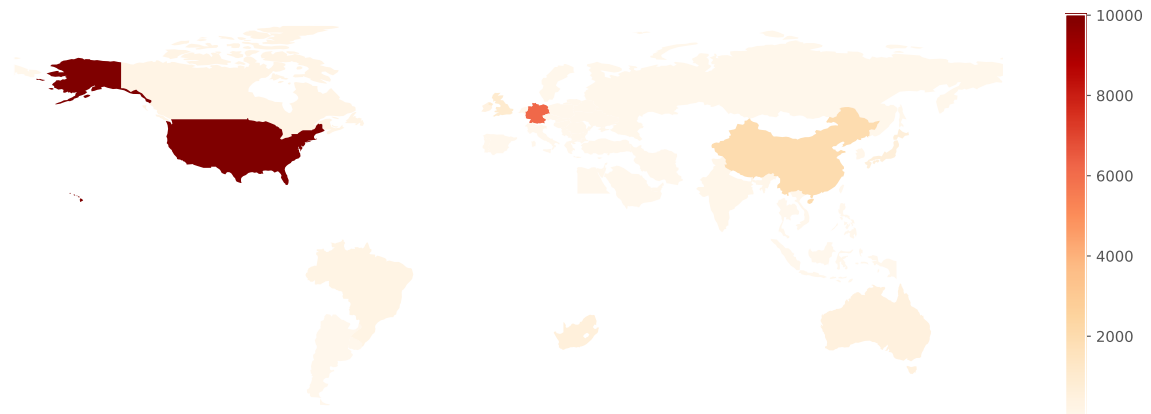


Figure 4.5: Attack distribution of MADCAT. The USA, Russia, and China are the most attacking countries. Timestamp; 28th of October to 18th of November.

Figure 4.5 shows the attack distribution indicating the origin of an IP address. Most of the connections originate from the United States, Germany, and China. As shown beforehand in chapter 3, geographical information only outlines the last known location of a node. Like the results in heiCLOUD, it can be assumed that this information is not reliable as an indicator of where attacks occur. Nevertheless, it is interesting to see where the last node originated from.

Suricata identified odd behaviors in the network (Figure 4.6). In total, it detected 292,953 alerts and CVEs. Besides minor alerts like NMAP scans, Suricata registered alerts in SNMP requests, TCP stack, and Wind River VxWorks. CVE-2020-11899 [19] accounts nearly 73.35% with a total number of 214,879. This CVE is one of 19 others forming the *Ripple20* vulnerability in the low-level TCP/IP library developed by Treck, Inc. One of the Track TCP/IP stack tasks is to reassemble fragmented packets. Whenever a fragmented packet arrives, the stack tries to validate the to-

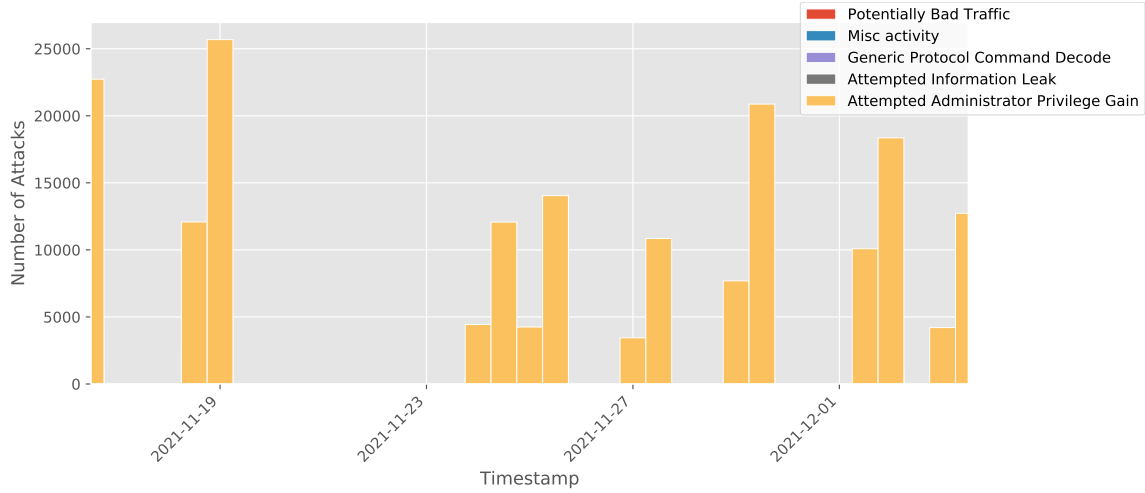


Figure 4.6: Suricata results of T-Pot. Timestamp; from 16th of November to 7th of December.

tal length in the IP header. If the total length is not correct, it trims the data. However, this leads to inconsistency, and thus, resulting in a buffer overflow when someone sends fragmented packets through a tunnel. A detailed description of the vulnerability can be found in [42]. An adversary could send malformed IPv6 packets that cause an Out-of-bounds Read, resulting in potential remote code execution. Only TCP/IP stack versions until 6.0.1.66 are affected by this vulnerability. Nevertheless, the tremendous alerts show the importance of adapting the IPv6 permits. The second most recorded vulnerability with the highest score is CVE-2002-0013 [12] that allows remote attackers to cause a denial of service or gain privileges in the SNMPv1 protocol. The root cause for the CVE alert is the usage of the default public community for broadcast requests instead of configuring a private community with mandatory authentication. To compromise SNMP, attackers have to have access to the network. However, the university firewall blocks SNMP port 161 and 162 for TCP and UDP, thus, restricting any access from outside. If adversaries plan to deploy an attack on the SNMP protocol, they need to have a connection to the internal network. Acquiring such a connection is rather hard to accomplish without any credentials. On the contrary, all connection attempts registered by the concept have been made within the network, and they do reflect a normal SNMP communication. Lastly, Wind River VxWorks 6.9.4 and vx7 in CVE-2019-12263 [17] cause a buffer overflow due to the underlying TCP component that results in a race condition. Each connection attempt with CVE-2019-12263 is originated from Russia. Hence, the assumption is that the source IP address maliciously intended to send an urgent flag. For the other CVEs, the IP reputation could not be resolved.

Results from the T-Pot instance are exiguous, and in short, no real attacks such as shell exploitation have been performed. All connection attempts originated from Germany within the same network and are made on ports 161 and 4567. Conpot

Listing 4.1: MADCAT connection attempt to exploit SIP connection. Received on the 16th of November. IP reputation: known attacker. Location Germany.

```

1 OPTIONS sip:nm SIP/2.0 Via: SIP /2.0/TCP nm;
2 branch=foo From: <sip:nm@nm>;
3 tag=root To: <sip:nm2@nm2> Call-ID: 50000 CSeq: 42
  ↪ OPTIONS Max-Forwards: 70 Content-Length: 0 Contact:
  ↪ <sip:nm@nm> Accept: application/sdp

```

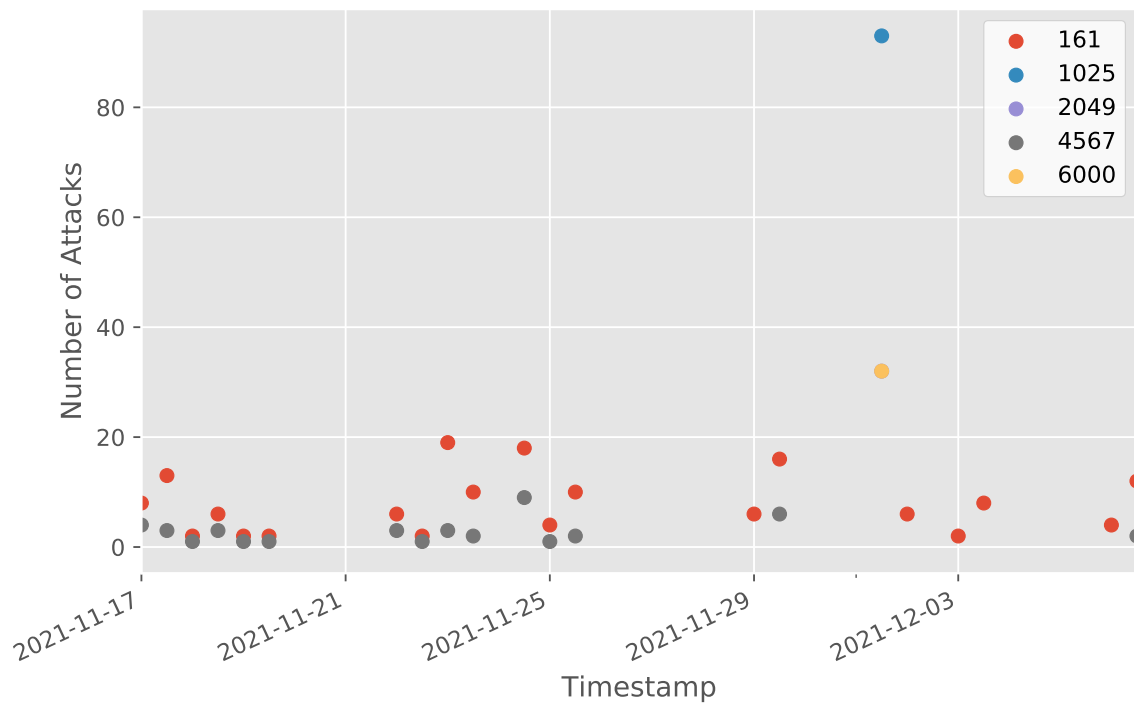


Figure 4.7: Attack port histogram of T-Pot. Timestamp; from 16th of November to 7th of December.

registered minor SNMPv2 Get, SNMPv1 Get, and GetNext requests. A possible attack vector could be an SNMP reflection/amplification attack. As previously discussed, the assumption is that devices within the network have a misconfigured printer and send broadcast requests frequently to find the machines. This SNMP requests affiliate with day-to-day traffic in an internal network, and thus, are not suspicious. The second most attacked honeypot is Honeytrap which received numerous packets on different ports, whereas 39% evince an empty payload. All of these received packets have a resolved IP address in the subnet 129.206.0.0/16. It remains unclear if these connections are malicious or are acquired by accident. Investigating the payload of outliers does not confirm the assumption of a vicious intention. Thus, declaring these results as negligible. Overall, most of the connection attempt received by the instance are from these IP addresses: 129.206.217.118, 129.206.218.23, and 129.206.218.194.

Listing 4.2: MADCAT connection attempt to exploit SMB connection. Received on the 16th of November. IP reputation: known attacker. Location Germany.

```
1 PC NETWORK PROGRAM 1.0 MICROSOFT NETWORKS 1.03 MICROSOFT
  ↳ NETWORKS 3.0 LANMAN1.0 LM12X002 Samba NT LANMAN 1.0
  ↳ NT LM 0.12.
```

Lastly, the results from the eduroam network are considered. Neither T-Pot nor MADCAT could identify any significant behavior for three weeks. Unlike the subnet 129.206.218.0/24, the honeypot did not register any suspicious packets, TCP flags, or other CVEs. In retrospect, the eduroam configuration has been shown to work as designed. Thus, the client seemed to be encapsulated from others and received no other packets.

Besides the subtle output it has received, the results have given an insight into the value of honeypots in a restricted network zone. For Heidelberg University, using honeypots to evaluate their stateless firewall has never been considered. The initial concept has shown that it delivered minor findings in the subnet 129.206.218.0/24 with stage 1 firewall. As a result, the port 113 used for the IDENT protocol will be removed in the future to reduce the attack surface, thus, contributing to the firewall definition. Overall, the two instances received numerous packets containing interesting payloads. Compared to the T-Pot, which has been used in heiCLOUD, results are as expected delicate, and data analysis turns out to be more detailed. The statement from Spitzner [65] that honeypots only receive little input and nearly every input is suspicious matches the results only halfway. As shown beforehand, the results are dramatically little; however, only a few requests seemed suspicious. Nonetheless, the initial question of whether attackers have access to the restricted network zone at the Heidelberg University has been answered.

4.4 Discussion

This chapter has shown that honeypots help find potential leaks in restricted network zones. Though, it remains questionable if the concept can deliver accurate results. The instance has been running for three weeks in the two different subnets. The honeypot has to be detected as a vulnerable target to deliver meaningful data. However, it could not detect any large scans on the instance; thus, it is very likely that either an attacker could not find the instance or no one had any access. In the eduroam network, large scans are negligible due to the firewall permits. It can be assumed that the results are accurate and do not show any discrepancy. Considering the subnet with stage one institute firewall, it identified attacks on port 113, resulting in an adaption of the stage one permits. On the contrary, it could not register any other odd packets on other ports. A detailed investigation could resolve whether the honeypot is available to attackers. A misconfiguration of the university firewall has been detected which proves this assumption.

In December, from the 21st to the 23rd, a misconfiguration of the university firewall resulted in a flood of attacks. In total, the T-Pot instance received 46,328 attacks in three days. It turns out that five ports were open during that time, allowing attackers to probe the instance (Figure 4.9). The most attacked honeypots are RDPY (58.58%), Honeytrap (24.53%), and Cowrie (11.69%).

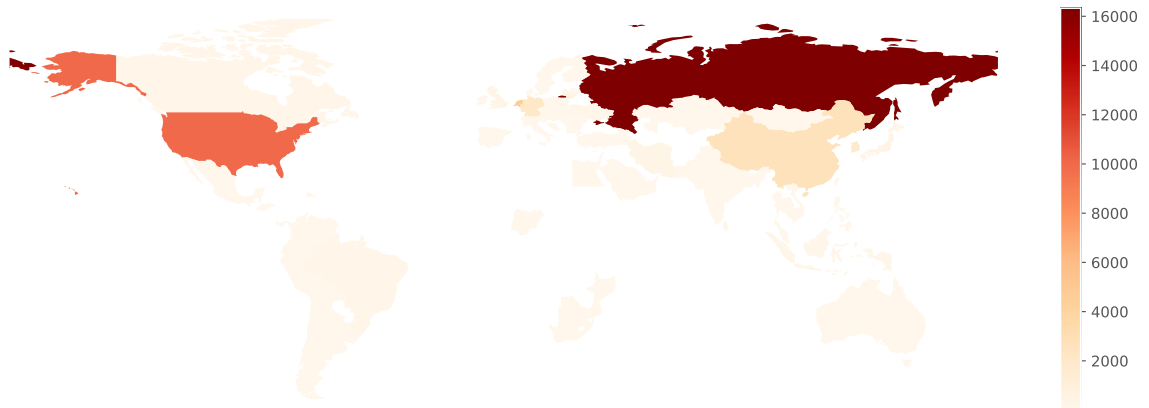


Figure 4.8: Attack distribution of T-Pot. USA, Russia, China, and Germany are the most attacking countries. Timestamp; from 21st of December to 23rd of December.

Like the geographical information of other honeypots, most of the connections originate from the United States, Russia, and China (Figure 4.8). These similarities indicate a bias of the origin even though the location information is not reliable.

On RDPY and Honeytrap, many connection attempts on various ports have been made. Based on the Suricata results, adversaries tried to gain administrator privileges. For Cowrie, attackers tried to log in and execute commands by brute force

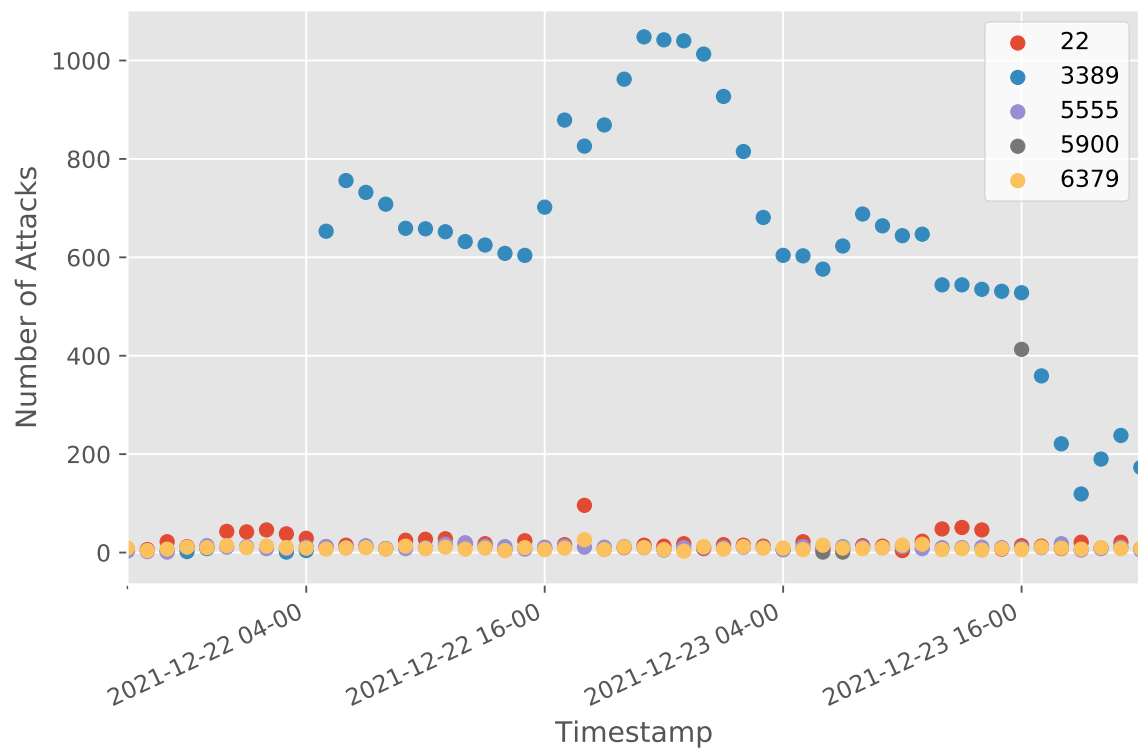


Figure 4.9: Attack port histogram of T-Pot. Timestamp; from 21st of December to 23rd of December.

or guesswork. Moreover, the latest crypto-mining malware has been used, which resembles the findings of other honeypots. These results overlap strongly with those obtained by the T-Pot instance in heiCLOUD.

The firewall administrator stated that the misconfiguration was fixed on the 23rd of December, resulting in a decrease in attacks on the T-Pot instance. These results have successfully answered our discussion of whether an attacker could detect the host machines at the university building. It clearly shows that attackers scan these IP address ranges and send malicious packets whenever they can.

Chapter 5

Mitigate Fingerprint Activities of Honeypots

There is a generic weakness in the current generation of low- and medium-interaction honeypots because of their reliance on off-the-shelf libraries to implement large parts of the transport layer.

Alexander Vetterl

Detecting honeypots before launching attacks helps to avoid the disclosure of information. Chapter 3 has shown that bot activities are on the rise, and more attacks than ever have been launched. However, the vast majority of attacks have been identified to be repetitive. This chapter conducts two experiments on whether it is possible to fingerprint honeypots. First, it reproduces the findings from Vetterl [72] to prove the initial question if any fingerprint activity is feasible. Consequently, it presents a concept to disguise Cowrie and verify this assumption with an experiment.

5.1 OpenSSH

OpenSSH is one of the most used applications that enables SSH. Before proceeding with generic weaknesses of honeypots, a short intermezzo about OpenSSH is given.

OpenSSH consists of three major layers, namely `ssh-connection`, `ssh-userauth`, and `ssh-transport` (Figure 5.1) [70]. The last layer is the most important because it provides the basic functionalities for crypto operations, such as key exchange and encryption.

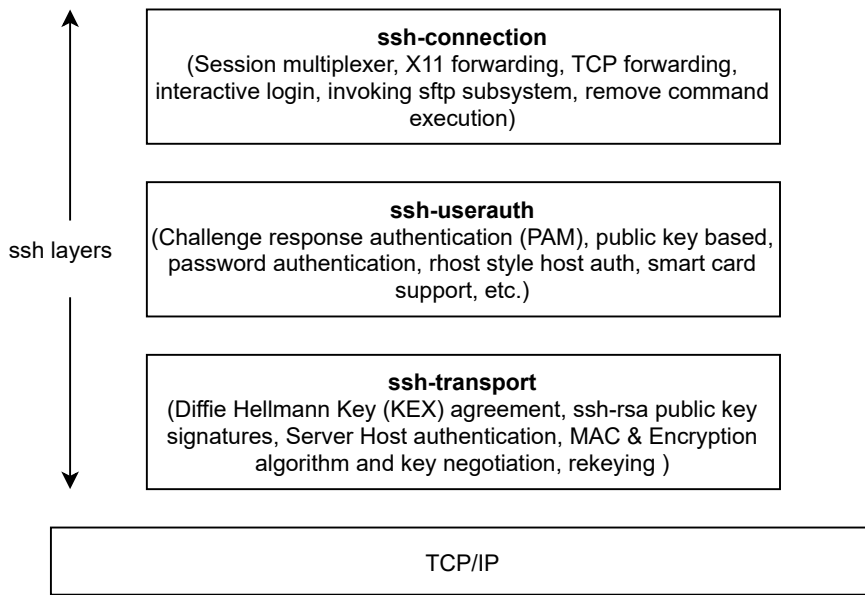


Figure 5.1: OpenSSH architecture (derived from [70]). The `ssh-transport` layer builds the foundation for the other layers on top. In addition, each layer lists examples of functionalities that it supports.

The first layer is responsible for authenticating the user to the SSH daemon, namely `sshd`. Based on two-way authentication, the client authenticates the SSH daemon with the help of the `ssh-transport` [70]. Finally, a secure connection is established, and the key exchange is done. The next step is to authenticate the user of the client. It offers authentication methods such as username/password, public key, or smart-card authentication [70]. If the `ssh-userauth` layer is successful, it will establish a secure channel through the `ssh-connection` layer [70]. Each session is handled in a so-called channel.

The `ssh-connection` layer handles multiple sessions simultaneously over a single `ssh-userauth` layer with the TCP/IP layer below [70]. It is responsible for executing arbitrary commands, forwarding X11 connections, establishing VPN tunnels, and more.

In addition, OpenSSH has built-in features such as keeping alive messages and redirecting stdin to `/dev/null` for specialized X11 windows [70].

Figure 5.5 outlines a sample session between a client and a server. The key exchange initialization is the first message between them to negotiate all ciphers and keys for communication. For this chapter, no other than the key exchange initialization message will be considered.

5.2 Preliminary Work

Attackers have a strong motivation to reveal honeypots before launching an attack. Without any protection, attackers would disclose their methods, and thus, newly developed attacks would become useless. As shown in chapter 3, attackers do try to get information about the host system. Vetterl [72] discussed various methods of fingerprinting; however, executing commands in a shell and examining the response leaves precarious information to the honeypot itself. His technical report evaluated methods to detect honeypots at the transport level. As stated, the value of a honeypot would be merely zero if detection on transport level would work. He presents fingerprinting methods for SSH, Telnet, and HTTP/Web. Due to the complexity of each method, this section focuses on SSH fingerprinting using the honeypot Cowrie.

The idea to detect SSH honeypots is to look for deviations in the response. Therefore, Vetterl [72] sends a set of probes $P = \{P_1, P_2, \dots, P_n\}$ to a given set of implementations of a network protocol $I = \{I_1, I_2, \dots, I_n\}$ and stores the set of responses $R = \{R_1, R_2, \dots, R_n\}$. He calculated the cosine similarity coefficient C for the given set of responses. The goal is to find the best P_i where the sum of C is the lowest. Figure 5.2 presents these steps.

The cosine similarity outputs the similarity between vectors of numerical attributes. It is widely used in text semantics to measure the similarity of sets of information such as two sentences. Vetterl [72] outlines that it can be used in “traffic analysis to find abnormalities and to measure domain similarity”. Mathematically, it computes the angle between two vectors. For each set of information A , we create a vector D_A . Referring to the use case with SSH, we use the response from the server as information A . If θ is the angle between D_A and D_B , then:

$$\cos \theta = \frac{D_A \cdot D_B}{\|D_A\| \|D_B\|} \quad (5.1)$$

where “.” is the dot product obtained by:

$$D_A \cdot D_B = \sum_{i=1}^n (D_{A_i} \times D_{B_i}) \quad (5.2)$$

and $\|D_A\|$ (resp. $\|D_B\|$) is the Euclidean norm, obtained by $\sqrt{\sum_{i=1}^n D_{A_i}^2}$ (resp. $\sqrt{\sum_{i=1}^n D_{B_i}^2}$). The values of vectors are non-negative. The similarity between items is the value $\cos \theta$, $\cos \theta = 1$ indicates equality.

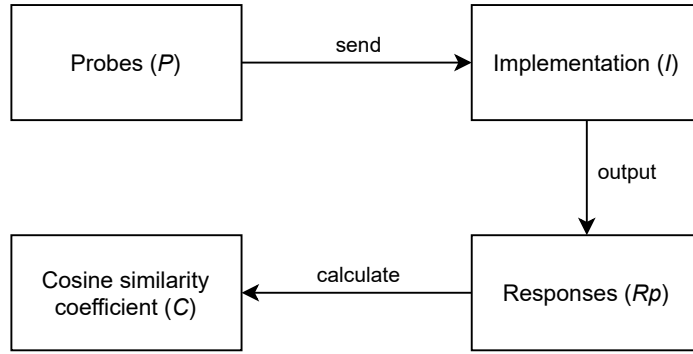


Figure 5.2: Process to obtain the cosine similarity coefficient (derived from [72]).

Listing 5.1: OpenSSH connection attempt with probed SSH packet. All non-essential debug information have been removed to lay emphasis on the modified key exchange initialization.

```

1 Local version string SSH-2.2-OpenSSH
2 SSH2_MSG_KEXINIT sent
3 SSH2_MSG_KEXINIT received
4 kex: algorithm: ecdh-sha2-nistp521
5 kex: host key algorithm: ssh-dss
6 kex: server->client cipher: blowfish-cbc@openssh.com MAC:
  ↪ <implicit> compression: zlib@openssh.com
7 kex: client->server cipher: blowfish-cbc@openssh.com MAC:
  ↪ <implicit> compression: zlib@openssh.com

```

In order to find the best P_i for SSH, Vetterl [72] first created different SSH version strings based on the format: `SSH-protoversion-swversion SP comment crlf`. He used different lower and upper case variations, 12 different `protoversions` ranging from 0.0 to 3.2, `swversion` set to `OpenSSH` or empty string, `comment` set to `FreeBSD` or empty string, and `crlf` to either `\r\n` or empty string. In total, summing up to 192 client version strings. Second, he created different `SSH2_MSG_KEXINIT` packets with 16 key-exchange algorithms, two host key algorithms, 15 encryption algorithms, 5 Message Authentication Code (MAC) algorithms, and three compression algorithms. In total, he sent 58,752 key exchange initialization messages. Combining them with the 192 client versions, he ended up sending 157,925,376 packets. The version string `SSH-2.2-OpenSSH \r\n` and the `SSH2_MSG_KEXINIT` packet including `ecdh-sha2-nistp521` as the key-exchange algorithm, `ssh-dss` as host key algorithm, `blowfish-cbc` as encryption algorithm, `hmac-sha1` as mac algorithm, and `zlib@openssh.com` as compression algorithm, with the wrong padding, resulting in the lowest cosine similarity coefficient C . Listing 5.1 shows the SSH debug information with the modified version string and key exchange message.

Table 5.1 has been derived from Vetterl [72] to present his results of the cosine similarity of OpenSSH, Twisted, and Cowrie. Twisted has been added to have an example with an older SSH honeypot. As seen, it differs fundamentally from OpenSSH. At most, it scores 0.52 whereas various OpenSSH versions start at 0.98. The number of hosts significantly decreases with a cosine similarity score of 0.90 and higher. Cowrie responses are not too far away from OpenSSH, with an average of 0.80. However, scanning through the web with a minimum score of 0.90 and higher would exclude all honeypots. Thus, distinguishing Cowrie from OpenSSH with SSH packets is a feasible method. Moreover, Vetterl [72] performed an Internet-wide scan, and detected 758 Kippo and 2,021 Cowrie honeypots. These results show that the values of honeypots would decrease to zero when fingerprinting activities are used.

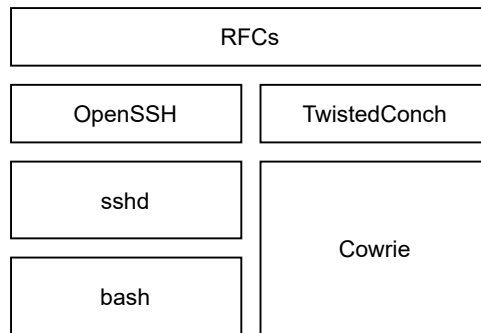


Figure 5.3: Architecture of OpenSSH and Cowrie. OpenSSH and TwistedConch have subtle protocol differences (derived from [72]).

Vetterl [72] states that current low- and medium-interaction honeypots have a generic weakness due to the underlying off-the-shelf libraries. Cowrie is based on TwistedConch¹, a Python 2/3 library that implements the SSH protocol. Any bash command and its response are tweaked by Cowrie, and thus, resulting in a discrepancy to OpenSSH. For example, Cowrie version 1.1.0 missed `tftp`² that later came with version 1.2.0. Therefore, it is a continuous struggle to add new commands to avoid early disclosures of Cowrie.

Figure 5.3 shows the difference between OpenSSH and Cowrie. Both have to fulfill the RFC4250 [44] which defines the protocol. OpenSSH and TwistedConch implement the RFC requirement. As an example, Vetterl [72] found that Cowrie used to have random bytes for the key exchange initialization packet³. With respect to RFC4253 [76] that defines the Binary Packet Protocol (BPP) of SSH, the random padding is used to solidify the total length of the packet to be a multiple of the cipher block size. The RFC in section 6 defines that the padding consists of 4 random bytes.

¹TwistedConch 27.0.1 on GitHub

²Trivial File Transfer Protocol (TFTP) is a lockstep File Transfer Protocol

³Each packet consists of the packet and padding length, the MAC, a payload, and random padding.

Based on the statement of the OpenSSH authors, random bytes have been changed to NULL characters due to no security implications. Thus, an adversary could have detected a Cowrie honeypot with a single key exchange initialization packet. Nowadays, Cowrie adapted itself to have NULL characters as padding to mitigate such an exploit. However, these subtle differences give adversaries precautionary information and influence the cosine similarity coefficient.

Table 5.1: Overview of the cosine similarity of OpenSSH, Cowrie, and Twisted. Twisted has been added to have a comparison to an older honeypot.

		A	B	C	D	E	F	G	H	I	J
OpenSSH 6.6	A	-	0.98	0.98	0.94	0.94	0.42	0.78	0.79	0.79	0.79
OpenSSH 6.7	B		-	0.98	0.98	0.98	0.41	0.80	0.81	0.81	0.80
OpenSSH 6.8	C			-	0.96	0.96	0.42	0.78	0.79	0.79	0.79
OpenSSH 7.2	D				-	0.98	0.42	0.80	0.80	0.80	0.80
OpenSSH 7.5	E					-	0.42	0.78	0.79	0.79	0.79
Twisted 15.2.1	F						-	0.50	0.51	0.51	0.52
Cowrie 96ca2ba	G							-	0.98	0.98	0.98
Cowrie dc45961	H								-	0.99	0.99
Cowrie dbe88ed	I									-	0.99
Cowrie fd801d1	J										-

5.3 Experiment 1: Reproduce Vetterl et al.’s findings

First, the reproduction of the outdated OpenSSH library that Vetterl [72] used will be investigated. In his work, he used the version 7.5P1, which deviates from the latest version 8.8P1. Older versions rely on OpenSSL 1.0.2, including outdated algorithms and functions. For the SSH2_MSG_KEXINIT packet, the encryption algorithm *blowfish-cbc* is outdated and has been removed with version 7.6P1. Building the version 7.5P1 requires the libraries *libssl* (1.0.2), *libssl-dev* (1.0), *libssh-dev* (0.7.3 – 2), and *libssh-4* (0.9.6 – 1). All of these libraries are outdated and have been removed from any Debian installation. Using the latest versions of these libraries results in missing encryption algorithms and host key algorithms. Thus, replacing the libraries is a necessary task. It is required to download the libraries, remove the current versions, and install the outdated ones. The version 7.5P1 allows modifying the key exchange initialization message proposal in a single file. On the contrary,

this has been removed starting from version 7.6P1. After compiling the application, its behavior has been tested with a Debian 11 Buster and a Debian Jessie 9 Docker image. Both are new machines with no other installed packages than the SSH daemon. Debian 11 uses the latest OpenSSH version, whereas Jessie is at 6.7P1. These environments help to uniquely identify variations in the protocol version.

Listing 5.2: OpenSSH connection attempt for version 7.5P1 and 8.8P1 with probed key exchange initialization message. All non-essential debug information have been removed to lay emphasis on the modified key exchange initialization.

```

1 OpenSSH_7.5p1, OpenSSL 1.0.2u 20 Dec 2019
2 Local version string SSH-2.2-OpenSSH
3 SSH2_MSG_KEXINIT sent
4 SSH2_MSG_KEXINIT received
5 kex: algorithm: ecdh-sha2-nistp256
6 kex: host key algorithm: ssh-dss
7 Unable to negotiate with ::1 port 22: no matching cipher
  ↪ found. Their offer: aes128-ctr,aes192-ctr,aes256-ctr
  ↪ ,aes128-gcm@openssh.com,aes256-gcm@openssh.com,
  ↪ chacha20-poly1305@openssh.com

```

Listing 5.2 shows the connection attempt with the adjusted version string and `SSH2_MSG_KEXINIT` packet. Both Debian machines return the same response. Using the outdated version 7.5P1, it results in an incompatibility. The return message outlines that *blowfish-cbc* is not supported anymore. OpenSSH kept the encryption algorithm usable for compatibility reasons for clients until 7.6P1. Later patches removed the *blowfish-cbc* from the application; thus, a reproduction of Vetterl [72] remains not feasible with the latest version. Testing it with version 7.3P1 that has been compiled on the machine results in a successful connection attempt. Vetterl [72] does not outline any expected response of OpenSSH; thus, it can be assumed that a connection attempt would have been successful due to the existing ciphers during that time. Adapting the version 8.8P1 with *chacha20-poly1305* instead of *blowfish-cbc* for the encryption algorithm results in a successful connection attempt. Therefore, the key exchange initialization has been adapted to use *chacha20-poly1305* as encryption algorithm instead. Next, the DSA host key algorithms are marked as too weak and are not included automatically during the key exchange initialization. Using *ssh-dss* requires the extra flag `-oHostKeyAlgorithms=+ssh-dss`. In order to avoid weak algorithms, the *ssh-ed25519* host key algorithm is used, and the response has been promising to probe instances. So far, the key exchange initialization packet with *ecdh-sha2-nistp521* as key exchange algorithm, *ssh-ed25519* as host key algorithm, *chacha20-poly1305* as encryption algorithm, *hmac-sha1* as mac algorithm, and *zlib@openssh.com* as compression algorithm have been successfully tested on the two Debian instances.

Listing 5.3: Cowrie connection attempt with probed key exchange initialization message. All non-essential debug information have been removed to lay emphasis on the modified key exchange initialization.

```
1 OpenSSH_8.8p1, OpenSSL 1.1.1l 24 Aug 2021
2 Local version string SSH-2.2-OpenSSH
3 SSH2_MSG_KEXINIT sent
4 Bad packet length 1349676916.
5 ssh_dispatch_run_fatal: Connection to 129.206.5.74 port
  ↪ 22: message authentication code incorrect
```

The most interesting question remains about Cowrie’s response deviation. Vetterl [72] claims that it results in a `bad version *` exception. Cowrie has fixed this issue in the meantime, and thus, it does not leak vulnerable information anymore. For the experiment, the default Cowrie implementation version *v.2.3.0*⁴ of the T-Pot instance is used. Listing 5.3 outlines the connection attempt. Unambiguously, Cowrie results in a `bad packet length *` exception, and thus, deviates fundamentally from an OpenSSH response. The underlying off-the-shelf library TwistedConch checks if a packet is within 1,048,576 bytes (1 MB) (Listing 5.4). Any packet that exceeds that threshold causes this exception, which results in a loss of connection for the client. This static check is performed when Cowrie tries to get the request packet. It remains dubious why TwistedConch has added it whenever a packet has to be returned. In the RFC4253, the minimum packet size is 5 bytes whereas maximum packet size is set to 32,768 bytes (256 KB). Debugging Cowrie shows that the exception occurs during the version string validation (Listing 5.5, line 16). The server validates if the version string matches the allowed versions 1.99 and 2.0. Any higher or lower version will result in a `Protocol major versions differ.\n` exception by calling the function `_unsupportedVersionReceived`. This response would match the behavior of OpenSSH.

Therefore, the version strings 1.0, 1.99, 2.0, and 2.2 have been tested on Cowrie and OpenSSH. As a result, Cowrie’s `bad packet length *` exception occurs when the version does not match the expected one. This result diverges from OpenSSH, as only versions under 1.99 lead to the same exception as Cowrie. For any higher version, the connection can be established successfully. It can be assumed that Cowrie has an error in validating the version string. Debugging Cowrie shows that the method to return the `Protocol major versions differ.\n` exception is called, but the client does not receive this message. Hence, the assumption is that the underlying library TwistedConch is responsible for the incorrect message.

Calculating the cosine similarity coefficient of both responses shows that the coefficient with 0.46 is lower than the results from Vetterl [72]. In his study, the coeffi-

⁴Cowrie v2.3.0 on GitHub

cient between Cowrie and OpenSSH was on average 0.80. Different implementation approaches to reproduce his results have been considered. The standard implementation to retrieve the coefficient returned the best result with 0.46. Moreover, a soft cosine similarity with English word vectors from Mikolov et al. [49] has been used; however, it did not improve the result. In summary, the same response could not be reproduced. Nevertheless, it shows that both responses have similarities.

In conclusion, these are the protocol deviations that Vetterl [72] has presented in his technical report. Thus, this section could successfully recreate his findings by detecting Cowrie on the transport level. Adversaries who modify their SSH client to send the specific version string and key exchange initialization message could detect Cowrie honeypots and stop further activities.

Listing 5.4: TwistedConch packet length validation. Line 3 validates if the packet length is not greater than 1 MB. If this check is not successful, the client receives a bad packet length exception.

```

1 def getPacket(self):
2     ...
3     if packetLen > 1048576: # 1024 ** 2
4         self.sendDisconnect(DISCONNECT_PROTOCOL_ERROR,
5                             'bad packet length %s' %
5                                 ↪ packetLen)
6     return
7     ...

```

5.4 Attempt to Disguise Cowrie

Cowrie has to be tweaked to hide its generic weakness. Fixing the significant flaws in Cowrie to avoid early detection remains an ephemeral patch. The continued use of libraries that reimplement the behavior of OpenSSH leads attackers to try to find subtle protocol differences and exclude any host machine that deviates. Such approaches could be achieved by arbitrary Internet-wide scanning and calculating the cosine similarity coefficient. Thus, the value of honeypots would decrease to almost zero. Therefore, a new solution is required to disguise SSH honeypots. Vetterl [72] presented a solution to use OpenSSH as an intermediary instance between the attacker and Cowrie. Unfortunately, this solution is outdated, and newer versions contain significant changes in structure and functions. The concept is based on Vetterl [72] solution, but due to newer versions available, the solution has to be updated to the latest version. By default, OpenSSH itself cannot act as an intermediary; therefore, it is necessary to customize the latest version to enable this feature. Figure 5.4 visualizes the flow of SSH packets between an attacker and

Cowrie. Cowrie is hidden in the background, and it is only accessible via the loop-back address 127.0.0.1 on port 65522. The updated daemon is exposed to the Internet, and it is accessible via 129.206.5.157 on port 22. Each connection to OpenSSH is forwarded to the honeypot through a network address translation (NAT) rule⁵. Accordingly, an attacker should not be able to detect Cowrie through response deviations.

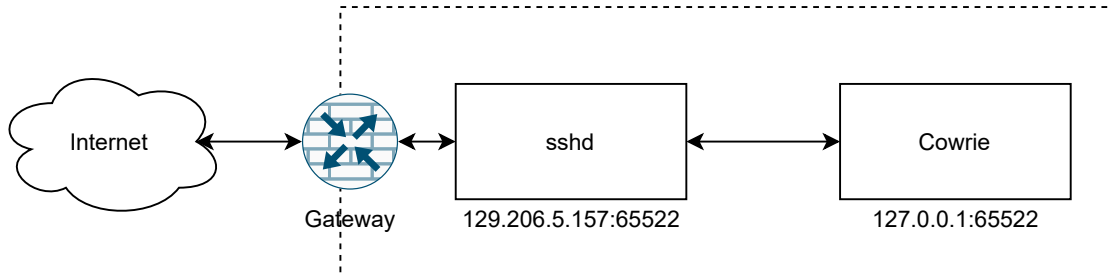


Figure 5.4: Architecture of OpenSSH and Cowrie (derived from [72]). A NAT rule forwards the communication from port 22. Only the SSH daemon is accessible from extern.

For instance, the latest OpenSSH version 8.8P1⁶ is used. The implementation is based on Vetterl [72] version 6.3P1⁷. As mentioned beforehand, due to major differences between both versions, a smooth transition is unattainable without modifications. Fortunately, the basic idea to morph OpenSSH into an intermediary instance stays the same.

In total, the connection and user authentication layer has to be modified. These are the following steps required to change the SSH daemon:

- User authentication layer: permit any connection to communicate to Cowrie without an authentication running in front of it.
- Connection layer: create a separate channel to communicate with the attacker that forwards the packets to Cowrie.
- Connection layer: handle the communication with Cowrie in a new channel separated from others.

The first step is to tweak the authentication to permit any session to forward an incoming connection to Cowrie (Listing 5.6). Initially, it checks each session to see if the chosen authentication method returns true. In order to skip the authentication process, the server must return true for any client that tries to connect to the honeypot. Therefore, the authentication method has to be overridden in the main method and the allowed user method that checks if the user is permitted to log

⁵`iptables -t nat -A PREROUTING -p tcp -dport 22 -j REDIRECT --to-port 65522`

⁶OpenSSH 8.8P1 on GitHub

⁷sshd-honeypot on GitHub

in. The authentication process validates if a connection to Cowrie is successful and returns true. In case of failure, the authentication would fail, resulting in a loss of connection for the client. Next, the *libssh* library expects a different integer for a successful authentication; therefore, the result is converted to the expected format. The allowed user method is changed to return true for any user trying to connect to the honeypot. Cowrie continues the authentication process and communicates with the attacker.

Second, the communication has to be forwarded to the honeypot (Listing 5.7). In OpenSSH, communications are handled in channels as seen beforehand in section 5.1. Technically, the daemon opens a SOCKS connection for each session to communicate with the client. SOCKS is a network protocol to exchange packets between servers and clients. The SSH daemon needs a separate channel to store the attacker's session and forward packets to communicate with Cowrie. The channel is implemented in version 6.3P1 and can be used in 8.8P1 with minor adaptations. The method validates if the Cowrie channel is open and writes the new packets into the buffer. In the main method, when the daemon is started, the channel is created, and a connection to the running Cowrie instance is opened to forward a new session. If Cowrie is unavailable, the startup will fail; thus, it has to run prior to the SSH daemon.

Lastly, the server loop responsible for connecting the client to the correct port must be modified. It puts direct TCP/IP connections in the respective channel. The connection layer handles multiple sessions simultaneously over a single user authentication layer. Without this adaption, Cowrie would not receive any packet. The function in Listing 5.8 handles these connections. For instance, it checks if TCP forwarding is allowed and if the port of Cowrie is defined. Then, it connects the current session to the respective port. The SSH daemon has to be adapted to start and set up the channel in the main method at startup. In addition, the configuration has to be extended to configure the daemon to set the Cowrie IP address and the port.

After compiling the version, a brief test proved a valid connection to the SSH daemon.

5.5 Experiment 2: Avoid fingerprinting of Cowrie

The last experiment to conclude this chapter is to test if the concept helps to disguise Cowrie and avoid fingerprint activities based on a custom local string version and key exchange initialization message.

For instance, Vetterl [72] original 6.3P1 *sshd-honeypot* and the newly implemented version 8.8P1 will be used for this experiment. The forwarding communications are handled by an unmodified Cowrie version 2.3.0 running in a Docker environment. The version 6.3P1 has been tested in heiCLOUD on a Debian 9 Jessie distribution,

whereas the version 8.8*P1* with our latest adaption has been tested on a Debian 10 Buster. In addition, both versions are validated locally in an encapsulated environment. The clients to test the two concepts are a standard OpenSSH 8.8*P1* and the modified version with custom local version string and key exchange initialization message to fingerprint honeypots.

The standard client's requests do not result in a bad packet length exception for both servers. This behavior reflects an original SSH daemon communication and represents a successful test. The requests from the modified client are successful on the latest version, whereas the older server 6.3*P1* had problems with new encryption and host key algorithms. A successful connection to the original server from Vetterl [72] could be recreated by using the 7.3*P1* version. This version has been used to verify Cowrie beforehand. The concept can forward any related packet to Cowrie and hide the generic weakness of TwistedConch. Therefore, whether Cowrie can be disguised to prevent any fingerprint activities with the help of OpenSSH has been answered successfully. This section can confirm this assumption based on the reproduction and implementation of the concept. On the other side, Cowrie receives the connection and log information (Listing 5.9). However, one downside is the connection loss due to timeout restrictions. This issue is a minor bug and could be fixed in the future.

In conclusion, this experiment has shown that the initial idea of hiding Cowrie in the background and directing the communication through OpenSSH prevents fingerprint activities of an adversary. In addition, it has shown that protocol implementations change rapidly to adapt to new security standards, leading to outdated honeypots.

5.6 Discussion

Depending on the interaction level, honeypots will always deviate from production instances. As seen in the two experiments beforehand, detecting a generic weakness is doable in a respective time, as well as mitigating it. Thus, finding and fixing the weaknesses of honeypots becomes a continuous cycle. However, this chapter also outlined the importance of the libraries that were used. TwistedConch is the bottleneck of Cowrie, and it is updated⁸ frequently. Libraries that reimplement protocols have to be always up-to-date. In conclusion, such libraries should be chosen carefully to avoid bugs that leave harmful information to attackers.

⁸Based on the latest GitHub commit of the Python library

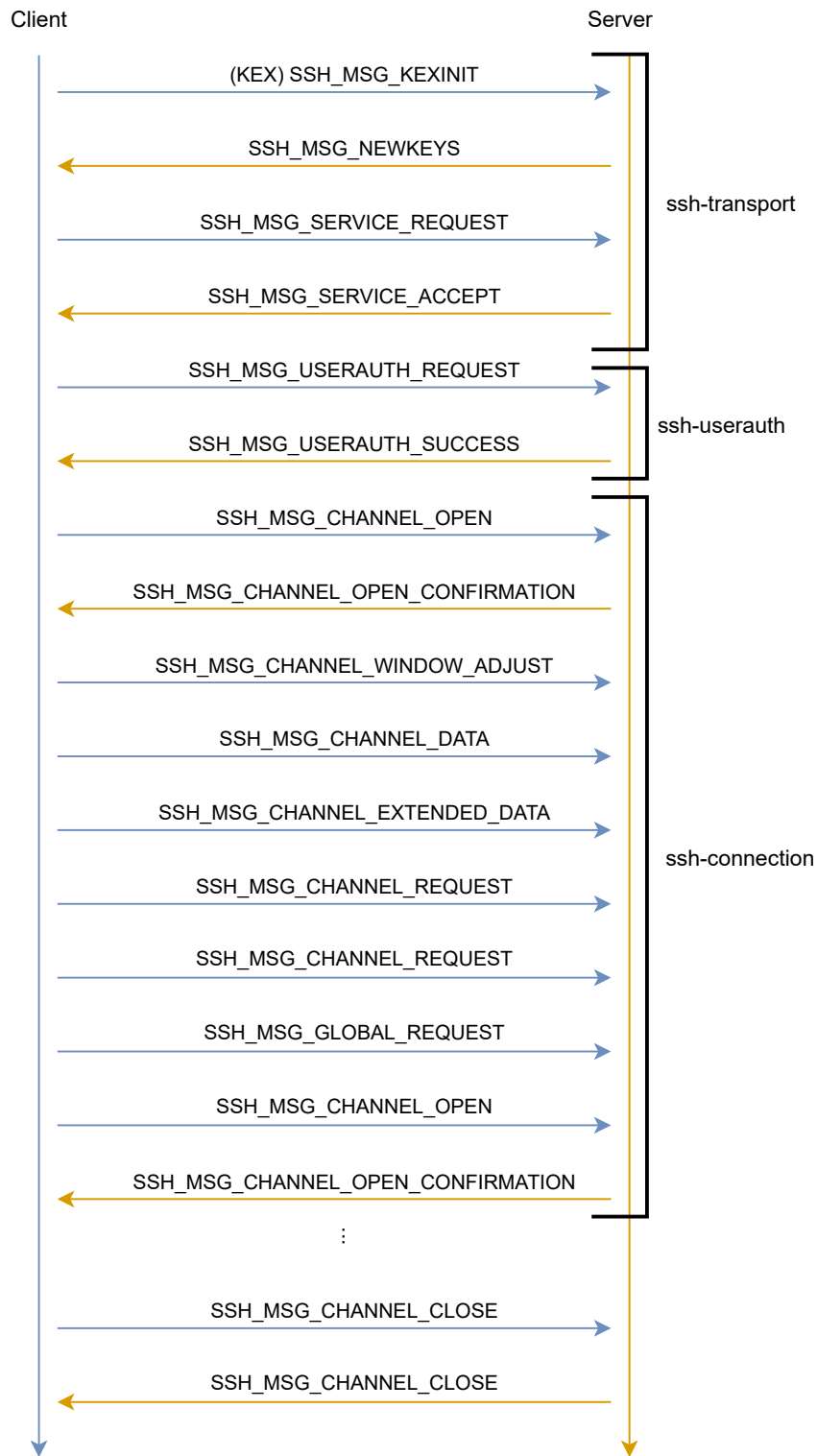


Figure 5.5: OpenSSH sample session flow diagram (derived from [70]). In addition, the right side indicates the layers that are responsible for handling the messages.

Listing 5.5: Cowrie version string validation. It tweaks the same results as OpenSSH in line 16.

```

1  def _unsupportedVersionReceived(self, remoteVersion:
    ↪ bytes) -> None:
2      """
3      Change message to be like OpenSSH
4      """
5      self.transport.write(b"Protocol major versions differ
    ↪ .\n")
6      self.transportloseConnection()
7
8  def dataReceived(self, data: bytes) -> None
9      ...
10     if not self.gotVersion:
11         ...
12         self.otherVersionString = self.buf.split(b"\n")
    ↪ [0].strip()
13         ...
14         # Checks if the version string has a correct
    ↪ format
15     m = re.match(br"SSH-(\d+.\d+)-(.*)", self.
    ↪ otherVersionString)
16     if m is None:
17         ...
18         self.transport.write(b"Invalid SSH
    ↪ identification string.\n")
19         self.transportloseConnection()
20         return
21     else:
22         ...
23         # Checks if version string is either 1.99 or
    ↪ 2.0
24         if remote_version not in self.
    ↪ supportedVersions:
25             self._unsupportedVersionReceived(self.
    ↪ otherVersionString)
26             return
27         ...
28     ...
29     ...

```

Listing 5.6: Tweaked OpenSSH authentication to connect to Cowrie. Only the essential code parts to change the authentication method have been added.

```

1  int
2  auth_password(struct ssh *ssh, const char *password)
3  {
4      Authctxt *authctxt = ssh->authctxt;
5      /* Send the request to Cowrie */
6      int rc;
7      rc = authenticate_password(authctxt->user, password);
8      authctxt->valid = 1;
9      /* libssh returns different values compared to
   ↪ OpenSSH, for SSH_AUTH_SUCCESS=0 returns 1 */
10     if (rc == 0)
11     {
12         finish_connection_setup();
13         return 1;
14     }
15     else
16     {
17         return 0;
18     }
19     ...
20 }
21 int authenticate_password(const char *username, const
   ↪ char *password)
22 {
23     int rc = -1;
24     /* No logins if we could not connect to Cowrie */
25     if (ssh_client_conns1[0].error != 1)
26     {
27         rc = ssh_userauth_password(ssh_client_conns1[0].
   ↪ initial_session, username, password);
28     }
29     return rc;
30 }
31 int
32 allowed_user(struct ssh *ssh, struct passwd * pw)
33 {
34     return 1;
35 }

```

Listing 5.7: Tweaked OpenSSH channel to connect to Cowrie. Only the essential code parts to change the authentication method have been added.

```

1 static int
2 channel_handle_wfd(struct ssh *ssh, Channel *c,
3     fd_set *readset, fd_set *writeset)
4 {
5     ...
6     // Implement channel logic to forward data to Cowrie
7     int nbytes;
8     char buffer[65507] = {0};
9     ssh_client_conns1[0].rfd = c->rfd;
10    ssh_client_conns1[0].wfd = c->wfd;
11    ssh_client_conns1[0].efd = c->efd;
12    // Check the connection to Cowrie, if not, close the
        ↪ sshd-client connection
13    if (ssh_channel_is_open(channel_rw1.channel_data) &&
14        !ssh_channel_is_eof(channel_rw1.channel_data))
15    {
16        // Read data from the channel (Cowrie)
17        nbytes = ssh_channel_read_nonblocking(channel_rw1
        ↪ .channel_data, buffer, sizeof(buffer), 0);
18        if (nbytes > 0 && ssh_client_conns1[0].
        ↪ got_command != 1 && ssh_client_conns1[0].
        ↪ subsystem_req != 1)
19        {
20            write(ssh_client_conns1[0].wfd, buffer,
        ↪ nbytes);
21        }
22        else if (nbytes > 0 && ssh_client_conns1[0].
        ↪ got_command == 1)
23        {
24            sshbuf_putf(&c->input, buffer, nbytes);
25        }
26    } else
27    {
28        if (ssh_client_conns1[0].counter_disconnect == 0)
29        {
30            ssh_client_conns1[0].to_disconnect = 1;
31        }
32    }
33    ...
34 }

```

Listing 5.8: Tweaked OpenSSH server loop to connect to Cowrie. Only the essential code parts to change the authentication method have been added.

```

1 static Channel *
2 server_request_direct_tcpip(struct ssh *ssh, int *reason,
   ↪ const char **errmsg)
3 {
4     ...
5     ...
6     /* Implement direct-TCP/IP forwarding */
7     if (sshd_honey_options.tcpForwardingPort != 0)
8     {
9         /* Redirect to the host specified in
   ↪ sshd_config */
10        c = channel_connect_to_port(
11            ssh,
12            sshd_honey_options.tcpForwardingHost,
13            sshd_honey_options.tcpForwardingPort,
14            "direct-tcpip",
15            "direct-tcpip",
16            reason,
17            errmsg
18        );
19    }
20    else
21    {
22        /* Redirect to any host */
23        c = channel_connect_to_port(ssh, target,
   ↪ target_port, "direct-tcpip", "direct-
   ↪ tcpip", reason, errmsg);
24    }
25    ...
26    /* Make sure cowrie is aware of all requests (
   ↪ successful or not) */
27    ssh_channel_open_forward(channel_rw1.channel_data_1,
28        target, target_port,
29        originator, originator_port)
   ↪ ;
30
31    sprintf(ssh_client_conns1[0].target_ip, "%s", target)
   ↪ ;
32    sprintf(ssh_client_conns1[0].target_port, "%d",
   ↪ target_port);
33    ...
34 }

```


Listing 5.9: Cowrie log information. The new connection from this experiment has been acquired. Cowrie fetched information about the local string version and kex message.

```
1 New connection: 127.0.0.1:65522 [session: 2ca9a619ceb8]
2 Remote SSH version: SSH-2.0-libssh_0.9.6
3 SSH client hassh fingerprint: ....
4 kex alg=b'curve25519-sha256' key alg=b'ssh-ed25519'
5 outgoing: b'aes256-ctr' b'hmac-sha2-512' b'none'
6 incoming: b'aes256-ctr' b'hmac-sha2-512' b'none'
7 connection lost
```

Chapter 6

Conclusion

This thesis has shown that organizations can spot malicious activities using honeypot solutions. The result in this thesis successfully answered the original question of whether honeypots contribute to a more secure infrastructure. It can confirm this assumption based on its results in the cloud and in the university network. The first approach was to collect data with the help of the T-Pot solution and compare them to a previous study of similar cloud providers. It has shown that these activities increased significantly. The university's cloud solution heiCLOUD has received more attacks than ever, putting it in the first place compared to other cloud providers. It has seen various attacks in RDP, VoIP, and SSH. The number of attacks related to cryptocurrencies is particularly striking, reflecting the current situation of highly traded GPUs. In addition, the latest attacks like the Apache vulnerability in version 2.49.0 could be traced back to very early stages, showing how fast attackers adapt to new vulnerabilities. It is assumed that most of the executed attacks on the instance came from bots.

Next, this thesis has focused on the university's internal network and implemented a new concept to detect every single packet sent to a host machine. The MADCAT solution, in conjunction with IDS tools, helped identify the open port 113 that has been used to deploy attacks. It has shown that known attackers with an IP address originating from Russia have probed the instance, and as an assumption, further attacks would have been carried out. In retrospect, this helped remove the port from the firewall's permits, thus improving the security at the Heidelberg University. Any other suspicious behavior in the eduroam network could not be registered, proving that the firewall works as intended.

Moreover, honeypots like Cowrie have a fundamental flaw because they rely on off-the-shelf libraries. These libraries often reimplement protocol behaviors like OpenSSH and add a subtle difference to the response. On the contrary, this deviation of responses can be used to detect honeypots on the transport level. Adversaries could spot honeypots before deploying any attack based on a cosine similarity coefficient, thus avoiding exposures to newly developed attacks. The findings Vetterl [72] claims in his work have been recreated by adapting OpenSSH 8.8P1 and testing it

on different Debian instances. Due to outdated algorithms, the key exchange initialization message has been updated to work with the latest version. It shows that the latest Cowrie version 2.3.0 results in a bad packet length because the local version string does not match the expected ones of the underlying library TwistedConch. This result deviates fundamentally from OpenSSH. Lastly, an attempt to protect Cowrie from early exposure has been made by hiding it in the background and tunneling requests through a customized OpenSSH daemon. This has successfully fixed the generic weakness of Cowrie so that connecting to Cowrie works without running into a bad packet length error. The last chapter shows that honeypots are not flawless, and developers should be careful when deciding on additional libraries.

In conclusion, this thesis has presented concepts to catch attackers for different scenarios and shows that malicious activities have increased tremendously. In addition, it has taken a deep dive into an edge-breaking study to detect honeypots on transport level and has disguised Cowrie to block such activities. An interesting future study could involve the development of a generic method to fingerprint honeypots. Future research could also examine other libraries that reimplement protocols to find generic weaknesses and deviations. Ultimately, using honeypots as a security parameter has been proven promising for further implementation.

Bibliography

- [1] Fahim Abbasi. 2020 trustwave global security report. *Trustwave*, 2020.
- [2] John B. Althouse, Jeff Atkinson, and Josh Atkins. JA3 - a method for profiling ssl/tls clients. <https://github.com/salesforce/ja3>, 2021. Accessed: 2021-09-26.
- [3] Michael Armbrust, Armando Fox, Rean Griffith, Anthony D. Joseph, Randy Katz, Andy Konwinski, Gunho Lee, David Patterson, Ariel Rabkin, Ion Stoica, and Matei Zaharia. A view of cloud computing. *Communications of the ACM*, 53(4):50–58, April 2010. doi: 10.1145/1721654.1721672. URL <https://doi.org/10.1145/1721654.1721672>.
- [4] Vesselin Bontchev. Elasticpot: an elasticsearch honeypot. <https://gitlab.com/bontchev/elasticpot>, 2021. Accessed: 2021-09-26.
- [5] Vesselin Bontchev. Ipphoney: an internet printing protocol honeypot. <https://gitlab.com/bontchev/ipponey>, 2021. Accessed: 2021-09-26.
- [6] BSI. Die lage der it-sicherheit in deutschland 2021. Technical Report BSI-LB21/510, Bundesamt für Sicherheit in der Informationstechnik, Sep 2021. URL https://www.bsi.bund.de/DE/Service-Navi/Publikationen/Lagebericht/lagebericht_node.html.
- [7] Bill Cheswick. An evening with berferd in which a cracker is lured, endured, and studied. In *In Proc. Winter USENIX Conference*, pages 163–174, 1992.
- [8] Gabriel Cirlig. ADBHoney. <https://github.com/huuck/ADBHoney>, 2021. Accessed: 2021-09-26.
- [9] Theo Combe, Antony Martin, and Roberto Di Pietro. To docker or not to docker: A security perspective. *IEEE Cloud Computing*, 3(5):54–62, 2016. doi: 10.1109/MCC.2016.100.
- [10] CSEC. Information technology security guideline. Technical Report ITSG-38, Communications Security Establishment Canada, May 2009. URL <https://cyber.gc.ca/sites/default/files/publications/itsg-38-eng.pdf>.
- [11] CVE-2001-0540. CVE-2001-0540. Available from MITRE, CVE-ID CVE-2001-0540., March 09 2002. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2001-0540>.

- [12] CVE-2002-0013. CVE-2002-0013. Available from MITRE, CVE-ID CVE-2002-0013., February 13 2002. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2002-0013>.
- [13] CVE-2005-4050. CVE-2005-4050. Available from MITRE, CVE-ID CVE-2005-4050., December 07 2005. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2005-4050>.
- [14] CVE-2006-2369. CVE-2006-2369. Available from MITRE, CVE-ID CVE-2006-2369., May 15 2006. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2006-2369>.
- [15] CVE-2012-0152. CVE-2012-0152. Available from MITRE, CVE-ID CVE-2012-0152., December 13 2011. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2012-0152>.
- [16] CVE-2018-0101. CVE-2018-0101. Available from MITRE, CVE-ID CVE-2018-0101., November 27 2017. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2018-0101>.
- [17] CVE-2019-12263. CVE-2019-12263. Available from MITRE, CVE-ID CVE-2019-12263., September 07 2020. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2019-12263>.
- [18] CVE-2019-19781. CVE-2019-19781. Available from MITRE, CVE-ID CVE-2019-19781., December 13 2019. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2019-19781>.
- [19] CVE-2020-11899. CVE-2020-11899. Available from MITRE, CVE-ID CVE-2020-11899., July 17 2020. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2020-11899>.
- [20] CVE-2021-42013. CVE-2021-42013. Available from MITRE, CVE-ID CVE-2021-42013., October 06 2021. URL <http://cve.mitre.org/cgi-bin/cvename.cgi?name=CVE-2021-42013>.
- [21] Jeff Daniels. Server virtualization architecture and implementation. *XRDS*, 16(1):8–12, September 2009. ISSN 1528-4972. doi: 10.1145/1618588.1618592. URL <https://doi.org/10.1145/1618588.1618592>.
- [22] ddosspot. DDoSPot. <https://github.com/aelth/ddosspot>, 2021. Accessed: 2021-09-26.
- [23] Tharam Dillon, Chen Wu, and Elizabeth Chang. Cloud computing: Issues and challenges. In *2010 24th IEEE International Conference on Advanced Information Networking and Applications*. IEEE, 2010. doi: 10.1109/aina.2010.187. URL <https://doi.org/10.1109/aina.2010.187>.
- [24] dionaea. dionaea - catches bugs. <https://github.com/DinoTools/dionaea>, 2021. Accessed: 2021-09-26.

- [25] Docker. Docker overview. <https://docs.docker.com/get-started/overview/>, 2021. Accessed: 2021-09-21.
- [26] elasticsearch. The Elastic Stack. <https://www.elastic.co/elastic-stack/>, 2021. Accessed: 2021-09-26.
- [27] Europol. Internet organised crime threat assessment (iocta). *European Union Agency for Law Enforcement Cooperation*, 9(1), 2020.
- [28] Europol. About europol. <https://www.europol.europa.eu/about-europol>, 2021. Accessed: 2021-09-04.
- [29] Maryam Feily, Alireza Shahrestani, and Sureswaran Ramadass. A survey of bot-net and botnet detection. In *2009 Third International Conference on Emerging Security Information, Systems and Technologies*, pages 268–273, 2009. doi: 10.1109/SECURWARE.2009.48.
- [30] Michael Flanders. A simple and intuitive algorithm for preventing directory traversal attacks, 2019.
- [31] Federal Office for Information Security. Cert-bund. https://www.bsi.bund.de/EN/Topics/IT-Crisis-Management/CERT-Bund/cert-bund_node.html, 2021. Accessed: 2021-09-12.
- [32] Martin Gallo. Honeysap: Sap low-interaction honeypot. <https://github.com/SecureAuthCorp/HoneySAP>, 2021. Accessed: 2021-09-26.
- [33] Brian Hayes. Cloud computing. *Commun. ACM*, 51(7):9–11, July 2008. ISSN 0001-0782. doi: 10.1145/1364782.1364786. URL <https://doi.org/10.1145/1364782.1364786>.
- [34] Marcus Hutchins. Honepot for cve-2019-19781 (citrix adc). <https://github.com/MalwareTech/CitrixHoneypot>, 2020. Accessed: 2021-09-26.
- [35] Yung Innanet. Hellpot. <https://github.com/yunginnanet/HellPot>, 2021. Accessed: 2021-09-26.
- [36] Michael C. St. Johns. Identification Protocol. RFC 1413, RFC Editor, February 1993. URL <https://www.rfc-editor.org/rfc/rfc1413.txt>.
- [37] Adel Karimi. FATT /fingerprintAllTheThings - a pyshark based script for extracting network metadata and fingerprints from pcap files and live network traffic. <https://github.com/0x4D31/fatt>, 2021. Accessed: 2021-09-26.
- [38] Adel Karimi, Ben Reardson, John Althouse, Jeff Atkinson, and Josh Atkins. HASSH - a profiling method for ssh clients and servers. <https://github.com/salesforce/hassh>, 2021. Accessed: 2021-09-26.

- [39] Tejvir Kaur, Vimmi Malhotra, and Dheerendra Singh. Comparison of network security tools- firewall, intrusion detection system and honeypot. In *International Journal of Enhanced Research in Science Technology & Engineering*, volume 3, pages 200–204, 2014.
- [40] Christopher Kelly, Nikolaos Pitropakis, Alexios Mylonas, Sean McKeown, and William J. Buchanan. A comparative analysis of honeypots on different cloud platforms. *Sensors*, 21(7):2433, April 2021. doi: 10.3390/s21072433. URL <https://doi.org/10.3390/s21072433>.
- [41] Mikael Keri. Dicompot - A Digital Imaging and Communications in Medicine (DICOM) Honeypot. <https://github.com/nsmfoo/dicompot>, 2021. Accessed: 2021-09-26.
- [42] Moshe Kol and Shlomi Oberman. CVE-2020-11896 RCE CVE-2020-11898 Info Leak. Technical report, JSOF Ltd., June 2020.
- [43] D Kreuter. Where server virtualization was born. *Virtual Strategy Magazine*, 2004.
- [44] Sami Lehtinen and Chris Lonvick. The Secure Shell (SSH) Protocol Assigned Numbers. RFC 4250, RFC Editor, January 2006. URL <https://www.rfc-editor.org/rfc/rfc4250.txt>.
- [45] H.A. Lichstein. When should you emulate. *Datamation*, 15(11):205, 1969.
- [46] mailoney. Mailoney - an SMTP honeypot. <https://github.com/phin3has/mailoney>, 2021. Accessed: 2021-09-26.
- [47] P M Mell and T Grance. The NIST definition of cloud computing. Technical report, National Institute of Standards and Technology, 2011. URL <https://doi.org/10.6028/nist.sp.800-145>.
- [48] Steve Micallef. Spiderfoot automates osint for threat intelligence and mapping your attack surface. <https://github.com/smicallef/spiderfoot>, 2021. Accessed: 2021-09-26.
- [49] Tomas Mikolov, Edouard Grave, Piotr Bojanowski, Christian Puhersch, and Armand Joulin. Advances in pre-training distributed word representations. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*, 2018.
- [50] Iyatiti Mokube and Michele Adams. Honeypots: Concepts, approaches, and challenges. In *Proceedings of the 45th Annual Southeast Regional Conference*, ACM-SE 45, page 321–326, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936295. doi: 10.1145/1233341.1233399. URL <https://doi.org/10.1145/1233341.1233399>.

- [51] Marcin Nawrocki, Matthias Wählisch, Thomas C. Schmidt, Christian Keil, and Jochen Schönfelder. A survey on honeypot software and data analysis. *CoRR*, abs/1608.06249, 2016. URL <http://arxiv.org/abs/1608.06249>.
- [52] Marco Ochse. T-Pot. <https://github.com/telekom-security/tpotce>, 2021. Accessed: 2021-09-26.
- [53] Michel Oosterhof. Cowrie SSH/Telnet Honeypot. <https://github.com/cowrie/cowrie>, 2021. Accessed: 2021-09-26.
- [54] Sylvain Peyrefitte. RdpY: Remote desktop protocol in twisted python. <https://github.com/citronneur/rdpy>, 2021. Accessed: 2021-09-26.
- [55] Niels Provos. Honeyd - a virtual honeypot daemon. In *10th DFN-CERT Workshop*, Washington, D.C., August 2003. USENIX Association.
- [56] Antonio Regalado. Who coined 'cloud computing'?, Feb 2020. URL <https://www.technologyreview.com/2011/10/31/257406/who-coined-cloud-computing/>.
- [57] Cymmetria Research. Cisco ASA honeypot. https://github.com/Cymmetria/ciscoasa_honeypot, 2018. Accessed: 2021-09-26.
- [58] Lukas Rist, Johnny Vestergaard, Daniel Haslinger, Andrea Pasquale, and John Smith. Conpot ics scada honeypot. <http://conpot.org/>, 2021. Accessed: 2021-09-26.
- [59] Lukas Rist, Johnny Vestergaard, Daniel Haslinger, Andrea Pasquale, and John Smith. Glutton: low-interaction honeypot. <https://github.com/mushorg/glutton>, 2021. Accessed: 2021-09-26.
- [60] Lukas Rist, Johnny Vestergaard, Daniel Haslinger, Andrea Pasquale, and John Smith. Snare: Super next generation advanced reactive honeypot. <https://github.com/mushorg/snare>, 2021. Accessed: 2021-09-26.
- [61] Lukas Rist, Johnny Vestergaard, Daniel Haslinger, Andrea Pasquale, and John Smith. Tanner: He who flays the hide. <https://github.com/mushorg/tanner>, 2021. Accessed: 2021-09-26.
- [62] Markus Schmall. Medpot: HL7 / FHIR honeypot. <https://github.com/schmalle/medpot>, 2021. Accessed: 2021-09-26.
- [63] Bruce Schneier. *Secrets & lies - IT-Sicherheit in einer vernetzten Welt*. Dpunkt-Verlag, Köln, 2004. ISBN 978-3-898-64302-3.
- [64] Pavol Sokol, Jakub Míšek, and Martin Husák. Honeypots and honeynets: issues of privacy. *EURASIP Journal on Information Security*, 2017, 02 2017. doi: 10.1186/s13635-017-0057-4.
- [65] Lance Spitzner. *Honeypots - Tracking Hackers*. Addison-Wesley, Amsterdam, 2003. ISBN 978-0-321-10895-1.

- [66] Clifford Stoll. *The Cuckoo's Egg: Tracking a Spy through the Maze of Computer Espionage*. Pocket Books, 2000. ISBN 0743411463.
- [67] suricata. Suricata. <https://github.com/OISF/suricata>, 2021. Accessed: 2021-09-26.
- [68] University Computing Center Heidelberg. heicloud - the heidelberg university cloud infrastructure. <https://heicloud.uni-heidelberg.de/heiCLOUD>, 2021. Accessed: 2021-09-02.
- [69] University Computing Center Heidelberg. Heicloud. <https://www.urz.uni-heidelberg.de/en/service-catalogue/cloud/heicloud>, 2021. Accessed: 2021-09-02.
- [70] Girish Venkatachalam. The openssh protocol under the hood. *Linux J.*, 2007 (156):6, apr 2007. ISSN 1075-3583.
- [71] Johnny Vestergaard. Heralding: Credentials catching honeypot. <https://github.com/johnnykv/heralding>, 2021. Accessed: 2021-09-26.
- [72] Alexander Vetterl. Honeypots in the age of universal attacks and the Internet of Things. Technical Report UCAM-CL-TR-944, University of Cambridge, Computer Laboratory, February 2020. URL <https://www.cl.cam.ac.uk/techreports/UCAM-CL-TR-944.pdf>.
- [73] Lizhe Wang, Gregor von Laszewski, Andrew Younge, Xi He, Marcel Kunze, Jie Tao, and Cheng Fu. Cloud computing: a perspective study. *New Generation Computing*, 28(2):137–146, April 2010. doi: 10.1007/s00354-008-0081-5. URL <https://doi.org/10.1007/s00354-008-0081-5>.
- [74] Christopher Wellons. Endlessh: an ssh tarpit. <https://github.com/skeeto/endlessh>, 2021. Accessed: 2021-09-26.
- [75] Tillmann Werner. Honeytrap. <https://github.com/armedpot/honeytrap/>, 2021. Accessed: 2021-09-26.
- [76] Tatu Ylonen and Chris Lonvick. The Secure Shell (SSH) Transport Layer Protocol. RFC 4253, RFC Editor, January 2006. URL <https://www.rfc-editor.org/rfc/rfc4253.txt>.
- [77] Michal Zalewski. p0f v3: passive fingerprinter. <https://github.com/p0f/p0f>, 2021. Accessed: 2021-09-26.