

# A case study of unauthorized login attempts against honeypots via remote desktop

Oscar Rehnäck

**Information Security, master's level (120 credits)**  
**2023**

Luleå University of Technology  
Department of Computer Science, Electrical and Space Engineering

[This page intentionally left blank]

# A case study of unauthorized login attempts against honeypots via remote desktop

Oscar Rehnbäck (oscreh-1@student.ltu.se)

*M.Sc. Information Security*

*Luleå University of Technology*

July 28, 2023

## **Abstract**

Remote service software is typically used to establish a connection to an asset on another network. There are a variety of services depending on which asset needs to be accessed and which information needs to be transferred. One of these is Remote Desktop Protocol (abbreviated RDP), a communication protocol that allows clients to connect to another computer over a network. Microsoft developed The protocol and introduced it in their operating systems in the late 90s. The most common authorization method is by using credentials. These can be created locally on the host or managed centrally via Kerberos / Active Directory.

RDP is an attack surface that is heavily exposed. There are several vulnerabilities against this protocol. One is the possibility of eavesdropping on credentials. However, the most common reason intrusions occur via RDP is not because malicious actors have obtained the credentials via eavesdropping. They have managed to guess those with a dictionary- or brute force attack. This observational study was performed with three honeypots that were exposed to attacks via remote desktop for 37 days. More than 120,000 login attempts were recorded and the first attempts occurred within 24 hours. One of the research questions being studied is how availability is affected for an asset that is applied with a login rate limit. As this kind of control can be abused and exploited as a denial of service attack.

One of the honeypots was configured with "Account Lockout Policy" which is an integrated feature into the Microsoft Windows operating system. The policy was configured according to Microsoft's recommendation. The results show that the brute force attacks had a small impact on the availability. However, this is mainly due to the fact that the most active malicious actors did not target the administrator's account in their attempts to gain access. If they had chosen to do so, availability would have been significantly more affected.

Another honeypot was configured to use a non-standard port for the remote service, to study whether attacks can be avoided by trying to hide that the service is active and available. This turned out, not to be a good security enhancement as the remote service on this honeypot was discovered after 15 days and login attempts were conducted by several different actors. Previous research on attacks against the remote desktop has shown that this is an attractive target and a common attack surface. The results of this study support and confirm this.

# Acknowledgement

I would like to thank Christer Åhlund (Chaired Professor) and Saguna Saguna (Associate Professor) at Luleå University of Technology for the support and all good advises with this study. Would also like to thank family and friends supporting me with this Master Thesis.

# Contents

<b>1</b>	<b>Introduction</b>	<b>6</b>
1.1	Background	6
1.1.1	Connection restrictions	7
1.1.2	Remote desktop	7
1.2	Research questions	9
1.3	Summary	9
<b>2</b>	<b>Methodology</b>	<b>10</b>
2.1	Process plan	10
2.2	Reliability	11
2.3	Validity	11
2.3.1	Bias	11
2.4	Summary	11
<b>3</b>	<b>Background and Related Work</b>	<b>12</b>
3.1	Literature review technique	12
3.2	Theoretical Background	14
3.2.1	Information security theory	14
3.2.2	Information- and cyber security	14
3.3	Summary	15
<b>4</b>	<b>Observation setup</b>	<b>16</b>
4.1	Honeypots	16
4.1.1	Honeypot 1	16
4.1.2	Honeypot 2	16
4.1.3	Honeypot 3	16
4.2	Data collection	17
4.3	Summary	17
<b>5</b>	<b>Results Analysis</b>	<b>18</b>
5.1	General honeypot data analysis	18
5.2	First attacks	19
5.3	Total volume and number of attempts per day	19
5.4	IP-addresses attacking	22
5.5	Frequency of attacks and sessions	24
5.6	Account lockout	25
5.6.1	Honeypot 1	25
5.6.2	Comparison between honeypot 1 and 2	27
5.7	Analysis of the impact of availability	27
5.8	Strategies and patterns	28
5.9	Final analyzes and observations	28
5.10	Summary	28
<b>6</b>	<b>Conclusions</b>	<b>29</b>
<b>7</b>	<b>Appendix A</b>	<b>32</b>
<b>8</b>	<b>Appendix B</b>	<b>33</b>
<b>9</b>	<b>Appendix C</b>	<b>34</b>
<b>10</b>	<b>Appendix D</b>	<b>35</b>

# List of Figures

1	Example of a client logging in to a remote service host applied with authentication rate limiting. . . . .	6
2	Rate limiting applied in the Google Analytics reporting API. [8] . . . . .	7
3	Number of hosts with port 3389 accessible indexed by Shodan [17]. . . . .	8
4	Number of hosts with RDP on port 5000 indexed by Shodan [17]. . . . .	8
5	Research process plan with 4 phases and 7 steps. . . . .	10
6	An overview of how data is created, processed, and stored. . . . .	17
7	Total number of login attempts per day for all honeypots. The graph shows an uneven and unpredictable pattern where individual actors contribute to the spikes depending on the intensity of the brute force attack. First day online, no login attempts were made therefore this date has been excluded from the figure. . . . .	20
8	Total number of login attempts per day: Honeypot 1. The spike on March 25 is depending on a single actor from IP address 37.46.x.x (LU - Luxembourg) who conducted a large number of attempts. This actor also made an attack against honeypot 2 on March 23 (Figure 9). . . . .	20
9	Total number of login attempts per day for honeypot 2. The spike around March 23 was due to a brute force attack from two IP addresses 37.46.x.x (LU -Luxembourg) and 212.102.x.x (DE - Frankfurt am Main). . . . .	21
10	Total number of login attempts per day for honeypot 3. The attempts last days online were coming from IP addresses 185.190.x.x (AX - Mariehamn) and 5.181.x.x (SC - Victoria). . . . .	21
11	Shows the total number of unique IP addresses conducting login attempts against all honeypots per day. Addresses that have sessions over several days are included as one unique per day. The graph shows that almost every day, new IP addresses were registered that conducted login attempts. . . . .	22
12	Number of unique IP addresses attacking the number of days. This shows that only a few actors made login attempts over several days. . . . .	23
13	Top 10 IP addresses and their origin, with the highest total number of login attempts against all honeypots. The two addresses with the highest number of login attempts accounted for approximately 50% of all recorded attempts. . . . .	23
14	This figure shows the average delay between login attempts per IP address with the number of attempts. The first data point at 20 seconds has three IP addresses with a total number of 11641 login attempts. . . . .	24
15	Session data for top 10 IP addresses with most login attempts. Row color marks similar addresses. The results are summed for all honeypots. The time format is in HH:MM:SS. A new session is categorized when there are 10 minutes or more between attempts. The reason to choose 10 minutes is because of the proposed 10/10/10 rule from Microsoft [13] regarding account lockout configuration where an account should be locked for 10 minutes after 10 invalid login attempts. . . . .	24
16	The lockout statistics simulate the account lockout if all login attempts would have been against the administrator account. This statistic is based on the delay between all login attempts on honeypot 1. Potential lockout statistics for the administrator account. . . . .	26
17	Comparison between honeypot 1 and 2. Time format HH:MM:SS. The addresses from US - Portland has relatively few login attempts with several minutes between each attempt. All these addresses follow the same pattern, which indicates that it could be the same actor who conducted the login attempts from different addresses. . . . .	27
18	Data collected in a log entry from Windows event log . . . . .	33

# List of Tables

1	Categories and metrics used in honeypot data analysis . . . . .	18
2	Elapsed time from the time where the honeypots are started to the first login attempt. . . . .	19
3	Total number of login attempts per honeypot. . . . .	19
4	Number of unique IP addresses per honeypot . . . . .	22
5	Number of IP addresses conducting login attempts against multiple honeypots. It was only the two honeypots that used the standard RDP port that had login attempts from the same IP address. . . . .	22
6	General honeypot details . . . . .	32
7	Categories and metrics used in honeypot data analysis . . . . .	34
8	Top 50 username . . . . .	35

# 1 Introduction

Since the concept of having digital assets available online, it has been a great challenge to keep these secure but still available for users or clients. Accessing resources remotely is a very common and natural occurrence today. Regardless if the resource is a document, image, or server, it needs to be protected against unauthorized access in many cases. Many security mechanisms exist to achieve this, such as locks (authentication) and gates (intrusion protection). Despite this many intrusions and attacks are initiated via vulnerabilities in the remote access mechanism. Often because of weak locks and deficient gates.

On every occasion a brute force attack is conducted, there is a risk that the attack will succeed. However, the likelihood is drastically reduced when using strong passwords [1]. Unfortunately, users often select passwords that are vulnerable against dictionary- or brute force attacks [2]. Tools used for these types of attacks such as NCRACK [3] and THC-Hydra [4] are open-source and can target several protocols including SMB, FTP, SSH, HTTP, and RDP. An attack can be initiated with a single command line and configured with connection delays and anonymous connections via a proxy network [5].

## 1.1 Background

When using an authentication method that applies something that a user knows, NIST recommends implementing controls against brute force attacks [6]. This control which is often called rate limit or throttling applies a limit for how many login attempts a user or client has. When the limit is reached, the user can be blocked, and/or the account can be locked. This restriction rule is usually applied for a limited period of time. Figure 1 describes an example where rate limiting is implemented for requests to a remote service.

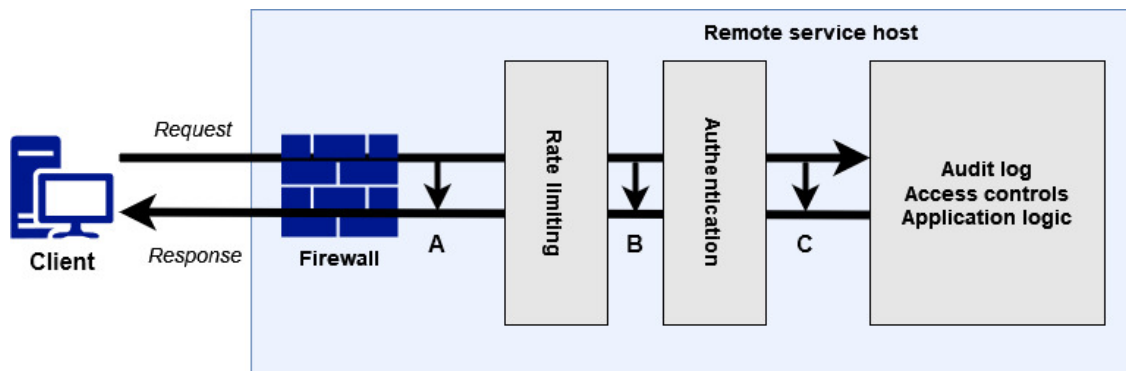


Figure 1: Example of a client logging in to a remote service host applied with authentication rate limiting.

In the example in Figure 1, a request can be denied at locations A, B, or C. In the firewall (A) due to restrictions for example the IP address, port number, or service requested. This request will not affect the rate limit mechanism. The authentication process will not be performed if the request passes the firewall but not the rate-limiting (B). If the request passes rate-limiting the authentication process (C) can be conducted.

Rate limiting is a protection mechanism commonly used for HTTP-based APIs [7]. The purpose can be business reasons or to protect hardware resources from being overloaded. Requests that initiate an operation that requires a lot of hardware resources like CPU, GPU, RAM. or bandwidth could otherwise be a target for denial of service (DoS) attacks.

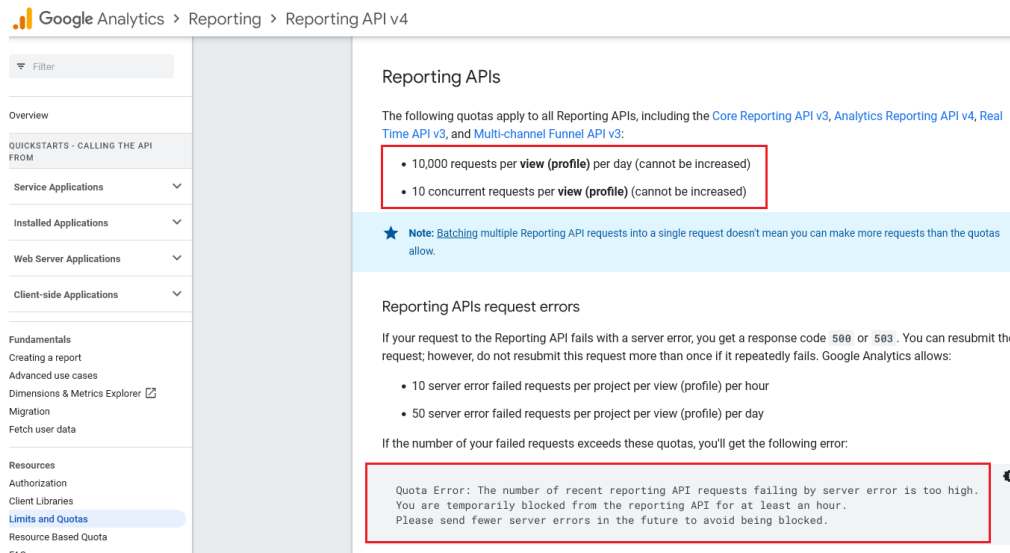


Figure 2: Rate limiting applied in the Google Analytics reporting API. [8]

### 1.1.1 Connection restrictions

A blocking mechanism is usually referred to restrictions of characteristics derived from the connection source such as IP address, user agent, patterns, or behavior in the TCP handshake. The block could be implemented and applied in a firewall or an IPS/IDS. Account lockout is a mechanism where all login attempts will be denied when a certain number of failed login attempts have been performed. When the account is locked there is not possible to log in even if the correct credentials are given. This control can be set with a time limit. When the time limit is reached, the blocking or locking mechanism is removed. When account lockout is implemented there is a potential risk for denial of service attacks [9], [10]. This occurs if a malicious actor constantly tries to brute force an account with a sufficiently high number of login attempts within a small period of time between each attempt. Defensive security controls are suggested to prevent this to happen but not all can be applied to existing remote services by default.

IP restrictions are not an effective security control against DoS attacks due to an attacker often having several machines with different IP addresses available. There is also a risk to restrict addresses from legitimate users [10][11].

### 1.1.2 Remote desktop

Remote service authentication, which is carried out via a network or communication link is one area where account lockout can be used. These services can be used through several different methods and with various types of network protocols. For clients that want to connect to a Microsoft Windows client is Remote Desktop Protocol (RDP) one method. This remote service was introduced with Microsoft Windows XP and is available for all succeeding versions of Windows. It provides an easy way via a graphical user interface for users to connect to a remote Windows machine by authenticating with for example a username and password, Other authentication methods exist but the use of credentials is the default option. The service uses port 3389 by default [12], for communication via TCP- or UDP-protocol which makes this service easy to discover and enumerate by malicious actors. RDP is also the most exposed attack surface within many industries and branches [13] [14]. "It is likely that any computer exposed to the internet via RDP is of interest to criminal hackers and the subject of frequent attacks" [15]. As an attack

surface is not a new trend but due to that it is very easy to attack it has been a popular target for several years. An attacker can for instance perform an RDP-based ransomware attack which is a method that has grown rapidly over the last couple of years [16]. When the attacker got access to a system via RDP the actor gets partly or full control of the system depending on the user's privileges. Based on this, the attacker can perform a lateral movement or privilege escalation and then install all kinds of malicious software.

Microsoft has recently taken measures by introducing a control aimed at reducing the possibility of a successful brute force attack. These are installed and activated in their latest operating systems (Windows 11) but have recently also been released via a major security update [13]. The disadvantage, however, is that users need to configure and activate the mechanisms manually. When installing the control from the update. This means that organizations that do not follow recommendations today will possibly not activate this protection. But as more and more devices get this protection activated automatically, this might affect the development of this trend.

Shodan reports that the number of hosts with port 3389 is over 3 million (February 2023).



Figure 3: Number of hosts with port 3389 accessible indexed by Shodan [17].

The default port of RDP can be changed to any number. It is not recommended as a prevention strategy against attacks but may prevent detection in some cases [18]. If an attacker only scans for the default RDP port number the service is not found, but can easily be discovered with other port scanning techniques.

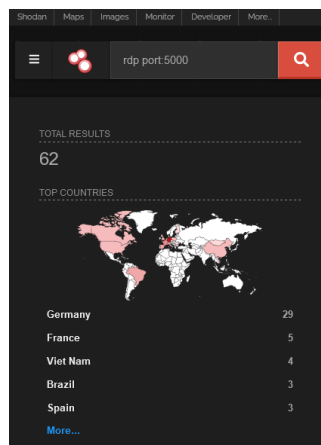


Figure 4: Number of hosts with RDP on port 5000 indexed by Shodan [17].

## 1.2 Research questions

Remote services are one of the most common attack surfaces that exist today. There are several attack vectors that malicious actors are using. One of these is by guessing login credentials for a specific account. Despite the fact that there are many recommendations and knowledge about remote services having a high likelihood of cyber attacks, there are still a large number of vulnerable targets. A defensive control against brute force attacks is account lockout which limits the number of failed login attempts. But this control can also pose a threat to the availability of the asset. There is little research regarding patterns and behaviors for attacks against RDP accounts in the wild that have been applied with account lockout.

This leads to the following research questions for this Master thesis:

- How does account lockout policy affect availability for an administrator account on a host with remote desktop?
- What impact does account lockout policy have on the number of recurrent attacks?
- Which retry and delay connection strategies are used by attackers when brute forcing?
- Do these strategies differ for targets that have applied an account lockout policy?
- How does changing the RDP port number affect the number of recurring attacks?

A case-control study will be performed to answer these questions where a systematic observation is conducted. The work with this master's thesis starts in January 2023 and ends in June 2023. Three honeypots will be deployed and observed in real-time. These honeypots will be publicly visible and exposed online. Remote desktop services will be activated and a developed software will monitor connection attempts. The purpose of this software is to collect unstructured data and store it structurally in an SQL database. Based on the collected data a quantitative uni-variate analysis will be conducted.

## 1.3 Summary

This chapter gives the background and current challenges related to remote services. It focuses on Remote desktop and known vulnerabilities that exists against this protocol today. The chapter ends with research questions included in this study with a motivation.

## 2 Methodology

The research process plan (figure 5) consists of seven steps and is divided into four phases. Each phase must be completed before the next phase can begin. This is especially important so that the results from the observation are based on the same honeypot configuration. A change in the configuration during the observation could affect the result. Also, the analysis cannot begin before the entire observation phase is completed.

### 2.1 Process plan

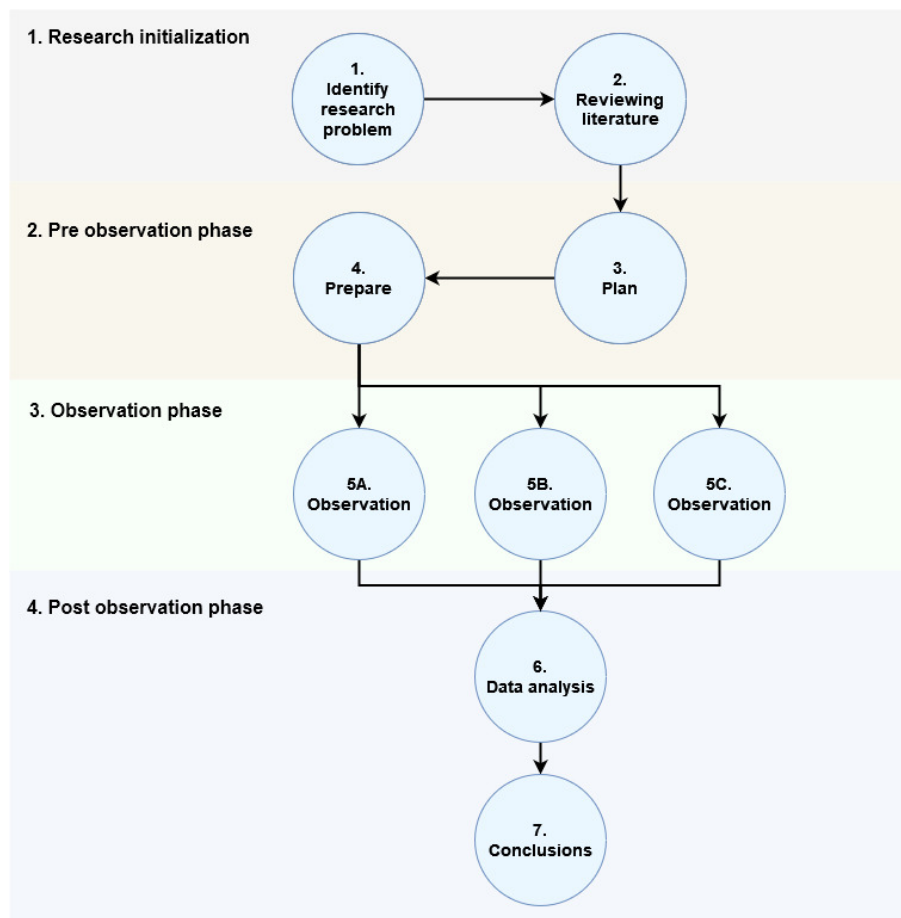


Figure 5: Research process plan with 4 phases and 7 steps.

Brief description of steps from the research process plan:

1. Identify research problems by examining and evaluating existing problems related to the topic that is described from publicly available sources.
2. Existing literature on the topic will be searched via scientific databases. I will primarily focus on literature that are published within the last 5 years.
3. There are several things that need to be considered and decided in the planning. The goal of this stage is to clarify what requirements that are needed in order to perform the observation and analysis. Based on the requirements a plan needs to be formed on how to prepare an environment that is able to perform the observation. This plan will specify

which environment, hardware, software, and other tools that will be used. As well as a specification of what type of data that will be collected and stored.

4. Before the observation can be started an environment needs to be prepared. Hardware and software need to be installed and configured. System settings needs to be applied such as firewall adjustments, settings for RDP connections, and account lockout policy.
5. When the necessary preparation has been completed, observations can begin. The observation is monitoring login attempts via RDP. Whenever a login attempt occurs via RDP, data will be generated and stored. This study will perform observations from three different hosts which are specified with observation locations 5A to 5C in the process plan.
6. Analysis can be started when the data collection process is completed. Due to that, the analysis considers data from a specific time period additional data should not be added after the analysis has started. If any data is supplemented after the analysis has begun the analysis needs to be restarted from the beginning.
7. Results and analysis are discussed and conclusions are drawn.

## **2.2 Reliability**

The literature review has been performed by reviewing articles, conference journals, and papers that are reviewed, published, and searchable in scientific databases. Creating and configuring honeypots as described, can be fully replicated. The hosting of the honeypots has taken place in a publicly available cloud environment. Data collection has been done by using integrated log management in the operating system. However, recreating the result using the same conditions is not possible as the honeypots are publicly exposed to everyone. Conclusions made have been based on the results reported in this thesis.

## **2.3 Validity**

Data from the collection process has not been calculated or transformed in any way. All data has been interpreted directly from what is reported in the log management from the operating system. The single data processing that has taken place is to transfer the information from a log entry to an SQL database. This process has been done through a clean copy, where the information has not been processed or transformed in any way. The copy process has been conducted through a custom-developed .NET application that was verified before the observation started. The verification consisted of checking that the correct information from the log entry was copied to the correct field in the database.

### **2.3.1 Bias**

Some knowledge may have been acquired without being referenced or cited. This can, for example, be based on lessons and lectures that took place during the course or Master's program. Since these are just assumptions without citing available research, there is a possibility that these can be disproved if the research can show other conclusions. When reviewing the literature, I try not to draw my own personal conclusions/opinions based on what is reported. However, there may be cases where I consider something obvious without mentioning it in the literature. These assumptions are not based on any self-conducted or published research but on knowledge gained when reviewing the subject.

## **2.4 Summary**

This chapter describes how the study will be conducted. An overall plan is presented where all the steps and phases are included. It also describes the literature review technique used and which keywords that are included in the searches. The chapter also includes a description of the data collection process and how each honeypot is configured. It ends with reliability, validity, and bias.

## 3 Background and Related Work

Several vulnerabilities related to RDP have been discovered that enable man-in-the-middle attacks or that authentication data could be eavesdropped. "Actually, there is not much to break. RDP is already completely broken by design" [19]. Clients that offer connections via RDP are a popular target for attacks and intrusion attempts [15]. Palo Alto Networks writes in its Breach Report from the year 2020 "the number one initial attack vector was through RDP services, occurring in 50% of our ransomware matters" [20]. In some cases, minor configuration updates are usually required to greatly improve the security of an RDP client. Despite this, there are a large number of clients that offer RDP with inadequate security settings. The reason why security-enhancing measures are not taken may be due to ignorance or that one is not aware of the risks this entails. Lack of knowledge about how the system should be configured can also be a contributing factor [21].

There are probably several reasons and factors why RDP composes such a large attack surface. As written in the paper from Sophos [15] it is easy to find suitable targets and easy to deploy an attack. In most cases, there are no exploits deployed or eavesdropping it is all about guessing a username and password with brute force techniques. Already back in the year 2003, it was recommended to use a "strong" password for RDP users [22]. It is often said that a strong password is one that does not exist in a dictionary and can thus easily be cracked via a dictionary attack [23]. The password must also not be able to be cracked via brute force within a reasonable time, which requires that the password has a certain number of characters [24]. NIST (National Institute of Standards and Technology) recommends that a password should consist of at least 8 characters [6]. There are of course exceptions. These passwords might look like they could be categorized as strong: "vfkfrbxt@9012259", "74111ummDMZE6". But in fact, if any of these passwords would be used they could be cracked within seconds due to that they exist in a popular password dictionary [25]. This shows how difficult it is for humans to find a secure and strong password. Even if a user believes she or he is composing a strong and secure password by adding numbers or special characters that may be completely wrong which is shown in the study by UR, Blase, et al [26].

Controls applied are intended to avoid or reduce a threat. But if the control is not installed or configured correctly, this can lead to the occurrence of new threats. For example, a system administrator that applies IP restrictions and includes addresses from legitimate users will implicate that the availability ceases for these users. A similar scenario can be used for account lockout where users are locked out after a certain number of failed login attempts. This in turn can be exploited by malicious actors to generate a DoS attack against users or systems [10].

### 3.1 Literature review technique

The literature review will identify what has been written in relation to the research problems. The review also aims to identify what has been written about honeypot analyses.

These databases and search engines have been used to search for relevant papers, articles, and conference journals for the literature review and related work:

#### Databases used in the literature review

- IEEE Xplore.
- Science Direct by Elsevier.

### **Search engines used**

- Semantic Scholar.
- Google Scholar.

### **Keywords**

One or more keywords were combined to conduct the searches. These keywords were used individually or in combination

- Remote services
- RDP
- Remote Desktop
- API
- Information security
- Cyber security
- Brute force
- Local account policy
- Account lockout policy
- Windows
- User awareness
- Behavior
- Honeypot
- Administrators
- Eavesdropping
- MITM
- Password
- Threat
- Intelligence
- Risk
- Cloud

## 3.2 Theoretical Background

The selected research problem can be considered from several different perspectives. Partly it can be seen from a technical perspective where something expected to function in a certain way does not. Partly how users implement and use technology. But it is also possible to study it from an alignment perspective where directives or recommendations given are not followed. Birman states that information security is not purely about technical issues it could be strategic as well [27]. Today, many would argue that the weakest link in an IT security chain is the user [28][29]. A user can contribute to increased security risks in several ways. Some of these are through a lack of knowledge or awareness. William Stallings et al. write in [9] that the key to getting employees to follow policies and procedures related to security is to increase awareness and knowledge. Employees more often tend not to follow policies if they do not have knowledge of the risks this entails.

Many organizations strive to get everyone to achieve the same common goal. Most often, this requires strategies and objectives that are implemented and used by internal and external stakeholders so that the entire organization is focused to achieve the shared goals. Lindström, John, et al. further states that problems with the alignment should be considered by all organizations in order to comply with security controls and standards [30]. NIST Special Publication 800-39 states that "Strategic alignment of risk management decisions with missions and business functions consistent with organizational goals and objectives". In addition, IS architecture is an embedded integral part of the enterprise architecture that is aligned with the enterprise's mission and strategic plans. This indicates that NIST assesses that a secure infrastructure is one that is in alignment with the other goals of the organization. Achieving strategic alignment can be a complex process that is often unique for each individual organization [31]. Alignment is an important process within organizations to reduce risks and avoid vulnerabilities. Alignment theory could have been applied to this study if the focus would have been to observe how organizations manage remote services attacks.

### 3.2.1 Information security theory

According to ISO 27002, the very core of IS/IT security is information. This is one of the primary assets that need to be protected. Information needs to be available, reliable, and in many cases confidential. The information does not only include data that is stored on a physical medium such as hard drives, tapes, or papers but also intangible assets such as knowledge, ideas, and concepts [32]. Horne et al. state that information that is protected with controls becomes a robust resource that is resistant against threats [33].

Information that has ineffective or missing security controls leads to information being degraded due to an increased risk that the information is not reliable, available, or confidential. Even if controls are implemented there always exists threats that are unknown which means that information can be degraded.

### 3.2.2 Information- and cyber security

Even if controls are implemented there will always be threats that are unknown which means that information can be degraded. Information security theory (IST) focuses on an assessment where the application of controls increases the likelihood that information can become a resource. A malicious actor who tries to get access and control over a system may have other intentions than getting access to the information on the system. For example, the intention could be to gain access to resources that the asset has such as computing power or bandwidth. For those cases, it is the underlying asset that has been affected not the availability, integrity, or confidentiality of the information [34]. R. Von Solms, J. Van Niekerk distinguishes the concepts of information security and cyber security. Where the term cyber security among other things refers to the protection of the environment and infrastructure that organizations and users have at their disposal. An attack against information security does not necessarily be the same thing as a cyber attack. A cyber attack is conducted by one or a group of people or with automated software where the attacker has certain knowledge about the target [35]. This constitutes a limited number of threats against the infrastructure or the information security due to that this

type of attack does not for example includes natural events or technical interruptions. In cases where a remote system is under attack via brute force, this starts as a cyber attack but can turn into an attack against information security.

This study primarily focuses on studying the patterns and behavior of cyber attacks against remote systems. Although this study implicitly includes access and availability of information, it is primarily attacks on the infrastructure that is studied. However, it can be argued that if someone has unauthorized access to information but does not affect it, they still have the opportunity to do it even if that is not their intention.

### **3.3 Summary**

This chapter gives a theoretical background against the research topic. A presentation about what has been studied before and a brief description of the results from some of these studies. The chapter includes a more detailed description of different theories in information- and cyber security.

## 4 Observation setup

This chapter will describe details about the honeypots and the data collection process.

### 4.1 Honeypots

In order to observe attacks against RDP, three honeypots will be configured and observed. These will serve as attack targets. The most common definition of a honeypot [36] is a "security resource whose value lies in being probed, attacked or compromised" [37]. These will be designed as low-interaction [38] server honeypots (LIHP) where the attacker is not able to log in and interact with the operating system. They are created in a cloud environment at the AWS data center located in Stockholm (further details in Appendix A). Initially when the instances are created all incoming ports are closed in the firewall from external access, except port 3389 (RDP) which is allowed only from one specific IP address.

To attract attackers to the honeypots there will be no IP- or other firewall restrictions against connecting via RDP. The hosts are responding to ICMP requests and all opened ports are reported as opened when conducting a port scan in Nmap [39]. To prevent attackers from successfully login via RDP, a randomized password with a length of 32 characters will be used. The honeypots are not registered or advertised in any way such as in internet forums or similar.

On each honeypot is a custom-developed .NET application installed which makes a lookup against the event log database. All log entries with ID 4625 will be collected and processed further. This application is configured to execute every 10 minutes.

All honeypots are configured to only use the built-in Administrator account. All other accounts are disabled. This is also reflected in the response to a login attempt. For example, trying to log in with the username "test" the response will be "bad username/password". With the username "administrator" the response will be "incorrect credentials". The response used is the default messages integrated with RDP.

#### 4.1.1 Honeypot 1

This honeypot was configured and activated with an account lockout policy according to the recommended 10/10/10 rule [13]. This means that an account gets locked after 10 invalid login attempts that occur within 10 minutes. The lockout is active for 10 minutes. This policy applies to both local login attempts and attempts via RDP.

#### 4.1.2 Honeypot 2

No account lockout policy configured or activated or any other rate-limiting restrictions related to login via RDP. The purpose of this configuration is to be able to compare this honeypot with honeypot 1. The idea is that this honeypot should be a more attractive target as brute force attacks can be used without limit.

#### 4.1.3 Honeypot 3

No account lockout policy configured or activated or any other rate-limiting restrictions related to login via RDP. The standard port for RDP was closed, and instead, port 6781 was opened for RDP access. The purpose of this configuration is to be able to study attack patterns when a none standard configuration is used.

## 4.2 Data collection

When a login attempt to a host via RDP is conducted, a security event log will be created by Windows. Failed login attempts via RDP are saved and marked with a specific log id (4625). In the log entry are the timestamp, username, and IP address registered among other things. Details about the content of a log entry can be seen in Appendix B. These log entries will form the base of the data collection and analysis. From the IP address collected in the log will a further lookup be conducted in order to collect more data about the location and whether the address belongs to a proxy network such as TOR or I2P. An illustration of the complete data collection process can be seen in Figure 6.

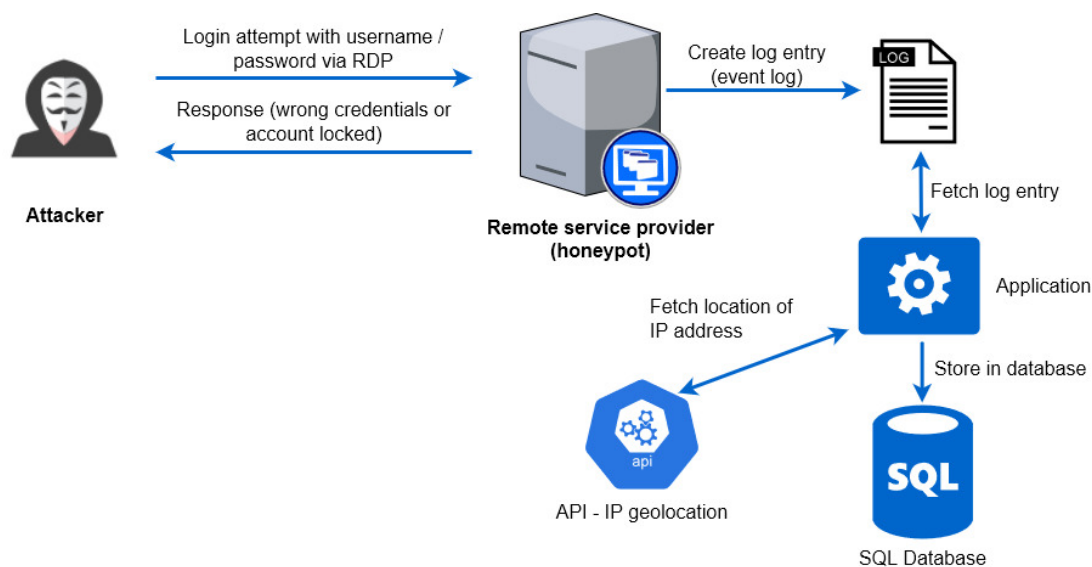


Figure 6: An overview of how data is created, processed, and stored.

### Description of figure 6

A client completes a login attempt via RDP to one of the honeypots. Login attempts will be denied with a response that the wrong credentials have been entered or that the account is locked. At the same time is a log entry with id 4625 created by the Windows operating system. An application does a lookup against the event log database to retrieve all logs with id 4625. Information from the logs is gathered and sent to be stored in an SQL database. If any new IP addresses are discovered they will be sent to an IP-geolocate API [40] to collect metadata about the IP address. This process together with saving this information is also conducted by the application. Details about which data elements are processed and stored in the database can be seen in Appendix B.

## 4.3 Summary

Three low-interaction (LIHP) honeypots will be configured and observed. They will be publicly exposed online with RDP service (Remote Desktop) activated. Data will be collected from Windows security event log and stored in an SQL database. Metadata about the IP-addresses conducting login attempts will also be collected.

# 5 Results Analysis

In this section will the results from the observation be presented. Most of these results are presented in tables, figures, or charts for improved readability. The chapter first presents a summary of the data collected and will be followed up by a more analytical view. The starting point is to analyze and interpret the results based on the research questions. Therefore has the analysis part been divided into sections related to the research problem.

All three honeypots were publicly available via RDP between March 15th and April 20th year 2023 which means that they were exposed to attacks for 37 days. The three honeypots are numbered/identified as 1,2 and 3, more information about the honeypot details can be seen in Appendix A.

## 5.1 General honeypot data analysis

Raw data that forms the basis of the analysis is created through log entries on the honeypots. Which metrics and with what methods to analyze depends of course on what type of honeypot is used and its purpose. Generally, there are more metrics to analyze the more interaction the honeypots offer. Due to that these honeypots are of a low-interaction type all metrics can not be applied such as vulnerability analysis, attack risk assessments and exploit detection. This data analysis will look into several metrics that are further described in [41]. These categories will be included in the analysis:

Table 1: Categories and metrics used in honeypot data analysis

Category	Description	Metrics
Attack source	Identify origin of attack.	IP-address, Country, City, ISP.
Attack target	Identify which honeypot that is attacked.	Time stamp, Hostname.
Attack frequency	Time until first attack and frequency of attacks.	IP-address, Hostname. Username, Sessions per time unit, Session duration, Time between sessions, Number of attacks per session, Number of attacks per time unit.
Propagation of attacks	Correlations between attacks on multiple honeypots.	Propagation graph, Attack graph.

One of the most important data elements that are collected is the IP address where the attack origins. But for all attacks, you need to consider that the IP address may be spoofed or is derived from a proxy network. For this study, unfortunately, there is no way to check or further explore whether an attacker's address is valid or not. Except for IP addresses that are registered to belong to an exit node for a proxy network. These addresses are most likely not the address from which the attacker initiated the connection. For this analysis, one unique IP address combined with the name of the workstation will be considered as one attacker. If an attack origins from one IP address but has different workstation names each "workstation" will be considered as an attacker even if the address is the same.

## 5.2 First attacks

The first attacks occurred on 16th March at 9 am from the city of Seoul which is around 15 hours after the honeypots are exposed. As previously mentioned, the IP addresses of these honeypots have not been marketed or otherwise been published publicly online. To be able to find this host, one therefore either needs to get information from the cloud provider or make a qualified or random guess with ping sweeps or a similar method.

Table 2: Elapsed time from the time where the honeypots are started to the first login attempt.

Honeypot	Elapsed time	Origin of IP-address
1	15 hours	Korea - Seoul
2	16 hours	Ukraine - Kyiv
3	15 days	Russia - Moscow

For two honeypots, the first login attempts were made within 24 hours of being publicly exposed online. Remote desktop for these two were configured against the default port (3389). This result supports previous studies [15][20] on initial attacks against hosts that expose remote services publicly.

For honeypot 3, the first login attempts were conducted after 15 days (table 1). This indicates that changing the default port for remote services is not an effective strategy to avoid detection.

That publicly connected computers are exposed to intrusion attempts within a very short time from the exposure is a fact, but to what extent does this occur? In the following section will results regarding the volume and number of login attempts per day be presented.

## 5.3 Total volume and number of attempts per day

All honeypots received a total number of 121.648 login attempts.

Table 3: Total number of login attempts per honeypot.

Honeypot	Number of login attempts
1	60925
2	12882
3	47841

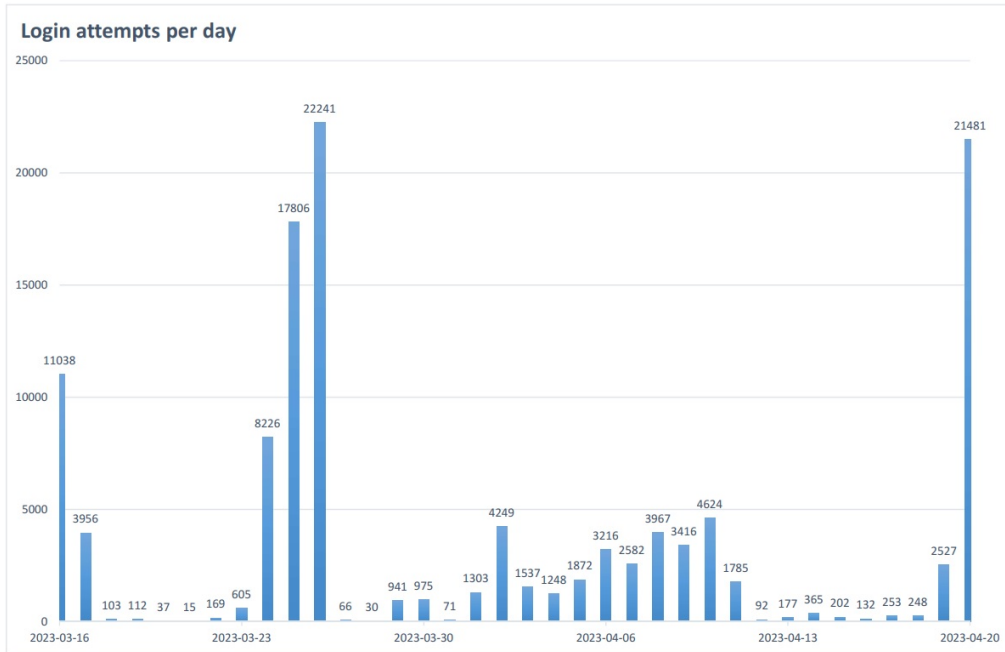


Figure 7: Total number of login attempts per day for all honeypots. The graph shows an uneven and unpredictable pattern where individual actors contribute to the spikes depending on the intensity of the brute force attack. First day online, no login attempts were made therefore this date has been excluded from the figure.

The figures below show the number of total login attempts per day for each honeypot.

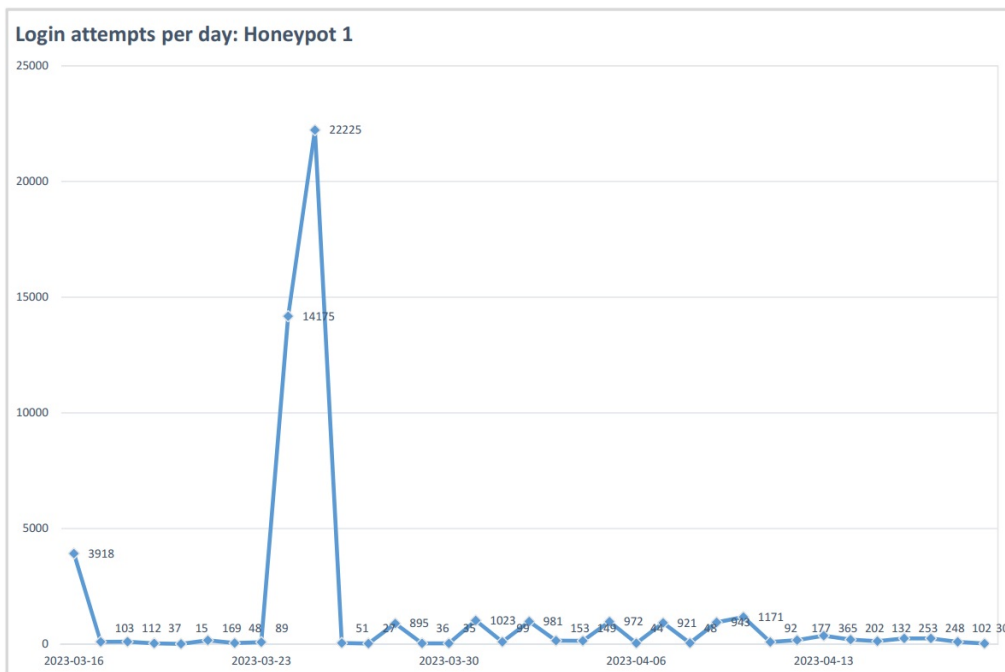


Figure 8: Total number of login attempts per day: Honeypot 1. The spike on March 25 is depending on a single actor from IP address 37.46.x.x (LU - Luxembourg) who conducted a large number of attempts. This actor also made an attack against honeypot 2 on March 23 (Figure 9).

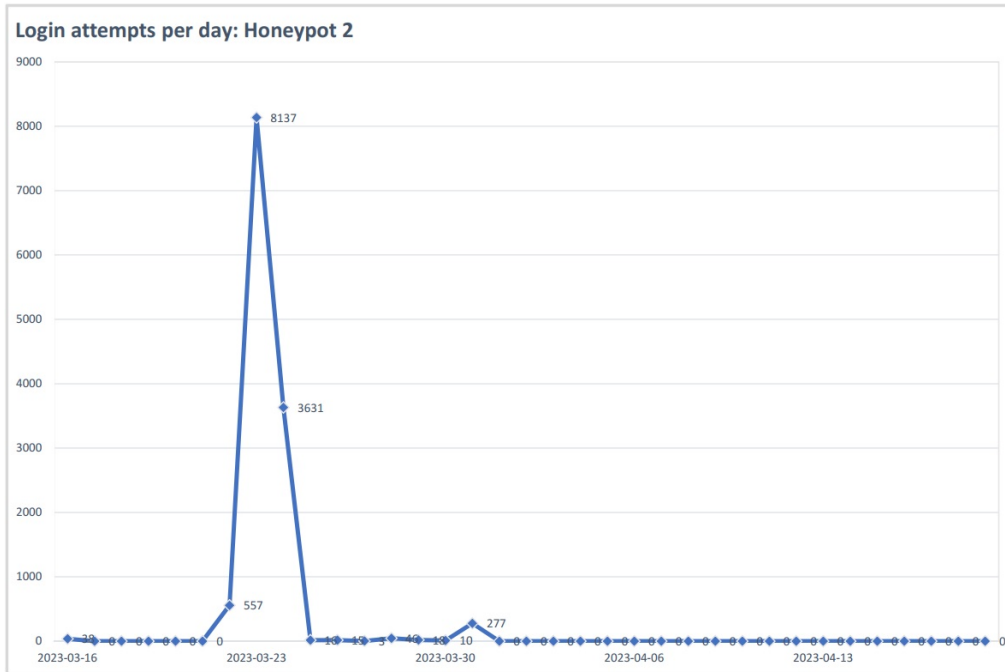


Figure 9: Total number of login attempts per day for honeypot 2. The spike around March 23 was due to a brute force attack from two IP addresses 37.46.x.x (LU -Luxembourg) and 212.102.x.x (DE - Frankfurt am Main).

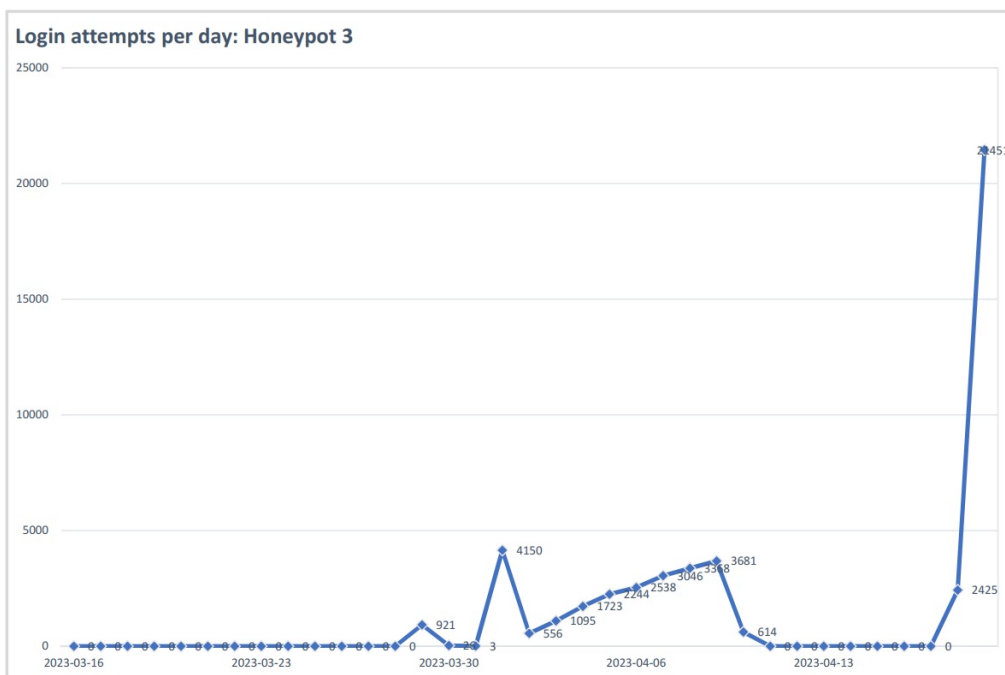


Figure 10: Total number of login attempts per day for honeypot 3. The attempts last days online were coming from IP addresses 185.190.x.x (AX - Mariehamn) and 5.181.x.x (SC - Victoria).

Honeypot 1 is the only host that had login attempts made every single day. For others, the login attempts were sporadic.

As shown, the volume in the number of login attempts is uneven and unpredictable. To study

and understand this pattern more deeply, you need to consider the origin of the actors conducting the login attempts. Is it one actor that has conducted all the attempts on different occasions or thousands? To find out the origin of an actor, the IP address is used. The results regarding IP addresses are followed up in the next section.

## 5.4 IP-addresses attacking

Login attempts were made from 165 different IP addresses across all honeypots. 15 of these addresses are associated with exit nodes for proxy networks. 71 addresses had more than 10 login attempts (43%). 57 addresses had 1 login attempt (approx. 35%). IP addresses that have less than 10 recorded login attempts will be ignored in the figures below. The reason for this is that an IP address with single or very few login attempts does not have to be malicious attempts. But can rather depend on typing errors, tests, or similar. Further details about the origin of the IP addresses can be seen in Appendix C.

Table 4: Number of unique IP addresses per honeypot

Honeypot	Number of IP-addresses
1	133
2	45
3	13

Table 5: Number of IP addresses conducting login attempts against multiple honeypots. It was only the two honeypots that used the standard RDP port that had login attempts from the same IP address.

Honeypots	1,2,3	1,2	1,3	2,3
Number of IP-addresses	0	26	0	0

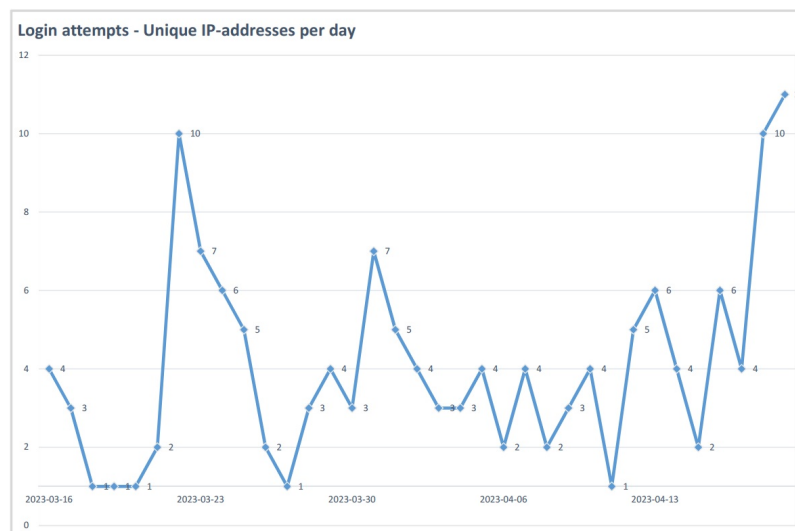


Figure 11: Shows the total number of unique IP addresses conducting login attempts against all honeypots per day. Addresses that have sessions over several days are included as one unique per day. The graph shows that almost every day, new IP addresses were registered that conducted login attempts.

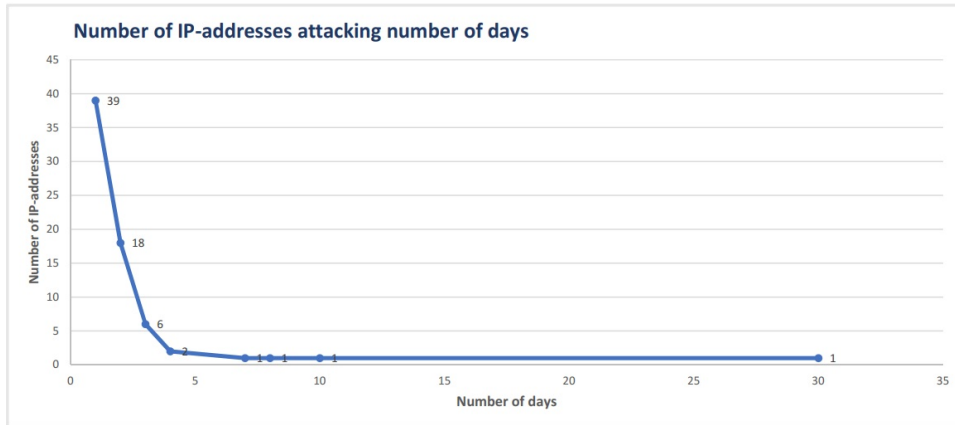


Figure 12: Number of unique IP addresses attacking the number of days. This shows that only a few actors made login attempts over several days.

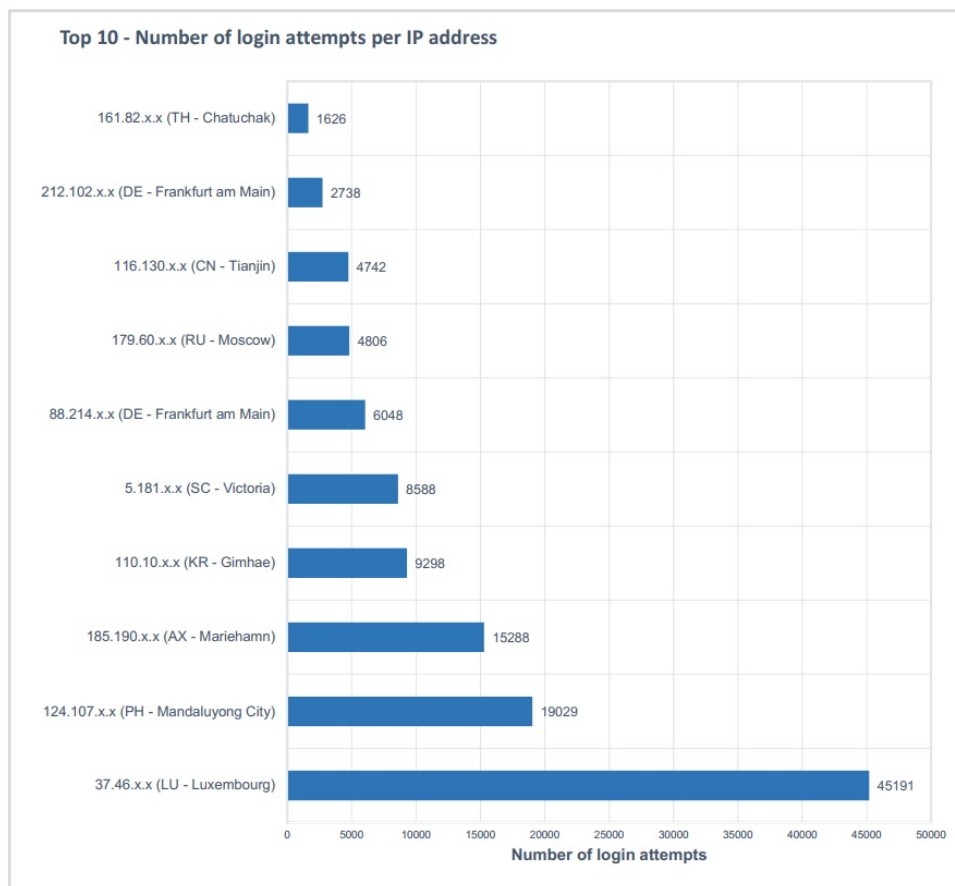


Figure 13: Top 10 IP addresses and their origin, with the highest total number of login attempts against all honeypots. The two addresses with the highest number of login attempts accounted for approximately 50% of all recorded attempts.

The volume and number of actors conducting the attacks only provide a basic picture of the pattern and strategy being used. To be able to analyze how this affects the availability of a resource, you need to look in more detail at the sessions. In the following section, results are presented regarding the frequency of the attacks and sessions.

## 5.5 Frequency of attacks and sessions

This section shows results related to the frequency and duration of attacks. Due to that, this study has a limited amount of time to gather and report results the number of IP addresses for this section has been limited. The results in this section include the top 10 IP addresses that have the highest number of login attempts. Top 10 stands for 100.752 login attempts which are 83% of the total number of attempts.

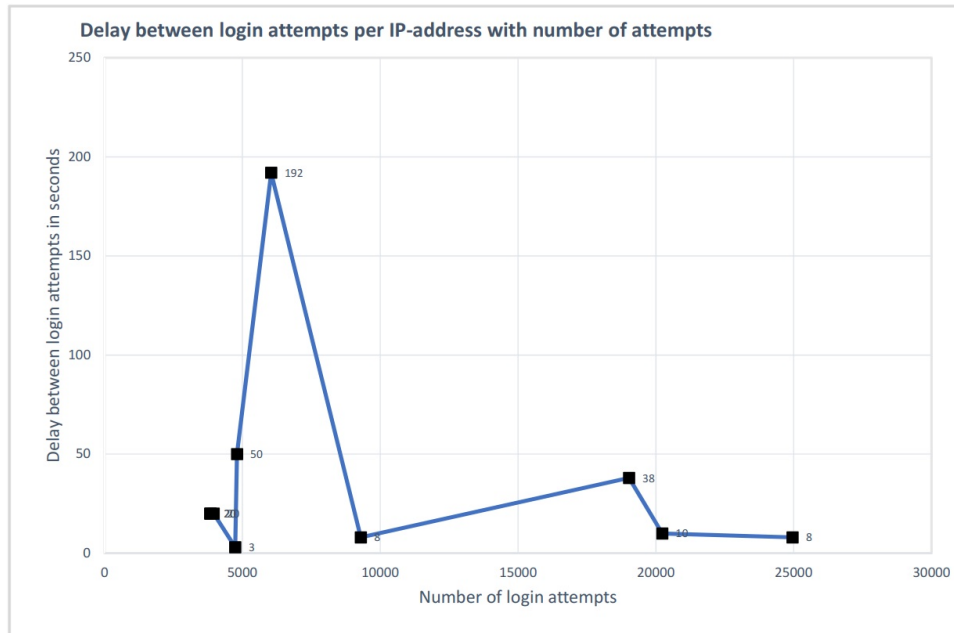


Figure 14: This figure shows the average delay between login attempts per IP address with the number of attempts. The first data point at 20 seconds has three IP addresses with a total number of 11641 login attempts.

IP address	Total number of login attempts	Number of sessions	Session duration			Time between sessions			Number of attempts per session			Average delay between attempts		
			Max	Min	Avg	Max	Min	Avg	Max	Min	Avg	Max	Min	Avg
37.46.x.x LU - Luxembourg	24959	1	02:26:48	02:26:48	02:26:48	-	-	-	24959	24959	24959	00:04:56	00:00:00	00:00:08
37.46.x.x LU - Luxembourg	20232	1	02:23:41	02:23:41	02:23:41	-	-	-	20232	20232	20232	00:06:10	00:00:00	00:00:10
124.107.x.x PH - Mandaluyong City	19028	2	06:59:27	00:31:26	03:45:25	00:54:50	00:54:50	00:54:50	18175	853	9514	00:02:31	00:00:12	00:00:34
110.10.x.x KR - Gimhae	9298	1	23:03:45	23:03:45	23:03:45	-	-	-	9298	9298	9298	00:00:23	00:00:00	00:00:09
88.124.x.x DE - Frankfurt	6048	7	00:16:57	00:14:53	00:16:34	71:42:22	27:01:56	53:33:53	864	864	864	00:01:55	00:00:01	00:00:01
179.60.x.x RU - Moscow	4804	9	09:19:28	00:00:08	01:35:35	30:54:04	00:10:03	05:13:41	3853	1	533	00:06:43	00:00:04	00:00:10
116.130.x.x CN - Tianjin	4742	1	04:04:53	04:04:53	04:04:53	-	-	-	4742	4742	4742	00:00:18	00:00:02	00:00:03
185.190.x.x AX - Mariehamn	3959	8	14:41:23	00:00:01	02:26:40	00:50:59	00:10:12	00:22:34	3398	2	493	00:09:33	00:00:04	00:00:18
5.181.x.x SC - Victoria	3850	10	11:54:09	00:14:51	01:56:44	00:31:41	00:10:48	00:17:16	2904	40	384	00:09:56	00:00:04	00:00:17
185.190.x.x AX - Mariehamn	3832	11	07:07:47	00:00:46	01:08:42	00:48:15	00:10:07	00:20:53	2822	2	348	00:09:47	00:00:04	00:00:17

Figure 15: Session data for top 10 IP addresses with most login attempts. Row color marks similar addresses. The results are summed for all honeypots. The time format is in HH:MM:SS. A new session is categorized when there are 10 minutes or more between attempts. The reason to choose 10 minutes is because of the proposed 10/10/10 rule from Microsoft [13] regarding account lockout configuration where an account should be locked for 10 minutes after 10 invalid login attempts.

In this chapter were data presented about the details of the sessions and frequency. In the following section will these results be connected to account lockout and availability for user accounts.

## 5.6 Account lockout

In this section will results related to account lockout and availability of accounts be presented. Results are based on honeypot 1 which had account lockout activated and honeypot 2 which had account lockout disabled or any other restrictions related to login rate-limiting.

When a login attempt via RDP is conducted against a Windows account that is locked, the host will include a lockout notification in the response. The client that tries to log in can therefore get information if the authentication has failed or if the account is locked. During an account lockout, the account is disabled therefore, there is no point to continue trying to log in even if the correct credentials are used. It is only possible to log in after the lockout period has expired.

### 5.6.1 Honeypot 1

The administrator account which was the only active and enabled account on the host, got 58 lockouts during the whole period of 37 days. Each lockout brought the account to be unavailable for 10 minutes. This means that the account was locked for 9 hours and 40 minutes, approximately 1% of the total up-time. The maximum time between a lockout was 21 hours and the minimum time was 16 minutes. The average time between lockouts was 6 hours and 47 minutes.

- Number of login attempts that received the response "Unknown user name or bad password" for all usernames: 60691. This response is supplied when trying to log in with the username 'administrator' and the wrong password is entered. But also when using another username that does not exist on the system. More about the most commonly used usernames are shown in Appendix D.
- Number of login attempts targeting Administrator account that received the response "Unknown user name or bad password": 1071.
- Number of login attempts that received the response "Account is locked": 230.
- Number of login attempts that received the response "Account is disabled": 4 (this response message is generated when conducting a login attempt with the disabled guest account).

If all login attempts on honeypot 1 would have been targeting the administrator account, the number of lockouts would be higher.

Date	Number of lockouts	Average delay between lockout	Total lockout time (HH:MM)	Total lockout time of day
2023-03-16	81	00:01:21	13:30	56%
2023-03-17	43	00:01:19	07:10	30%
2023-03-18	0	-	-	0%
2023-03-19	0	-	-	0%
2023-03-20	0	-	-	0%
2023-03-21	0	-	-	0%
2023-03-22	0	-	-	0%
2023-03-23	0	-	-	0%
2023-03-24	1	-	00:10	1%
2023-03-25	80	00:00:38	13:20	56%
2023-03-26	81	00:07:10	13:30	56%
2023-03-27	0	-	-	0%
2023-03-28	0	-	-	0%
2023-03-29	2	01:25:06	00:20	1%
2023-03-30	0	-	-	0%
2023-03-31	0	-	-	0%
2023-04-01	3	06:59:02	00:30	2%
2023-04-02	0	-	-	0%
2023-04-03	2	10:08:41	00:20	1%
2023-04-04	1	-	00:10	1%
2023-04-05	1	-	00:10	1%
2023-04-06	2	02:45:27	00:20	1%
2023-04-07	0	-	-	0%
2023-04-08	2	01:38:05	00:20	1%
2023-04-09	0	-	-	0%
2023-04-10	2	00:58:33	00:20	1%
2023-04-11	4	03:02:20	00:40	3%
2023-04-12	0	-	-	0%
2023-04-13	1	-	00:10	1%
2023-04-14	3	11:00:37	00:30	2%
2023-04-15	0	-	-	0%
2023-04-16	0	-	-	0%
2023-04-17	2	12:24:23	00:20	1%
2023-04-18	2	10:05:27	00:20	1%
2023-04-19	1	-	00:10	1%
2023-04-20	0	-	-	0%

Figure 16: The lockout statistics simulate the account lockout if all login attempts would have been against the administrator account. This statistic is based on the delay between all login attempts on honeypot 1. Potential lockout statistics for the administrator account.

## 5.6.2 Comparison between honeypot 1 and 2

A total of eight IP addresses conducted login attempts on both honeypots 1 and 2 against the 'Administrator' account. The total number of attempts and delay between attempts for those addresses is shown in Figure 16.

IP-address	Honeypot 1		Honeypot 2	
	Total attempts	Average delay between attempts	Total attempts	Average delay between attempts
52.91.x.x (US - Ashburn)	27	00:09:43	13	00:21:03
52.208.x.x (IE - Dublin)	126	00:11:11	62	00:22:35
34.216.x.x (US - Portland)	63	00:06:13	32	00:12:13
34.213.x.x (US - Portland)	37	00:06:24	18	00:13:12
34.212.x.x (US - Portland)	32	00:06:19	31	00:06:31
34.212.x.x (US - Portland)	28	00:06:17	14	00:12:35
212.102.x.x (DE - Frankfurt)	54	00:00:11	2683	00:00:25
18.237.x.x (US - Portland)	66	00:06:21	32	00:12:55
<b>Sum</b>	<b>433</b>	<b>00:06:35</b>	<b>2885</b>	<b>00:12:41</b>

Figure 17: Comparison between honeypot 1 and 2. Time format HH:MM:SS. The addresses from US - Portland has relatively few login attempts with several minutes between each attempt. All these addresses follow the same pattern, which indicates that it could be the same actor who conducted the login attempts from different addresses.

## 5.7 Analysis of the impact of availability

A system that uses rate-limiting, for example, to limit usage or the number of incorrect logins, has a potential risk to become unavailable in the event of brute-force attacks. This can be both unintentional or exploited as a DOS attack by a malicious actor. In this study was honeypot 1 unavailable for approximately 1% of the total up-time. The average time between the lockout was 6 hours and 47 minutes. This means that the brute force attacks on the honeypot had little effect on the availability of the administrator account. However, this honeypot had the highest number of login attempts. The reason why availability was very little affected is that the login attempts were conducted with usernames other than administrator. If all login attempts were targeting the administrator, this honeypot would have been unreachable 56% of the day for some days (Figure 16).

## 5.8 Strategies and patterns

When performing an unsuccessful login attempt via RDP, you get information from the response whether an account is currently locked. This information allows you to adapt your attack strategy to avoid making login attempts against a locked account. When comparing the actors who conducted login attempts against both honeypots 1 and 2, this strategy has no clear differences. However, the data set is very small and is based on a small number of login attempts. But for all of those who tried to log in to both honeypots, the delay between attempts was higher on honeypot 2 than on honeypot 1 (Figure 17).

For IP addresses where several similar IP addresses have conducted login attempts, the session duration and delay between attempts had a similar pattern (addresses 37.46.x.x and 185.190.x.x in Figure 15). This indicates that an actor directs an attack from different machines that are configured in the same way, or several actors that share the same configuration at the time of the attack. The IP addresses 185.190.x.x and 5.181.x.x originated in different countries and the addresses were not registered as an exit node in a proxy network. When you study the patterns in Figure 15, I suspect that IP address 5.181.x.x are initiated from the same actor as IP 185.190.x.x. The login attempts were conducted at the same time, against the same honeypot, and had a similar pattern regarding the length of the session and average delay between attempts. But as this is only a suspicion, they have been marked with different colors.

The fact that an actor uses several attacking machines naturally leads to more login attempts being made during a period. But this is only effective if the target has not been configured with an account lockout policy or other rate-limiting restrictions. For these cases, the use of multiple attack machines rather contributes to the attack becoming more complex as they need to be synchronized to avoid locking the target account.

Regarding persistence and endurance, these have been relatively uncommon in this study. 5 IP addresses out of 165 were registered to complete more than 5.000 attempts and 4 IP addresses had recurring attempts for more than 5 days. By far the most common pattern was to make a small number of login attempts during one session. Those who conducted a few login attempts (tens) usually had a longer delay between attempts (minutes). Those who completed many (thousands) of attempts usually had a delay between 5-15 seconds.

## 5.9 Final analyzes and observations

It is clear that no actor does not have done any significant reconnaissance of the honeypot before the attack. This indicates that the attacks were initiated and conducted with a pre-configured machine. For example, it is easy to look up which country the honeypots are located in and which username could possibly correspond to the administrator account. Instead, most attempts have been conducted against the administrator user in English, Portuguese ("Administrador"), and French ("Administrateur"). In the case of a targeted attack, it can be assumed that login attempts should have been conducted against the local administrator username, which for Sweden is 'administratör'. However, no login attempts with this username have been registered.

## 5.10 Summary

This chapter presents data and analysis that are collected from the observations in this study. Initial attacks confirm that publicly connected computers are exposed to intrusion attempts within a very short time, which support previous studies. A few actors conducted the majority of the login attempts and approximately a third of the total number of actors only had one login attempt. The availability of the administrator account was not significantly affected. But it could have been significantly more affected if other strategies had been used by the actors.

## 6 Conclusions

During the 37 days, that the honeypots were exposed to attacks, a total of 121,648 login attempts took place from IP addresses originating in 40 different countries. It took less than 24 hours before the first login attempts were conducted despite the IP addresses not being marketed or published online. This result shows the importance of securing an RDP account before public access is granted. The most secure and effective protection against unauthorized access for an RDP account is to use a strong password. This makes it very difficult for a brute-force attack to succeed.

In this study, it was unfortunately not possible to see which passwords were tested, but the username used for each login attempt was recorded. The majority of all login attempts were against the administrator user. Several attempts have been made to log in with "administrator" translated into French and Portuguese. This indicates that the actor who carried out the attack did not carry out any reconnaissance of the object being attacked. This is further supported by the fact that the honeypots were located in Sweden and no login attempts were made with the username "administratör" which is the Swedish translation of "administrator".

There is a significant difference in the number of login attempts performed between honeypots 1 and 2 (Table 2). Where honeypot 1 had about 60,000 attempts and honeypot 2 had about 13,000 attempts. Although honeypot 1 had an account lockout policy enabled and the number of failed login attempts was limited by locking the account. This could be because a host configured with this type of security mechanism may be seen as a more attractive target. But since all the attacks were seemingly carried out at random, this theory is weakened. It is above all one actor with IP address 37.46.x.x (LU - Luxembourg) who conducted the majority of the login attempts against honeypot 1 (around 36,000 attempts) and 2 (around 9,000 attempts). However, no conclusions can be drawn about the reason why four times more attempts were made against honeypot 1 versus honeypot 2.

Trying to avoid the risk of attacks by changing the default port used for a remote service is not an effective security mechanism. Although it may take longer for actors to discover that the service exists and is available, the results have shown that it will be discovered and exposed to unauthorized login attempts (Table 1).

The availability may be affected when an account is applied with a locking mechanism, to prevent login attempts via brute force. This was studied via honeypot 1. The administrator account for this honeypot was locked out for a total of 9 hours and 40 minutes over 37 days due to failed login attempts. This means an impact on the availability of around 1% of the total uptime (section 4.5.1). Although this honeypot had approximately 61.000 failed login attempts, the majority of these were not directed to the administrator account, meaning that this account was not locked. If all login attempts had been directed to the user "administrator", availability would have been affected significantly more. On some days, this would have meant that the account would have been locked 56% of the day (Figure 16). The reason why most attempts were made against users other than "administrator" could be because an administrator is assumed to choose a stronger password than normal users and thus be easier to succeed with their brute force attack. It is clear that the actors who carried out login attempts against these honeypots tried to find users with very weak passwords. The perception according to the results is that they tested different usernames against a relatively short word list. If the actor failed with all login attempts, they did not continue or expand the attack but choose to give up. Figure 12 shows that there are only a few actors who conducted login attempts over several days, the majority have carried out their attempts during the same day and then given up. However, no clear conclusions can be drawn regarding this as an actor can use different IP addresses. A clear conclusion that can be drawn is that a host open to public RDP connections continues to be an attractive attack surface for malicious actors. With brute force attacks being the most common attack vector, the single most important and effective countermeasure is the use of strong passwords for all users.

# References

- [1] L. Bošnjak, J. Sreš, and B. Brumen. “Brute-force and dictionary attack on hashed real-world passwords”. In: *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. 2018, pp. 1161–1166. DOI: [10.23919/MIPRO.2018.8400211](https://doi.org/10.23919/MIPRO.2018.8400211).
- [2] Matt Weir et al. “Testing metrics for password creation policies by attacking large sets of revealed passwords”. In: *Proceedings of the 17th ACM conference on Computer and communications security*. 2010, pp. 162–175.
- [3] *Ncrack is a high-speed network authentication cracking tool*. <https://github.com/nmap/ncrack>. Accessed: July 28, 2023.
- [4] *Tool Hydra - GitHub*. <https://github.com/vanhauser-thc/thc-hydra>. Accessed: July 28, 2023.
- [5] Sanjib Sinha, Sanjib Sinha, and Karkal. *Beginning Ethical Hacking with Kali Linux*. Springer, 2018.
- [6] Paul A Grassi et al. “NIST special publication 800-63b: digital identity guidelines”. In: *National Institute of Standards and Technology (NIST)* (2017).
- [7] Donatella Firmani, Francesco Leotta, and Massimo Mecella. “On Computing Throttling Rate Limits in Web APIs through Statistical Inference”. In: *2019 IEEE International Conference on Web Services (ICWS)*. 2019, pp. 418–425. DOI: [10.1109/ICWS.2019.00075](https://doi.org/10.1109/ICWS.2019.00075).
- [8] Google. *Limits and Quotas on API Requests*. URL: <https://developers.google.com/analytics/devguides/reporting/core/v4/limits-quotas>. Accessed: July 28, 2023.
- [9] Stallings William Lawrie Brown. *Computer Security : Principles and Practice 4th edition*. Global Edition. New York: Pearson, 2018. ISBN: 9781292220611.
- [10] Yu Liu et al. “Account Lockouts: Characterizing and Preventing Account Denial-of-Service Attacks”. In: *Security and Privacy in Communication Networks: 15th EAI International Conference, SecureComm 2019, Orlando, FL, USA, October 23–25, 2019, Proceedings, Part II 15*. Springer. 2019, pp. 26–46.
- [11] Raj Badhwar. “Security Controls for Remote Access Technologies”. In: *The CISO’s Next Frontier: AI, Post-Quantum Cryptography and Advanced Security Paradigms*. Springer, 2021, pp. 99–104.
- [12] Microsoft - *Security guidance for remote desktop adoption*. <https://www.microsoft.com/en-us/security/blog/2020/04/16/security-guidance-remote-desktop-adoption/>. Accessed: July 28, 2023.
- [13] Microsoft. *KB5020282—Account lockout available for built-in local administrators*. URL: <https://support.microsoft.com/en-us/topic/kb5020282-account-lockout-available-for-built-in-local-administrators-bce45c4d-f28d-43ad-b6fe-70156cb2dc00>. Accessed: July 28, 2023.
- [14] Cortex. *Cortex Xpanse Attack Surface Threat Report*. Palo Alto Networks, 2022.
- [15] Matt Boddy, Ben Jones, and Mark Stockley. “RDP Exposed-The Threat That’s Already at Your Door”. In: *Sophos White paper. Sophos, Inc* (2019).
- [16] ZiHan Wang et al. “Automatically traceback RDP-based targeted ransomware attacks”. In: *Wireless Communications and Mobile Computing 2018* (2018), pp. 1–13.
- [17] Shodan.io. *Search engine for Internet-connected devices*. URL: <https://www.shodan.io>. Accessed: July 28, 2023.
- [18] Michael Gough. *Detecting and Protecting when Remote Desktop Protocol (RDP) is open to the Internet*. URL: <https://research.nccgroup.com/2021/10/21/detecting-and-protecting-when-remote-desktop-protocol-rdp-is-open-to-the-internet/>. Accessed: July 28, 2023.
- [19] A. Vollmer. “Attacking RDP”, How to Eavesdrop on Poorly Secured RDP Connections”. In: (2017).
- [20] Bret Padres Palo Alto Networks. *Incident Response and Data Breach Report. 2020*. URL: [https://www.paloaltonetworks.com/apps/pan/public/downloadResource?pagePath=/content/pan/en\\_US/resources/research/2020-unit42-incident-response-and-data-breach-report](https://www.paloaltonetworks.com/apps/pan/public/downloadResource?pagePath=/content/pan/en_US/resources/research/2020-unit42-incident-response-and-data-breach-report). Accessed: July 28, 2023.

- [21] Joshua B Gross and Mary Beth Rosson. “Looking for trouble: understanding end-user security management”. In: *Proceedings of the 2007 Symposium on Computer Human interaction For the Management of information Technology*. 2007, 10–es.
- [22] M. Sheltz S. Sage London M. Foust J. Chellis. *MCSE Windows Server 2003 Network Infrastructure Planning and Maintenance Study Guide: Exam 70-293*. Wiley, 2007, p. 735. ISBN: 9780471997924.
- [23] Laxmi B Pandya. “Cyber Security” When Working From Home”. In: *Emerging Trends in Commerce & Management* (), p. 36.
- [24] L Bošnjak, J Sreš, and Bosnjak Brumen. “Brute-force and dictionary attack on hashed real-world passwords”. In: *2018 41st international convention on information and communication technology, electronics and microelectronics (mipro)*. IEEE. 2018, pp. 1161–1166.
- [25] Daniel Miessler. *GitHub - SecLists/Passwords at master danielmiessler/SecLists*. URL: <https://github.com/danielmiessler/SecLists/tree/master/Passwords>. Accessed: July 28, 2023.
- [26] Blase Ur et al. “I added ‘!’ at the end to make it secure”: Observing password creation in the lab”. In: *Proc. SOUPS*. 2015.
- [27] Kenneth P Birman. “The next-generation internet: Unsafe at any speed?” In: *Computer* 33.8 (2000), pp. 54–60.
- [28] Tracey Caldwell. “Training—the weakest link”. In: *Computer Fraud & Security* 2012.9 (2012), pp. 8–14.
- [29] Zheng Yan et al. “Finding the weakest links in the weakest link: How well do undergraduate students make cybersecurity judgment?” In: *Computers in Human Behavior* 84 (2018), pp. 375–382.
- [30] John Lindström et al. “The need for improved alignment between actability, strategic planning of IS and information security”. In: *International ITA Workshop: 04/06/2008-06/06/2008*. 2008, pp. 14–27.
- [31] Virginie Goepp and Oscar Avila. “An Extended-Strategic Alignment Model for technical information system alignment”. In: *International Journal of Computer Integrated Manufacturing* 28.12 (2015), pp. 1275–1290.
- [32] *Information security, cybersecurity and privacy protection — Information security controls*. Standard. International Organization for Standardization, Feb. 2022.
- [33] Craig A Horne, Atif Ahmad, and Sean B Maynard. “A theory on information security”. In: (2016).
- [34] Rossouw von Solms and Johan van Niekerk. “From information security to cyber security”. In: *Computers Security* 38 (2013). Cybercrime in the Digital Economy, pp. 97–102. ISSN: 0167-4048. DOI: <https://doi.org/10.1016/j.cose.2013.04.004>.
- [35] Rui Zhuang et al. “A theory of cyber attacks: A step towards analyzing MTD systems”. In: *Proceedings of the second ACM workshop on moving target defense*. 2015, pp. 11–20.
- [36] Fabien Pouget, Marc Dacier, and Hervé Debar. “White paper: honeypot, honeynet, honetytoken: terminological issues”. In: *Rapport technique EURECOM* 1275 (2003), p. 09.
- [37] Lance Spitzner. *Honeypots: tracking hackers*. Vol. 1. Addison-Wesley Reading, 2003.
- [38] Iyatiti Mokube and Michele Adams. “Honeypots: Concepts, Approaches, and Challenges”. In: *Proceedings of the 45th Annual Southeast Regional Conference*. ACM-SE 45. Winston-Salem, North Carolina: Association for Computing Machinery, 2007, 321–326. ISBN: 9781595936295. DOI: [10.1145/1233341.1233399](https://doi.org/10.1145/1233341.1233399). URL: <https://doi.org/10.1145/1233341.1233399>.
- [39] nmap.org. *Network Mapper*. URL: <https://nmap.org/>. Accessed: July 28, 2023.
- [40] ip api.com. *IP Geolocation API*. URL: <https://ip-api.com/>. Accessed: July 28, 2023.
- [41] Marcin Nawrocki et al. “A survey on honeypot software and data analysis”. In: *arXiv preprint arXiv:1608.06249* (2016).

## 7 Appendix A

Three honeypots were used all within the same provider, location, and size of the instance.

Table 6: General honeypot details

Provider	Amazon Web Services (AWS)
Location	Europe Stockholm (eu-north1)
Number of honeypots	3
Instance name	t3.micro
Hardware	2 vCPU, 1 GB RAM

### Honeypot details

Open ports in the firewall which could be discovered publicly online.

Honeypot 1 - Account Lockout Policy On (HP1): 3389

Honeypot 2 - Account Lockout Policy Off (HP0): 3389

Honeypot 3 - None standard RDP port (HP2): 6781

## 8 Appendix B

Log entry details and details about data collection elements.

Level	Date and Time	Source	Event ID	Task Category
Information	2023-03-17 03:24:25	Microsoft Windows security auditing	4625	Login
Information	2023-03-17 03:18:16	Microsoft Windows security auditing	4625	Login
Information	2023-03-17 03:10:06	Microsoft Windows security auditing	4625	Login

Field	Value
SubjectUserSid	S-1-0-0
SubjectUserName	
SubjectDomainName	
SubjectLogonid	0x0
TargetUserSid	S-1-0-0
TargetUserName	VINAYAK
TargetDomainName	
Status	0xc000006d
FailureReason	96%2313
SubStatus	0xc0000064
LogonType	3
LogonProcessName	NtLmSsp
AuthenticationPackageName	NtLm
WorkstationName	
TransmittedServices	
LmPackageName	
KeyLength	0
ProcessId	0x0
ProcessName	
IpAddress	34.212.181.133
IpPort	0

Figure 18: Data collected in a log entry from Windows event log

The highlighted elements show that the event is a failed login attempt (log id 4625). The username provided is "VINAYAK". Status and reason indicate if the login attempt is denied due to wrong credentials or if the account is locked. IP address shows the address from where the login attempt is made.

These data elements are collected and stored in the SQL database.

### SQL Table 1 - Event log data

Created On - Time stamp when the record has been created.

Event Index - Unique id of the log entry.

Time Generated - Time stamp when the log entry has been created in the event log database.

Machine Name - Host name of the honeypot.

Workstation Name - Name of the workstation from the client that conducts a login attempt.

Failure Reason - Reason code of the failed login generated by Windows.

Status - Status code of the failed login generated by Windows.

Sub status - Sub status code of the failed login generated by Windows.

Logon Type - Method id of logon (interactive, network, service, etc).

Process Id - Id of processes related to other processes.

Username - String used as username for the login attempt.

Ip Address - Client IP address (IPv4) from login attempt.

Ip Port - Client port used for the login attempt.

### SQL Table 2 - IP meta data

Created On - Time stamp when the record has been created.

Ip Address - IPv4 address.

Country Code - Code of the country (2 characters).

Region Name - Name of the region.

City - Name of the city.

Isp - Name of the IP provider.

Org - Name of the organization providing the IP.

Proxy - Boolean if the IP belongs to a proxy or VPN network.

Hosting - Boolean if the IP is related to hosting, co-located, or data center.

## 9 Appendix C

The number of login attempts by the origin of the IP address. The table below shows the origin of the top 20 with the highest number of login attempts.

Table 7: Categories and metrics used in honeypot data analysis

Country code	City	Number of login attempts	Is exit node*
LU	Luxembourg	45191	Yes
PH	Mandaluyong City	19029	No
AX	Mariehamn	15288	No
KR	Gimhae	9298	No
DE	Frankfurt am Main	8787	Yes
SC	Victoria	8614	No
RU	Moscow	5036	No
CN	Tianjin	4742	No
TH	Chatuchak	1626	No
US	Portland	812	No
KR	Gangseo-gu	376	No
UA	Kyiv	356	No
IN	New Delhi	283	Yes
US	Ashburn	234	No
US	Denver	221	No
IE	Dublin	216	No
CN	Beijing	188	Yes
JP	Tokyo	169	No
US	Boydton	141	No
IN	Mumbai	123	No

\*Exit node refers to a known IP address that is associated with a proxy network.

# 10 Appendix D

Top 50 most used usernames for login attempts across all honeypots.

Table 8: Top 50 username

Username	Number of login attempts
ADMINISTRATOR	45817
Administrador	6615
EC2AMAZ-BKRO3TR *	5258
ADMIN	662
USER	300
Test	180
SCANNER	177
USER1	123
BACKUP	96
PC	96
SCAN	94
SUDHIR	72
ADM	69
OFFICE	65
POSTGRES	63
OFICINA	61
LOCALADMIN	60
DELL	60
ITADMIN	60
REMOTO	60
UTILISATEUR	59
VPN	59
LENOVO	58
UTENTE	57
SCANS	56
MASTER	55
WEB	55
SHARE	55
SALES	54
STAFF	54
SERVER	54
ADMINISTRATEUR	53
POS	53
FTP	52
RECEPTION	51
JEFF	48
SUPPORT	47
BEHEERDER	47
USER01	47
TEMP	47
INFO	47
PRINTER	47
ATELIER	46
ROOT	46
HP	45
MANAGER	45
CS	45
TERMINAL	44
HOME	43
SECRETARIAT	43

\* This corresponds to the machine- / host name for one of the honeypots.