Master of Science in Software Engineering
May 2022

A comparison between on-premise and cloud environments in terms of security

With an emphasis on Software-as-a-Service & Platform-as-a-Service

Oliver Byström

Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden

This thesis is submitted to the Faculty of Computing at Blekinge Institute of Technology in partial fulfilment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 20 weeks of full time studies.

The authors declare that they are the sole authors of this thesis and that they have not used any sources other than those listed in the bibliography and identified as references. They further declare that they have not submitted this thesis at any other institution to obtain a degree.

Contact Information:
Author(s):
Oliver Byström
E-mail: oliver.bystrom@outlook.com

University advisor:
Dr. Emiliano Casalicchio
Department of Computer Science

Faculty of Computing
Blekinge Institute of Technology
SE–371 79 Karlskrona, Sweden

Internet : www.bth.se
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

Background: Cloud- and on-premise environments have been compared in terms of security several times. Many of these comparisons based their assessments on qualitative data rather than quantitative metrics. Some recent articles have considered comparing environments by using quantitative data. These methodologies are often complicated and based on incident simulations that might not be relevant in a real-life scenario. Therefore it could be troublesome for a company to evaluate and compare two environments before deciding which environment they would prefer in terms of security. Before making a decision to use a specific environment to host service, it is decisive to know if that environment has been a target for recent cyberattacks. Unfortunately, this data is not available to the public.
Objectives: This study aims to provide the reader with an overview of the environmental aspects of the victims of recent cyberattacks. It will reveal what environment cybercriminals have targeted the most. The study will also propose a methodology to compare two environments to each other based on quantitative measurements. The measurements were based on cybersecurity metrics that quantified the threats in each environment.
Methods: A structured literature- and dataset review was conducted to find how much each environment had been exposed to cybersecurity incidents. Several expert interviews were held to help explain the findings made in the reviews. A threat analysis was used as the foundation for the proposed comparison methodology. A case study of a recent environment migration was used to test the proposed comparison methodology.
Results: The results show that on-premise environments have been more exposed to cybersecurity incidents during recent years than cloud environments. The proposed methodology showed that the cloud environment was the preferred choice in the conducted case study.
Conclusions: In recent years, cloud environments have been the preferred choice in terms of security as long as the cloud consumer takes heed to best practices. There is a knowledge gap when it comes to cloud environments. It has been the same for both cloud consumers and cybercriminals. However, according to recent threat reports, cybercriminals have started to improve. Therefore there will likely be more cloud-related incidents in the future. It was determined that the proposed methodology could represent the security posture of each environment. However, a decision should not be based entirely on this methodology because it has not been tested on a large scale.

Keywords: on-premise, cloud, comparison, cybersecurity, metric

Bakgrund: Moln- och on-premise-miljöer har jämförts vad gäller säkerhet flera gånger. De flesta jämförelser baserade sina bedömningar på kvalitativ data snarare än kvantitativa mått. Några nya artiklar har jämfört miljöer med hjälp av kvantitativ data. Dessa metoder är ofta komplicerade och baserade på incidentsimuleringar som kanske inte är relevanta i ett verkligt scenario. Därför kan det vara besvärligt för ett företag att utvärdera och jämföra två miljöer innan de bestämmer sig för vilken miljö de skulle föredra vad gäller säkerhet. Innan en miljömigrering är det avgörande att veta om den miljön har varit ett mål för de senaste cyberattackerna. Tyvärr är denna information inte tillgänglig för allmänheten.
Syfte: Denna studie syftar till att ge läsaren en översikt av miljöaspekterna hos offren för de senaste cyberattackerna. Det kommer att avslöja vilken miljö cyberkriminella har riktat sig mest mot. Studien kommer också att föreslå en metodik för att jämföra två miljöer med varandra baserat på kvantitativa mått. Mätningarna baserades på cybersäkerhetsmått som kvantifierade hoten i varje miljö.
Metod: En strukturerad litteratur- och datasetgranskning genomfördes för att ta reda på hur mycket varje miljö har varit utsatt för cybersäkerhetsincidenter. Flera expertintervjuer hölls för att förklara resultaten som gjorts i granskningarna. En hotanalys genomfördes för att ge underlag för den föreslagna jämförelsemetodiken. Jämförelsemetoden testades i en fallstudie av en nyligen genomförd miljömigrering.
Resultat: Resultaten visar att on-premise miljöer har varit mer utsatta för cybersäkerhetsincidenter under de senaste åren än molnmiljöer. Den föreslagna metoden visade att molnmiljön var det föredragna valet i den genomförda fallstudien.
Slutsatser: Under de senaste åren har molnmiljöer varit det föredragna valet när det gäller säkerhet så länge som molnkonsumenten tar hänsyn till bästa praxis. Det finns en kunskapslucka när det kommer till molnmiljöer. Det har varit samma sak för både molnkonsumenter och cyberkriminella. Men enligt de senaste hotrapporterna har cyberkriminella börjat kommit ikapp. Därför kommer det troligen att finnas fler molnrelaterade incidenter i framtiden. Det fastställdes att den föreslagna metoden kunde representera säkerheten för varje miljö väl. Ett beslut bör dock inte baseras helt på denna metodik eftersom den inte har testats i stor skala.

Nyckelord: on-premise, molnet; jämförelse, cybersäkerhet, mått .

Acknowledgments

I would like to thank Giovanni Abbiati, Silvio Ranise, Antonio Schizzerotto and Alberto Siena for creating the dataset that was used in this study. I also want to thank Asurgent AB for supplying this thesis with a real life case as well as expert opinions when conducting the case study. I am also very thankful to the experts from BTH, Cloud Security Alliance and Microsoft that gave their time to participate in the interviews of the study. Last but not least, I want to thank my supervisor, Dr. Emiliano Casalicchio, for his support and feedback during this thesis.

Contents

Abstract **i**

Acknowledgments **v**

1 Introduction **1**
1.1 Cloud- and on-premise environments 4
1.2 The aim and objective of the thesis 4

2 Related Work **7**
2.1 The research gap 8

3 Method **11**
3.1 Choice of methodologies 11
3.2 Structured literature review 12
3.2.1 Database selection 14
3.2.2 Incident inclusion criteria 14
3.3 Incident dataset review 15
3.3.1 Sample criteria 15
3.3.2 Sampling strategy 17
3.3.3 Third party source validation 18
3.4 Interviews 19
3.4.1 Finding participants 19
3.4.2 Selection and motivation of questions 19
3.4.3 Ethical considerations 21
3.5 Threat modeling 22
3.5.1 Scenario and motivation 22
3.5.2 Threat model process 22
3.6 Cybersecurity metrics 22
3.6.1 Metric weights 23

4 Results **27**
4.1 Structured literature review 27
4.2 Incident dataset review 27
4.3 Interview sessions 29
4.3.1 Statement 1 (Answers) 30
4.3.2 Statement 2 (Answers) 31
4.3.3 Statement 3 (Answers) 32
4.3.4 Statement 4 (Answers) 34

4.3.5 Statement 5 (Answers) . . . 35
4.3.6 Statement 6 (Answers) . . . 36
4.3.7 Statement 7 (Answers) . . . 37
4.3.8 Statement 8 (Answers) . . . 38
4.4 Threat modelling . . . 39
4.4.1 Architecture overview . . . 39
4.4.2 Decomposition of environment (On-premises) . . . 39
4.4.3 Decomposition of environment (Cloud) . . . 43
4.4.4 STRIDE analysis . . . 46
4.5 Cybersecurity metrics . . . 46
4.5.1 Measurements . . . 51

5 Analysis and discussion **55**
5.1 Incident history and interviews . . . 55
5.1.1 Defense capabilities . . . 56
5.1.2 Incident transparency . . . 56
5.1.3 Lack of competence . . . 57
5.1.4 Causes for cybersecurity incidents . . . 58
5.1.5 Physiological factors of environments . . . 58
5.1.6 Cybercriminal preference . . . 59
5.1.7 Future developments . . . 59
5.2 Case study . . . 60
5.2.1 Validity of measurements . . . 60
5.2.2 Validity of the methodology . . . 63
5.3 Threats to validity . . . 63
5.3.1 External validity . . . 63
5.3.2 Internal validity . . . 64
5.3.3 Construction validity . . . 64
5.3.4 Conclusion validity . . . 64

6 Conclusions and Future Work **67**
6.1 RQ1 . . . 67
6.2 RQ2 . . . 67
6.3 RQ3 . . . 68
6.4 Concluding remarks . . . 69
6.5 Future works . . . 69

References **71**

A Supplemental Information **83**
A.1 Data processing and sampling . . . 83

List of Figures

1.1 The responsibilities between the different environment models based on information given in the NIST 800-144 specification [59]. 2

3.1 The process of the structured literature review. The numbers are the result of the search, they are explained in the result chapter 4.1. . . . 13

4.1 The data flow diagram of the on-premises environment. 40
4.2 The data flow diagram of the cloud environment. 41
4.3 The STRIDE analysis of the components that are present in both environments. $Ex = Entity(x)$, $Sx = System(x)$, $Dx = Datasource(x)$, $Fx = Flow(x)$. 47
4.4 The number of threats from the STRIDE analysis that each metric represents. 51

List of Tables

3.1 The features of the dataset that was reviewed in this study. 15

4.1 Incidents (37) identified in the structured literature review meta-analysis. 28
4.2 Incidents (60) sampled from the compiled dataset together with additional sources and the result from the incident analysis 29
4.3 The metrics that were used to compare the two environments. 46
4.4 The mapping between threats defined in the STRIDE threat analysis and their corresponding metric IDs. 50
4.5 The measured values for each cybersecurity metric defined in table 4.3 52
4.6 The measurement results for the cloud environment. 53
4.7 The measurement results for the on-premises environment. 54

Acronyms

APT Advanced Persistent Threat. 38, 59, 60

BLI The Data Breach Level Index. 15

CSP Cloud Service Provider. 1, 3, 4, 32–34, 36, 38, 55–61, 63, 64

DDoS Distributed Denial of Service. 46, 61

DFD Data Flow Diagram. 22, 61, 67

IaaS Infrastructure-as-a-Service. 1

IT Information Technology. 14

ITRC The Identity Theft Resource Center. 15

MFA Multi Factor Authentication. 23, 60

MSP Managed Service Provider. 3, 4, 20, 33, 58, 59

NIST National Institute for Standards and Technology. 7

OT Operation Technology. 14

PaaS Platform-as-a-Service. 1, 4, 20–22, 27, 30, 55–59, 61

PIM Privileged Identity Management. 61

PRC Privacy Rights Clearinghouse. 15

SaaS Software-as-a-Service. 1, 4, 7, 22, 27, 56, 58

SLA Service Level Agreement. 56, 61

SME Subject Matter Expert. 11, 12, 48–50

Chapter 1

Introduction

The cloud phenomenon has become increasingly popular in recent years. The cloud industry allegedly grew from a $70 billion evaluation in 2015 to more than $203 billion at the end of 2020 [21]. Both private citizens and companies use cloud services today. Recently government agencies have joined in to utilize cloud services [111]. One recent example of a large government agency that uses cloud technology is the Pentagon defense agency [79]. It can be extra vital for government agencies to make proper infrastructure comparisons, espectially before transitioning to a CSP they intend to use to manage confidential data. If such infrastructure were to be compromised by a threat actor, it could be a matter of national security [59].

Today there are many public CSPs that a consumer can choose. Examples include Google Cloud Platform, Microsoft Office365/Azure, Oracle, Amazon Web Services, etc. [12, 13, 90, 111]. The most common models that are offered by these cloud providers are:

- Software-as-a-Service - Both software and the infrastructure is handled by the CSP.
- Platform-as-a-Service - The infrastructure and operating system is handled by the CSP while applications hosted on this environment are handled by the cloud consumer.
- Infrastructure-as-a-Service - The hardware is managed by the CSP while operating systems and applications are managed by the cloud consumer.

Figure 1.1 provides an overview of these shared responsibilities. Everything that falls on the responsibility of the cloud consumer has to be updated and maintained by the cloud consumer. The same applies to the CSP. Amazon Web Services define in their shared responsibility model that the CSP has the security responsibilities OF the cloud while the cloud consumer has the security responsibilities IN the cloud [14].

While these different models provide different functionalities to the cloud consumer, they also vary in security and risk management. Because a cloud consumer has the sole responsibility of a virtual machine running as a IaaS instance, they do not only have to patch security vulnerabilities inside the applications they use but also the operating system. Using a SaaS alternative frees a cloud consumer of such concerns. Instead, other issues present themselves in the forms of multi-tenancy factors and CSP misconfigurations [59, 111]. A PaaS solution is somewhere in between where the cloud consumer is responsible for any security patches that apply to the software or application that runs on top of the platform operating system. The CSP

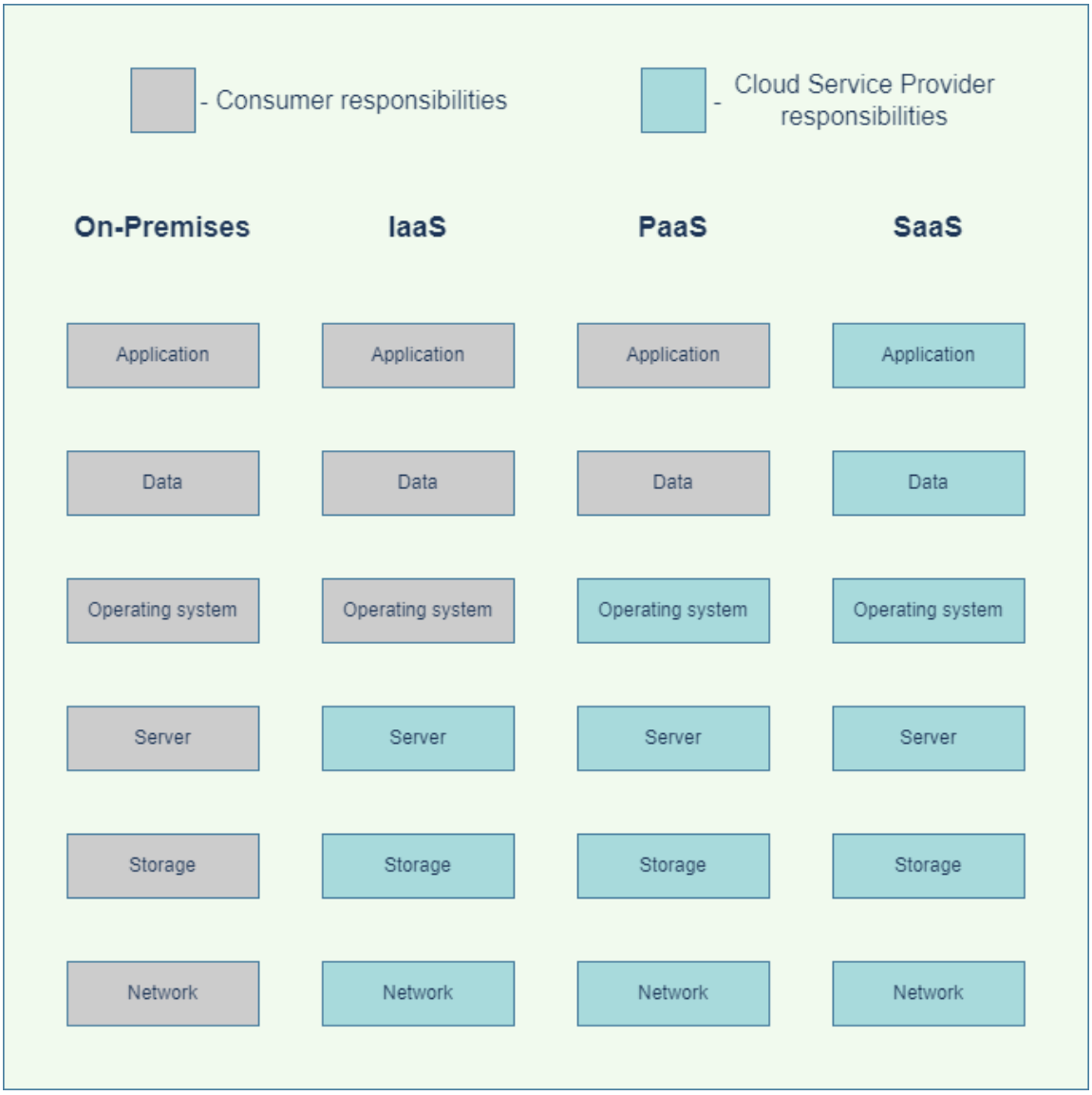

Figure 1.1: The responsibilities between the different environment models based on information given in the NIST 800-144 specification [59].

is then responsible for applying security patches to the underlying operating system that hosts the application.

When using an on-premise solution where everything is managed in-house or outsourced to a third party MSP, a company needs to apply the security patches for the entire technology stack (hardware to application) themselves. Unfortunately, this is still a big concern today because threat actors still leverage vulnerabilities that should already be patched [32].

Even though patch management is a big part of the security of an IT environment, it does not cover the whole picture. Many papers discuss issues related to cloud security. Despite this, security concerns still exist in the cloud. One concern organizations have regarding public clouds is their multi-tenancy nature [59, 111]. Today there are guidelines a cloud consumer can use to find a CSP that will satisfy their needs. These guidelines exist in publications from the National Institute for Standards and Technology (NIST), Cloud Security Alliance (CSA), and independent researchers [30, 59, 84]. It is one action a current or future cloud consumer should make before migrating a system or an environment to the cloud.

However, there might be scenarios where a consumer cannot use these guidelines. One of these scenarios is when a consumer wants to compare two alternatives using quantitative measurements. There is much information about risk management, but there are no methods documented that a consumer can use to see the impact of a transition to the cloud concerning their current security posture.

One scenario could be a company that uses a booking system where employees can schedule resources for their customers. The service has become slow because the company has grown with the number of customers that use the system. The issues and incidents regarding the performance of the booking system are starting to become more frequent and take up more and more time from the in-house IT staff. Therefore, the company is planning an infrastructure change. The senior management is considering either moving the in-house service to the cloud or expanding the current solution with additional on-premise resources. They favor the first option because a cloud alternative will be a long-term solution where they can either increase or decrease the capacity of the service. The CTO warns the senior management about the qualitative risks of moving the service to a public cloud. One of these risks is that their booking system would run on the same hardware as another tenant. A security advisor highlights that running the service in the cloud will still be more secure than running it on-premises. The CTO and the security advisor agree that the senior management should make a decision that will satisfy the security needs of the company in the long run. To make this decision, the senior management would need data to compare the two environments with each other in regards to security.

This thesis aims to help in this kind of situation by providing a methodology to compare the security aspects of two different environments. The methodology is demonstrated on two real-life samples to illustrate how the method can be applied to an arbitrary pair of environments.

The thesis will also include an overview of what environment cybercriminals have targeted the most to find trends based on environmental data. It can be interesting to see if an environment stands out in recent incidents before migrating to that environment.

The remainder of the report has the following structure: Chapter 2 will cover

previous work on on-premises and cloud comparisons and risk assessment methodologies. Chapter 3 describes the methods used to reach the objectives and answer the research questions. Chapter 4 will present the results of the thesis and relevant data. Chapter 5 contains an analysis of the results, the validity of the results, and the study itself. Finally, chapter 7 concludes the paper and lists some topics for future works.

1.1 Cloud- and on-premise environments

One definition of a cloud is that it is a service that gives a customer (cloud consumer) access to on-demand computational resources that they can provide themselves with minimal or no interaction from the CSP [73]. In this thesis, a public cloud is defined as a cloud that is available to the general public. This definition aligns with existing definitions done by NIST and CSA [12,73,109]. Anyone can sign up and deploy their services on a public cloud. The hardware resources are normally shared between the cloud consumers as well.

In this thesis, an on-premises environment is an environment that is not a cloud environment. Therefore, for a user to provision more resources to their service or infrastructure that runs inside an on-premises environment, the owner of that infrastructure needs to take action. One example of this is when a company uses an MSP to host their services, or they host their services internally with their own equipment.

The main distinction between the environments is that on-premise environments normally use a exclusive responsibility model while a cloud environment uses a shared responsibility model. The different responsibility models are illustrated in figure 1.1. The responsibilities range to maintenance of the environment to securing the assets inside it.

1.2 The aim and objective of the thesis

The aim of this study is to compare security aspects of on-premises and public cloud environments. The thesis has two main objectives that will complement each other to reach this goal.

The first objective is to investigate the history of disclosed cyberattacks and determine if the vulnerability that the threat actor exploited was a part of an on-premises or cloud environment. It will provide an overview of the past and current security relationship between these two environments. It will reveal which of the two environments has been more targeted by cybercriminals than the other. The thesis will be limited to incidents that occurred after 2008 because it was around this time cloud services started to increase in popularity in the commercial market [12,109,115].

The second objective is to propose a methodology to compare two environments against each other based on threat modeling. The threat models will be used to define cybersecurity metrics. These metrics will represent quantitative estimations of how well a particular environment fairs for a specific threat. The methodology was evaluated using two environments which were the products of a recent environment migration. One of these environments was an on-premise environment and the other a cloud environment with PaaS and SaaS services.

Considering the information and arguments that has been supplied so far, the following research questions were defined for this thesis:

- **RQ1** - To what extent have cybercriminals exploited on-premises environments compared to cloud environments between 2008 and 2021?
- **RQ2** - Which cybersecurity metrics are appropriate to use when comparing on-premises and cloud (Paas/SaaS) environments?
- **RQ3** - How are these cybersecurity metrics affected in an infrastructure migration from an on-premise environment to a cloud environment?

Chapter 2

Related Work

As mentioned previously, there has been a lot of research comparing on-premises and cloud security. Most of these studies rely on qualitative rather than quantitative data. The problem with qualitative data is that you cannot quantify the risk. Only using qualitative data can lead to misinterpretations where a consumer chooses a specific environment over another. Without quantifying risks, a consumer could exclude an environment based on rumor or fear.

The NIST 800-144 paper is one classic example of an on-premise and cloud risk comparison. In this paper, NIST released research and documentation on what to consider when an organization utilizes cloud computing [59]. Even though the publication is ten years old, the challenges presented in the study still hold to this day. NIST also briefly discusses the problems of moving old systems into the cloud. However, comparisons and recommendations are based on qualitative assessments and do not provide any methodology to assess risks in each environment using quantifiable means.

L. F. Plá, N. Shashidhar, and C. Varol make a comparison between cybersecurity solutions deployed on-premises and in the cloud [90]. They conclude that on-premise environments are best suited to consumers as long as they have the resources to maintain the systems in that environment. At the same time, they also mention that small and medium-size organizations usually do not have these resources. Therefore, they conclude smaller organizations are often better off using a cloud cybersecurity solution.

T. August et al. discuss the concept of directed and undirected attacks/risks when comparing on-premises and cloud SaaS environments [13]. A directed attack-/risk is when an actor focuses on finding a vulnerability in a system running a specific vulnerable software version. The authors argue that a SaaS cloud implementation where there is only one software version available is a classic example where a consumer should consider directed risks. They define an undirected risk as when an actor focuses on multiple systems and tries to find one vulnerable version. According to the authors, undirected risks often occur in on-premise environments with many software versions. The authors also mention that small threat actors are less likely to use directed attacks because these attacks might require a zero-day exploit. Again the authors base their comparison on qualitative data even though they put an interesting perspective on how different types of attacks might target different environments.

M. Ouedraogo and H. Mouratidis argue there is evidence that cybercriminals sometimes favor cloud providers as their target because if they are successful, it

could bolster the amplitude of their attack [84]. The authors argue that this is one advantage an on-premise environment has over a cloud environment. Their main argument is that detecting malicious activity in cloud applications is hard. If a cybercriminal has successfully infiltrated a cloud environment, their actions would be undetected by ordinary anti-virus solutions [84,90]. They also bring up limitations regarding cloud forensics when conducting incident response in the cloud. Again, the authors of this study compared the environments using qualitative data.

There are some studies regarding quantitative cybersecurity risk assessments. L. Allodi and F. Massacci discuss in their work the limitations of qualitative risk assessments and why a quantitative method is needed to make a more realistic risk assessment [10]. Their study proposes a method to determine the likelihood of an attack in a particular environment using various probability measurements. These measurements originate from the environment and the resources the adversary supposedly has. Their method can be complicated when used to compare two environments. Their methodology also requires both environments have IDS/IPS solutions in place to gather this data.

A. Montemaggio et al. explore in their work a methodology to measure the security of a self-protective system [74]. They conducted an experiment where they instructed a group of subjects to identify intrusions created by automatic tools. The metric in this report was the team's performance in handling the incidents in each environment. One problem with this approach is that the experiment does not necessarily represent an authentic incident. Another issue is that incident data from both environments is needed. A company can't improve these measurements if they have not experienced any attacks. Metrics based on data not related to specific events will make it possible to measure the current security stance. Therefore the methodology in this thesis will focus on environment data rather than event-based data.

A. Zieger et al. base their work on the "Time to Compromise" metric, which they also improve by making modifications to the measurement algorithm [122]. The method they present can measure the security of a system without detailed information. However, all vulnerabilities from all components need to be defined when applying this method. Because of this, their methodology will quickly become cumbersome if the environment is of considerable size. Even though not much detail is needed to find vulnerabilities for each component inside an environment, it can take considerable time. Because the time allocated to this thesis is limited, this study had to use another methodology.

Because none of the related articles propose a methodology that fits this study, this thesis will propose a novel methodology that measures the security of the components. Instead of calculating the probability of an attack, this novel methodology will measure the hardening of the environments. The theory is that this will provide more value and is more intuitive than the other methods.

2.1 The research gap

Despite discussing the risks and challenges of each environment, the related articles never measure the differences. Existing methodologies that calculate quantitative

risks are complicated and demand specific environment settings.

Therefore, it can be difficult for a cloud consumer to know the security impact of migration if they only have qualitative data to back their decision. A cloud consumer should quantify risks between two environments before migrating from one environment to the other. To see how values from cybersecurity metrics differ when comparing an on-premises environment to a cloud alternative is one way of doing this.

Because many security concerns in the previous studies involved multi-tenancy factors, this thesis will focus on public clouds where multiple cloud consumers could use the same hardware.

This thesis will test a more simple form of comparison based on threat model analysis. After reviewing several threat model strategies, the STRIDE threat model methodology became the preferred choice. The arguments of this choice are in the Method section.

A simple methodology would allow a cybersecurity specialist to present the security impact of an enviroment migration to senior management. It is vital for the executives that make these decisions to understand this despite not having any previous experience in cybersecurity.

There is little information regarding the environmental details in cybersecurity incidents available to the public. Information such as this could help researchers to see incident trends regarding a specific environment type [1, 84]. It would also reveal how cloud and on-premises risks have evolved.

Chapter 3

Method

Three different methodologies were used to reach the first objective. These include a structured literature review, a dataset review, and interviews with several SMEs.

A case study was conducted to reach the second objective. Because the objectives and the different methodologies support each other, they were conducted in the following order:

1. Structured literature review
2. Dataset review
3. Interviews
4. Case study

The thesis started with a structured literature review to find trends regarding environments of disclosed cyberattacks. The idea is that published articles will have accurate data on the incidents. Scientific papers are reliable as long as they have been peer-reviewed. Therefore, a structured literature review became a natural choice for this thesis. The purpose of the dataset review was to complement the results from the structured literature review. The result from both reviews will answer RQ1.

The interviews were conducted after the literature and dataset review because the interviews will make sense of the result. Therefore the reviews must take place before the interviews. The results from the interviews will give a more in-depth and updated view regarding cloud security from sources close to the industry. Therefore the interviews serve two purposes in this thesis. The first is making sense of the results from the reviews. The second purpose is to gain a nuanced view of what separates both environments regarding security which will help answer RQ2.

The case study is the last method used in this thesis. Information from the other methodologies will help define relevant cybersecurity metrics to compare the different environments. The purpose of the case study and the second objective was to answer RQ3.

3.1 Choice of methodologies

A structured literature review was chosen instead of a systematic literature review mainly because the needs of a systematic literature review cannot be satisfied in this study [65]. One requirement is that at least two people are involved in a systematic

literature review [66]. Because a single individual wrote this thesis, it is impossible to satisfy this requirement.

Another reason is that a systematic literature review demands a considerable amount of time to perform. Because the time allocated to this study is finite, it is not possible [66]. It would be a risk to rely solely on the results from a literature review to satisfy the first objective. Therefore it was not possible to put too much time into this methodology.

There is also a risk the scientific papers do not describe incidents in enough detail to distinguish any environmental parameters. There is also a risk that the scientific papers discuss only a handful of incidents. Therefore an incident dataset was included in the study as well.

Another literature review could have explained the results instead of conducting interviews. However, interviews provide information from SMEs that are active inside the cybersecurity industry. Given how fast the cybersecurity industry changes, information gathered in a literature review could be outdated. Another reason is that few academic papers found during the pre-study of this thesis described the differences in terms of modern security practices. Therefore interviews with industry experts would bring the most value.

A survey is another methodology that could have been used instead of a series of interviews. However, it would have been impossible to discuss the subjects with the participants. A discussion with the participants will reduce the bias of the answers and provide more value to the result [16].

Instead of a case study, an experiment would also be a viable method. The problem with choosing an experiment is that it would take longer to set up the conditions to perform it [16]. Creating multiple environments is one of those conditions. It would take a lot of time because these environments would have to be realistic while also being different from each other.

A case study would also bring more value to this thesis because the selected sample would be the product of a real-life migration. It provided a unique condition where the cloud architects that made the migration could give feedback during the case study. The cloud architects also evaluated the cybersecurity metrics defined in the study. It created a unique opportunity to gain feedback on the metrics from industry professionals who knew the environments well.

3.2 Structured literature review

The structured literature review will achieve its objective by transforming qualitative data into quantitative data. Meta-analysis was performed on scientific articles to find this data. One example is that an incident started inside an on-premises exchange server. The meta-analysis will then list that a lack of security in an on-premise environment was the cause of this incident. The entire process is illustrated in figure 3.1.

The review process started by using keywords to make a series of unstructured searches in different databases to find relevant keywords. A unstructured search can also be thought of as a free text search without any inherent structure [65]. The keywords were used to create search blocks to make the searches more precise. By

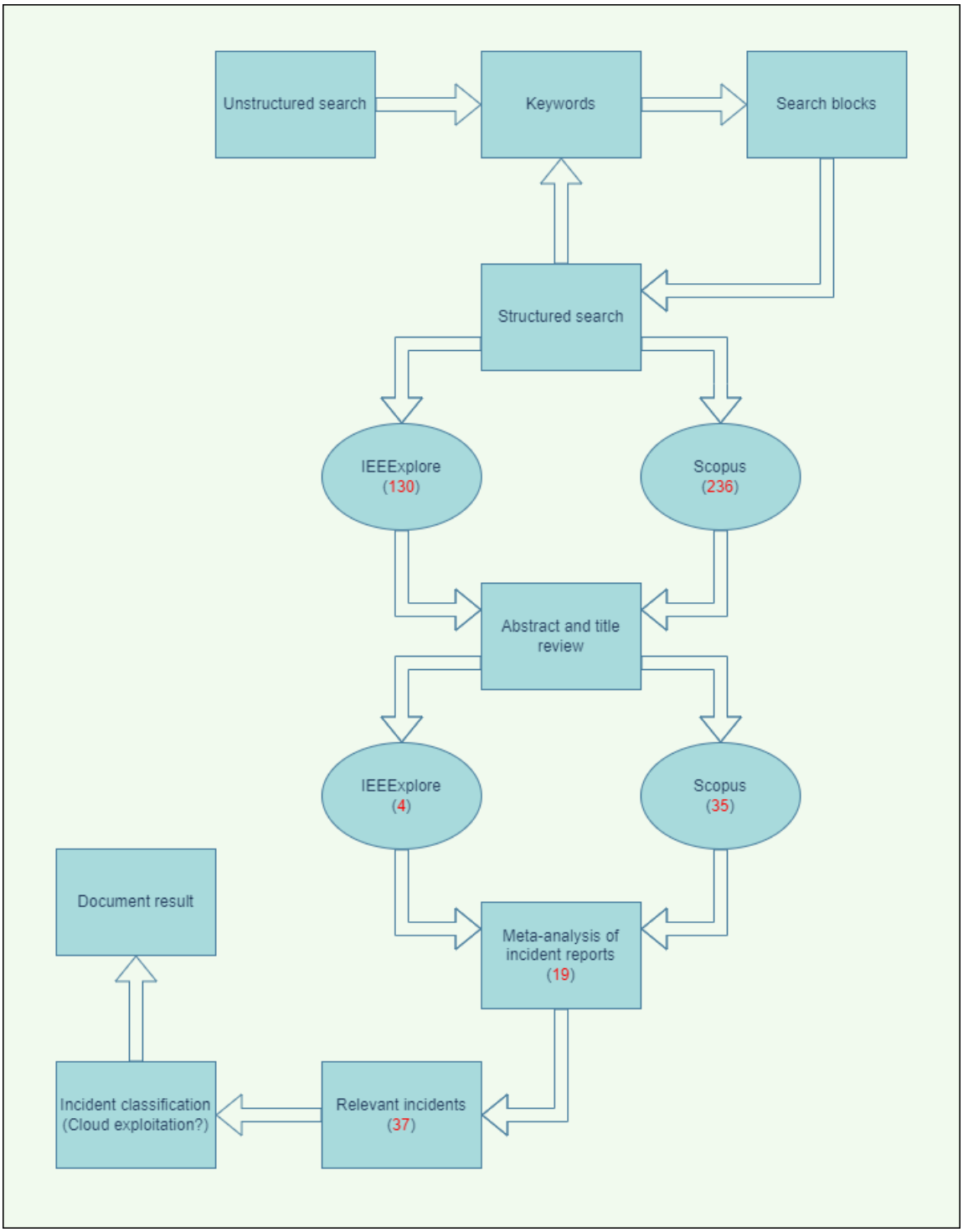

Figure 3.1: The process of the structured literature review. The numbers are the result of the search, they are explained in the result chapter 4.1.

adding more keywords to the search blocks, the searches became more and more structured. All incidents that occurred before 2008 and after 2021 were excluded to satisfy the requirement formulated in RQ1. The process was aborted when the number of articles reached the desired limit. The limit was set to 250 articles per database because this was considered to be enough to satisfy the scope while not taking up too much time from the study.

The result of the process was a series of search blocks that became the structured search query.

The abstracts of the articles left from the structured search query were then further processed by reading through their abstracts more thoroughly. This process filtered out any unrelated articles.

The next step was finding out which articles researched incidents. This information was obtained by doing a meta-analysis of the article. The meta-analysis was done by searching for the incident sections. If an article discussed an incident using insufficient information, it was ignored. It was also ignored if it proved to be irrelevant to the study.

The last step of the process involved classifying and documenting the incidents that were relevant to the study.

3.2.1 Database selection

Five different databases, *Scopus*, *Summon@BTH*, *IEEExplore*, *Google Scholar* and *Springer*, were evaluated in the unstructured search to see which one had the most relevant content. After performing several unstructured searches, *Scopus* and *IEEExplore* were the databases that had the most relevant content for this thesis. Google Scholar and Summon@BTH gave too many search hits even with a rigorous search block structure. It made the process of producing usable search blocks too cumbersome. Both Springer and Google Scholar did not handle search blocks very well. It made them difficult to use because they produced inaccurate results.

3.2.2 Incident inclusion criteria

Including a cybersecurity incident in this study demands it has to be discussed or disclosed inside the article in sufficient detail to determine the nature of the attack. Not having enough information about the incident made it harder, or impossible, to classify it. It is why this was a requirement. One example would be an article mentioning an incident without information about which year it happened.

Backward snowballing was used when an article discussed an incident using an external source. These external sources were also reviewed to determine if they were trustworthy or not. The external sources were only listed in the results of this study if the information inside the article was insufficient. The external source would need to live up to the same inclusion criteria applied to the articles.

Another limitation is that the victim of the incident was using an IT infrastructure. If the victim of the incident used an OT environment, the incident would be out of scope. The reason for this constraint is that OT systems cannot be replaced entirely by a cloud alternative. Therefore it is not fair to consider these types of incidents for this thesis.

Incidents should also occur because of a vulnerability exploit. This criterion excludes incidents where users install malware themselves. If a user can install malware on their system, it is vulnerable to such exploits. However, it is outside the scope of this thesis because no infrastructure exploits are needed to conduct such an attack. One example of such an attack is the Koobface worm [58].

If the same incident were discussed in multiple articles, that incident was considered to be a duplicate. All duplicate incidents were documented to be considered by the first article that mentioned it.

3.3 Incident dataset review

The dataset used in this study was created in another recent study conducted by G. Abbiati et al. by merging data from three different sources: BLI, PRC and ITRC [1]. The dataset contains 22328 instances of different categories. Table 3.1 displays each feature together with a short description.

Table 3.1: The features of the dataset that was reviewed in this study.

Feature	**Description**
Year	The year the incident occurred.
Location	The country in which the incident occurred.
Records	The number of records (personal identifiable information) that were compromised in the incident.
Entity	The victim that was targeted in the incident.
Industry	The industry of the victim.
Cause	The cause for the incident.
URL	The source for the knowledge of the incident.
cluster	The cluster ID of the merging process. An artefact from the creation of the dataset.

The dataset used in this thesis was the largest one available for public use. It is the primary reason this particular dataset was used in this study. Because the dataset contains instances from multiple sources, it reduces the bias towards one specific source.

The samples were taken from the dataset using the *pandas* Python package [80]. The code that enforced the inclusion criteria and sampled the incidents from the dataset is in the appendix section A.1. Python and pandas were selected because they are openly available to the public and will also make it less tedious to replicate the results.

3.3.1 Sample criteria

The dataset was processed before extracting the samples to filter out any irrelevant incidents. Each subsection explains the motivation behind these filters.

3.3.1.1 No invalid values

The dataset contains instances that have *NaN* values either in the 'Cause' or 'URL' columns. No undefined values could be listed in the 'Cause' column because this column was used to filter out irrelevant incidents.

The 'URL' column could identify incidents with too little information. This column indicated if there were any third-party sources to validate the instances. If an instance in the dataset does not have an entry in this column, it does not have a third-party source that can validate the incident. Therefore any instance with no URL source could be excluded from this study.

This filter reduced the size of the dataset to a total of 10808 instances.

3.3.1.2 Only incidents post 2008

This restriction exists because of the limit defined in RQ1. The dataset was reduced to 10729 after this restriction was applied.

3.3.1.3 Only incidents with 10.000 or more records

It is reasonable to assume that cloud incidents would be of considerable size. In such an incident/breach, many records would be affected. Therefore, to reduce the bias of the sampling process, the dataset study would only include incidents involving more than 10.000 records.

This restriction reduced the number of instances in the dataset to 1606.

3.3.1.4 No 'Inside job' or 'Paper data' incidents

The 'Cause' column in the dataset could also identify incidents that were out of the scope of the study.

Incidents caused by 'Inside jobs' were filtered out because they were out of scope. Incidents in this category do not need to exploit the environment. It makes them irrelevant to this study.

Incidents regarding 'Paper data' are not relevant to this study because they are physical. It is reasonable to assume physical attacks happen exclusively in an on-premise environment. Therefore this category of incidents is not fair to include in this study.

This filter reduced the number of instances to 1504.

3.3.1.5 Only URL sources with valuable information

Some URL values referred to external sources that never contained any usable information. The URLs played a valuable role when deciding if an incident was caused by an on-premise or cloud exploit. Therefore, incidents with useless URLs were filtered out.

After excluding these events, the size of the dataset was reduced to 1016.

3.3.1.6 No duplicate incidents

The authors that created the dataset cannot guarantee that there are no duplicates [1]. The authors of the paper tried to mitigate this by measuring the Jaro-Winkler distance of the alphanumeric entries. Their theory was that this distance could be used to measure how different each word is to each other to find duplicate names. When it came to numeric entries they measured the difference between two instances. Unfortunately, multiple duplicate incidents still existed in the dataset. It was a complicated issue because no feature in the dataset could properly identify duplicate incidents. One example was that an incident could refer to a company "Big Company ltd." and another to "Big Company". It was the same entity, but with a slightly different name. It means the 'company' field could not be used to filter out any duplicates in the dataset. Several combinations of different features were tested ('Year' & 'Record' & 'Cause' & 'Industry' & ...). Finally, 'Year' and 'Records' were deemed the most accurate combination to identify duplicates. This combination makes sense because there is a low chance of 2 incidents having the same values in these two fields.

By using this combination, some incidents could have been filtered out by mistake, but that is an acceptable loss as not many incidents were considered duplicates based on this criteria.

After filtering out duplicate incidents, there were 966 instances left that matched the requirements of this study.

3.3.2 Sampling strategy

The technique used to extract samples from the dataset was *simple random sampling* [27, 121]. This technique takes random samples from a population without replacement. Therefore, each time a random sample was selected, that sample was removed from the original population. This procedure will ensure there are no duplicate instances in the sample subset.

Formula 3.1 and 3.2 were used to determine the recommended sample size. These formulas come from the works of W. G. Cochran [27]. The original purpose of the formulas is to determine the sample size in market research. Even though the subject of the study differs, the formulas fulfill the same purpose in this thesis.

The variables in both formulas have the following meaning:

- n_0 - The recommended sample size for a large or unknown population.
- n - The recommended sample size for the population size N.
- Z - The Z score value which can be obtained from a t-table given that the sample population follows a normal distribution.
- P - The expected proportion of the population in the dataset. How many incidents were estimated to have been caused by a exploit?
- E - The margin of error that is allowed when determining the sample size.
- N - The size of the population from which the samples where taken.

First formula 3.1 was used to determine the sample size of a larger population. If the population size is expected to be very high or unknown, it is enough to apply this formula. However, the population size was known in this thesis. Therefore this formula was not enough to determine the recommended sample size.

$$n_0 = \frac{Z^2 \cdot P \cdot (1 - P)}{E^2} \tag{3.1}$$

The result of formula 3.1 was used inside formula 3.2 as the variable n_0. The result of formula 3.2 (n) will be the final sample size that is recommended based on the size of the population.

$$n = \frac{n_0}{1 + \frac{n_0 - 1}{N}} \tag{3.2}$$

The confidence interval was set to 90% and the margin of error to 10%. Using a high confidence interval and a low margin of error would make the sample size estimation more accurate. However, increasing the accuracy would also increase the number of samples to analyze. It would increase the time necessary to conduct the dataset analysis. Therefore, the values had to be adjusted to satisfy the time frame of the dataset review task.

Because the actual distribution of the dataset for environment exploitation is unknown, the dataset is estimated to have a normal distribution. Normal distribution and a confidence level of 90% corresponds to a Z score of 1.645. Therefore, in this study, Z will be set to 1.654.

Because the proportion between the two categories (on-premises and cloud) is unknown, it is assumed to be equal. Therefore the value for P will be 0.5.

With a margin of error of 10%, the value for E will be 0.10.

The *resample* function from the `sklearn.utils` library was used to gather the samples from the dataset.

3.3.3 Third party source validation

Many incidents inside this dataset had to be investigated by relying on third-party sources. Therefore these third-party sources had to come from trustworthy publishers. One example of such a requirement is that a news article about the incident had to be reported by a respected newspaper. News articles were also required to cite the leading investigators in the case. It made the information more trustworthy because the information would come directly from the source of the incident responders. News articles were considered trustworthy because the victim would likely correct the newspaper if it published incorrect information.

Data breach notices printed by representatives of the victim were also considered trustworthy sources. These notices were trusted because there are incentives for them to be truthful. For example, fines tied to HIPAA are considerably higher if the company exercises "willful neglect" [57]. Other than purposely choosing an incorrect security best practice, this term can also include withholding information about the incident.

Blog posts were also considered trustworthy third-party sources. The majority of incidents were on the *databreaches.net* site. This site provides sources of recent data breaches and summarizes them.

3.4 Interviews

The questions asked during the interviews were designed to be open. It means the questions required more than a yes or no answer. Counterarguments were stated during the interviews to provoke a discussion of the subjects. The purpose was to gain more information and insight into why they gave a particular answer. It also mitigated any biased opinions and forced the interviewee to provide logical reasoning to support their arguments.

Before each interview, the participants were informed of the study, its goals and objectives, to give them a them a foundation to base their decision to participate or not. The questions were then sent to the participants that agreed so that they could review them. This approach gave the interviewee time to reflect on the questions before the interview started. This approach reduced the risk of the interviewee leaving out events or incidents that could have been important to the study.

The interviews were recorded using the built-in record function in Teams and Zoom. Recording the interviews shifted the focus from taking notes to concentrating on the interview itself. It would also minimize the risk of misinterpreting answers and losing information during the interview.

Summaries were made after the interviews based on the conclusions from the recordings. These summaries were then sent to the interviewee so that they could correct any misinterpretations.

3.4.1 Finding participants

Candidates for the interviews were found by searching for relevant projects and organizations. It was important that the participants would have extensive knowledge in cloud security technology and have also had previous experiences with managing on-premise environments. Initially, eight people were asked to participate. Of these eight people, four were from OWASP, one was from Cloud Security Alliance, two were from an on-premise/cloud developer, and one was from the academic community. None of the candidates from OWASP were available, resulting in a total of 4 participants.

3.4.2 Selection and motivation of questions

During the interview, the participants were provided with one statement and then one or more questions based on that statement. Cloud security was the focus of the interviews because cloud-related incidents were so rare during the incident review stage in the study. The idea was to identify different aspects of why cloud security-related incidents appear to be less frequent than other vulnerabilities. Because the interview was limited to 40 minutes, there was only time for eight statements. These eight statements contained a total of 14 questions. The inspiration for the statements

and questions were the finding made during the literature and dataset review in conjunction with relevant surveys.

The statements and their questions are listed below in their respective subchapter.

3.4.2.1 Statement 1

"There have been numerous cyberattacks in the last decade that were initiated by vulnerabilities not specific to the cloud." 4.1

1. Do you recall any cybersecurity incidents that were based on exploiting public cloud environments (PaaS or otherwise)?

2. If you have read any disclosure of any of these incidents, what kind of vulnerability enabled the attack to happen?

3.4.2.2 Statement 2

"According to an article published by CSA, the most common cause for data leakage in the cloud are misconfigurations done by the cloud consumer or MSP." [29].

1. Do you agree that misconfigurations is the most prominent vulnerability in cloud environments (PaaS)? Why?

2. Would you consider misconfigurations a bigger problem than the multi-tenancy nature of public cloud environments?

3.4.2.3 Statement 3

"According to the same study, "prevent loss or leakage of data" is the category that public cloud consumers trust the least." [29]

1. Why do you think that this is the main concern of public cloud consumers?

2. Do you think the result would differ if the same survey was done to customers that are using on-premises solutions? Why, why not?

3.4.2.4 Statement 4

"The study also noted that 18% of all organizations that answered their survey did not know if any cybersecurity incident had occurred during the last year in the public cloud they were using." [29]

1. Do you think this is a problem that is prominent for public cloud consumers? Why, why not?

2. What is the difference when it comes to surveillance of activity inside on-premises versus public cloud PaaS environments?

3. How transparent is a public cloud environment today?

3.4.2.5 Statement 5

"Ransomware is a malware type that is common today (causes about 17% of all incidents 2017). It has made a lot of damage and have made several headlines during the pandemic." [3, 120]

1. From your experience, how does the ransomware threat differ inside on-premises and public PaaS cloud environments?

3.4.2.6 Statement 6

"More and more businesses are migrating from on-premises to the cloud." [13]

1. Do you think that cloud environments will be targeted more extensively in the future because of this?

3.4.2.7 Statement 7

"There are different forms of cyberattacks that can happen, targeted and untargeted." [13]

1. Which environment (PaaS cloud or on-premises) would you say is most exposed towards untargeted attacks? Why?

3.4.2.8 Statement 8

"Because there is no reliant dataset of targeted attacks where environment data is available, the target ratio between on-premises and cloud environments today is pretty much unknown." [1]

1. Based on your expertise, which environment do you think cyber criminals have targeted the most (on-premises or clouds)? Why?
2. Do you think that large cloud providers such as Microsoft or Amazon "scare" cyber criminals from targeting services that run in their clouds?

3.4.3 Ethical considerations

The following ethical aspects were considered and stated to the interviewee before the interview started:

- The interviewee was asked to permit the recording before the interview started. If they did not comply, notes would be taken instead.
- Only the interviewer will have access to the recorded interview and it will not be shared on any medium/platform or with any person.
- The material will only be used by the interviewer to review and summarize the conclusion for each question.
- After the thesis is complete, any recording taken during the interviews will be deleted.

3.5 Threat modeling

STRIDE was the basis of the proposed methodology in this thesis [100]. STRIDE has been widely used in the security industry to discover and deal with security threats in a dynamic way. These two features make it a good fit for this thesis [17]. Another threat model methodology considered during the study was DREAD [82]. However, STRIDE was chosen as the preferred method because DREAD is no longer a standard method that is used for threat modeling [99]. Another reason is that the DREAD methodology focus on possible attacks for which the components are vulnerable. Considering the measurement strategy to quantify threats, this made STRIDE once again the preferred methodology.

3.5.1 Scenario and motivation

The thesis will consider a scenario migrating an on-premise environment to the cloud. The on-premise environment consisted of a booking system with some general security services. SaaS and PaaS cloud solutions replaced the old on-premise services. This scenario matches the one described in the introduction chapter. The company had internal websites/services they wanted to upgrade but were unsure if they should have kept the current on-premise environment or migrated to the cloud. The environment samples in this thesis come from a real migration case which makes them relevant when testing the comparison methodology. The purpose of the comparison is to show if the company made the right choice based on the threats mentioned in this thesis.

3.5.2 Threat model process

The threat model process started with creating a DFD for each environment. The DFDs in this thesis were based on an abstract representation of the on-premises and PaaS environments which were the products of recent migration. The abstractions represent the components in the two environments and how these components interact with each other. Acronyms such as "D" (datastore), "F" (flow), or "S" (service) identified the purpose of each component inside the diagrams. These acronyms were followed by a unique number, for example, "S2", to help identify the different components in the threat model process.

The threat models also contained different trust zones (TZ). The trust zones helped illustrate the trust level of the data while it travels through the environment. Once the data has made it past a firewall, the level of trust for that data would increase as there is a smaller chance that it is malicious.

The DFDs will serve as a reference point to decompose each environment. The result will be an overview of data flows that happen between components.

3.6 Cybersecurity metrics

Both the DFDs and the continuous audit metric catalog defined by Cyber Security Alliance created the foundation for the cybersecurity metrics [24].

The theory is that making a model of the environment will reveal some architectural features of that environment. These features could then give clues about relevant cybersecurity metrics. These metrics would be based on the environment architecture rather than what software and hardware exist within that environment. A methodology to define cybersecurity metrics did not exist in any related literature reviewed during this thesis. There was also a lack of architecture metrics in the existing literature. Therefore, it was relevant to evaluate such a method during this thesis.

However, it is not possible to ignore metrics that consider the software and hardware features of the environment. In theory, measuring environment security features related to the architecture, software, and hardware will create a more accurate picture of the strengths and weaknesses of that environment. It is the reason why measurements from the recent issue of the Continuous Audit Metrics Catalog also were reviewed during this study [24].

The level of detail in each environment was based on the allocated time for this part of the thesis. Therefore, data flows, and the interaction between components and trust zones were the only elements used in this study.

There were some irrelevant metrics listed in the Continuous Audit Metric Catalog. It does not mean that some of these metrics are insignificant when measuring the security of an IT environment but rather that they did not fit the sample environments of this study. One example of such metrics measures how well an organization adapts best practices in software development, which is unrelated to this study because there is no software development process in the sample environments.

Each cybersecurity metric was associated with one or more threats defined in the threat model analysis. The metrics and threats were associated with each other in collaboration with industry specialists. It would make the associations as relevant and accurate as possible. One example is a metric that measures how many users have MFA enabled. This metric could be associated with a threat where a malicious actor can brute-force themselves into the environment.

3.6.1 Metric weights

Because each metric could be associated with a corresponding threat, it was possible to quantify the importance of each metric. It could help a practitioner evaluate the result of this methodology even if both environments were the preferred choice the same number of times. Without a weight system, the result would rely on opinions rather than facts. Therefore this will be a vital part of the comparison methodology.

There are several different ways to weigh measurements against each other. One way could be to use the MITRE ATT&CK framework to rate threats against each other based on the severity of the vulnerabilities enabled by the threats. This process would be complex because one would have to identify each vulnerability before weighing the measurements.

Another way is to let the practitioners themselves rate the threats based on how serious they would be in their particular case. However, for this thesis, this strategy would be all too specific. Therefore it would be hard to argue if the comparison methodology can be generalized. Another problem with this method is that the

practitioner must be familiar with the target environment. It was not possible in this thesis because the environment owner was not directly involved.

A third option would be to impose weights on the measurement results based on how many threats a metric represents. This method was easier to generalize because it ranked the metrics based on how much of the environment they measured. With this logic, a metric that could be associated with all threats in an environment would carry more weight than a metric that measures only 1 of these threats.

The third option was used in this thesis. Therefore, the weights were the number of threats associated with a specific metric.

The following example shows the significance of the weighted measurements. (*Metric 1*) can measure five different threats and (*metric 2*) can measure 3 of the threats. *Environment 1* were the preferred choice when comparing the environments in regards to *metric 1* while *environment 2* were the preferred choice when comparing them using *metric 2*. Even though both environments were the preferred choice, the metrics have different impacts on the overall security of the environment. Because *metric 1* affect more threats than *metric 2*, *metric 1* would affect a bigger part of the environment. Therefore the result of *metric 1* should weigh more than the result of *metric 2*.

The weighted measurement score can be calculated using formula 3.3.

$$w_{score} = \sum_{m=1}^{n} p_m \cdot w_m \tag{3.3}$$

w_{score} is the result of the weighted measurement of all cybersecurity metrics in an environment.

p_m is the comparison result between the environments for cybersecurity metric m. p_m can only take on the three values

- 0 - The environment is not the preferred choice in this particular comparison.
- 1 - The environment is the preferred choice in a particular comparison.
- 0.5 - The metric shows the same number for both environments.

w_m is the number of threats that the cybersecurity metric m can measure. As mentioned above, this will act as the weight.

n will represent the total number of metrics.

Using the above example, formula 3.3 would generate the following results:

$$env_1 : p_1 \cdot t_1 = 1 \cdot 5 = 5$$

$$p_2 \cdot t_2 = 0 \cdot 3 = 0$$

$$w_{score} = 5 + 0 = 5$$

$$env_2 : p_1 \cdot t_1 = 0 \cdot 5 = 0$$

$$p_2 \cdot t_2 = 1 \cdot 3 = 3$$

$$w_{score} = 0 + 3 = 3$$

In this case, the weighted measurement states that *environment 1* is considered more secure than *environment 2* because $5 > 3$.

Chapter 4

Results

4.1 Structured literature review

The resulting search blocks and keywords used in this thesis are listed below. These search blocks were the foundation of the final search strategy during the structured literature review.

Each search block was separated with a logical AND.

- (cybersecurity)
- (breach OR campaign OR case OR incident)
- (attack OR defense)
- (review OR discuss OR study OR investigate)

The structured search resulted in 130 and 236 articles inside the IEEExplore database and Scopus. Based on the titles and abstracts, four articles inside the IEEExplore database and 35 articles inside Scopus were relevant to the study.

By further analyzing these 39 articles, 19 of them described incidents in enough detail to determine the nature of the incident. There were 37 incidents inside these 19 articles that matched the inclusion criteria for this thesis.

One ($\approx 3\%$) of the incidents was associated with cloud exploitation. The other 36 ($\approx 97\%$) incidents were caused by vulnerabilities unrelated to PaaS or SaaS solutions. All 37 incidents are listed in table 4.1.

4.2 Incident dataset review

There were 791 incidents left after applying all inclusion criteria defined in the method section for the dataset review 3.3. By using formula 3.1 inside formula 3.2 with the predetermined values set in the method section, 63 became the recommended sample size. Unfortunately, only 20 samples were possible to extract from one of the relevant categories in the dataset. Since there were only three relevant categories and the result should represent all of them, the sample size was reduced to 60. Approximately 8% of the relevant samples were represented in this study.

The preexisting sources for each sample would sometimes not contain any information regarding the incident. Because of this, additional sources were used to find the necessary information. However, it was not always possible. Some of the

Table 4.1: Incidents (37) identified in the structured literature review meta-analysis.

Year	Victim	AttackReference	AttackType	Article	ExtraInfo	CloudCause
2015	Ukraine	BlackEnergy3	Trojan	[87]	-	FALSE
2015	SingHealth	Orangeworm	Spyware	[61]	-	FALSE
2020	USA Healthcare	Trickbot	Trojan	[61]	-	FALSE
2018	Ministry of Health Singapore	-	Data exfiltration	[61]	-	FALSE
2019	Indian health care site	APT22	Data exfiltration	[61]	-	FALSE
2010	Iran nuclear facilities	Stuxnet	Component hazard	[48]	-	FALSE
2017	UK NHS	WannaCry	Ransomware	[4]	[2]	FALSE
2012	Unknown	DUQU	Trojan	[72]	-	FALSE
2012	Saudi Aramco	Shamoon	Wiper	[72]	-	FALSE
2017	Maersk	NotPetya	Wiper	[72]	-	FALSE
2021	Colonoal Pipeline	Darkside	Ransomware	[72]	-	FALSE
2018	Global	VPNFilter	Wiper	[72]	-	FALSE
2016	Kemuri	-	Data exfiltration	[72]	-	FALSE
2021	Elekta	-	Ransomware	[60]	[8]	**TRUE**
2019	Orion customers	-	Data exfiltration	[9]	-	FALSE
2016	Dyn	Mirai	Botnet	[97]	-	FALSE
2015	Anthem	-	Data exfiltration	[18]	[101]	FALSE
2011	Sony Pictures	-	Data exfiltration	[3]	[103]	FALSE
2014	Sony Pictures	Guardians of Peace	Data exfiltration	[3]	[46]	FALSE
2014	JP Morgan	-	Data exfiltration	[3]	-	FALSE
2015	Ashley Madison	-	Data exfiltration	[3]	[52]	FALSE
2016	Hollywood Presbyterian Medical Center	-	Ransomware	[69]	[7]	FALSE
2020	UCSF	Netwalker	Data exfiltration	[120]	-	FALSE
2016	krebsonsecurity	Mirai	Botnet	[89]	-	FALSE
2020	University of Vermont Health Network	-	Ransomware	[60]	[77]	FALSE
2016	Hollywood Presbyterian Medical Centre	-	Ransomware	[78]	-	FALSE
2020	Argentina Ministry of Public Health	-	Data exfiltration	[22]	-	FALSE
2020	Brazil's Ministry of Health	-	Data exfiltration	[22]	-	FALSE
2020	Chilean citizens	-	Phishing	[22]	-	FALSE
2020	Dominican republic citizens	-	Phishing	[22]	-	FALSE
2020	Colombian citizens	-	Phishing	[22]	-	FALSE
2020	Costa rica citizens	-	Ransomware	[22]	-	FALSE
2020	Mexican citizens	-	Phishing	[22]	-	FALSE
2016	Bank of Greece	Anonymous	Unknown	[11]	-	FALSE
2020	Kaseya	Revil	Ransomware	[44]	-	FALSE
2013	Target	Target breach 2013	Data exfiltration	[71]	-	FALSE
2014	KHNP	No Nuclear Power Plant Group	Wiper	[19]	-	FALSE

incidents were too generic. In other cases, the incidents were never fully disclosed to the public. Therefore information regarding some incidents was impossible to find.

The samples and results from the dataset review are visible in table 4.2. The 'Location', 'Industry', and 'cluster' features were excluded because they do not contribute to the result of this study. The 'URL' feature was renamed to 'Source' because this name makes more sense in the context of this study. The 'CloudCause' feature indicates if the incident was caused by a cloud-related exploit or not.

Three ($\approx$ 5%) incident samples were related to cloud exploitation. 45 ($\approx$ 75%) incident samples were unrelated to cloud exploitation. It was impossible to determine the cause for the remaining 12 ($\approx$ 20%) incidents. It means that 20% of the sampled incidents were associated with vulnerabilities that could exist inside both on-premises and cloud environments. Most of these incidents were the result of web exploitation. A website could be hosted in the cloud and on-premise. Therefore, if it is not explicitly mentioned where the victim's website was hosted, it is impossible to say what environment was exploited.

Table 4.2: Incidents (60) sampled from the compiled dataset together with additional sources and the result from the incident analysis

Year	Records	Entity	Cause	Source	CloudCause
2018	150000	[24]7.ai.	Hacking or malware	[85]	FALSE
2018	10396	Aflac	Unintended disclosure	[40]	FALSE
2016	12738	Afrikadating.com	Hacking or malware	[25]	FALSE
2018	22661	Applied plan administrators	Other / Unknown	No info	N/A
2014	400000	AVAST Software	Hacking or malware	[106]	N/A
2016	50000	Baton Rouge Police	Hacking or malware	[81]	FALSE
2018	33644	Blue beacon international	Other / Unknown	[15, 56]	FALSE
2016	324000	BlueSnap/ Regpack	Hacking or malware	[64]	N/A
2013	21054	Boston Public Schools	Unintended disclosure	[119]	FALSE
2018	25450	Bronson nutritionals llc	Other / Unknown	[56, 104]	FALSE
2018	12049	Centris federal credit union	Other / Unknown	[56, 93]	FALSE
2018	11247	Cityof goodyear arizona	Other / Unknown	[49, 56]	N/A
2016	60000	Coast Central Credit Union	Unintended disclosure	[67]	FALSE
2014	15000	Colorado health	Unintended disclosure	No info	N/A
2018	662475	Comply right inc	Other / Unknown	[56, 68]	FALSE
2018	42587	CPT Group Inc.	Other / Unknown	[53, 56]	FALSE
2018	825000	Delta air lines inc	Other / Unknown	[56, 108]	N/A
2015	21000	Department of Health and Community Services / Newfoundland and Labrador's prescription drug program	Unintended disclosure	[33]	FALSE
2018	246167	Department of Homeland Security	Hacking or malware	[62]	FALSE
2015	30263	Drjizz	Hacking or malware	[94]	FALSE
2014	40029	ejhb.info	Hacking or malware	No info	N/A
2016	260000	EMR4all/Rehab Billing Solutions (RBS)	Unintended disclosure	[36]	**TRUE**
2016	33407472	Evony Gaming Company	Hacking or malware	[118]	N/A
2015	87314	Eye Associates of Pinellas	Hacking or malware	[37]	FALSE
2018	657529	Fast health corporation	Other / Unknown	[41, 56]	FALSE
2018	15000	Fresno State	Other / Unknown	[28, 70]	FALSE
2018	10796	Funding circle USA Inc	Other / Unknown	[55, 56]	**TRUE**
2018	23020	Health equity inc	Other / Unknown	[56, 116]	FALSE
2014	123000	Heart and Stroke Foundation of Canada (The Email Company)	Unintended disclosure	[96]	FALSE
2016	20800	HEI Hotels & Resorts for Starwood / Marriott / Hyatt and Intercontinental	Hacking or malware	[95]	N/A
2016	17841	Henry County	Hacking or malware	[38, 113]	FALSE
2018	42000	Huntington Hospital/Cohen / Bergman / Klepper & Romano	Unintended disclosure	[31]	FALSE
2015	576274	Ifit/ NordicTrack/ ICON	Unintended disclosure	[35]	FALSE
		Health & Fitness			
2018	34097	L a fashion enterprise l t d	Other / Unknown	[47, 56]	FALSE
2016	2919651	Louisiana Voters	Unintended disclosure	[75]	FALSE
2015	14500	LSU Health New Orleans	Hacking or malware	[6]	FALSE
		School of Medicine			
2018	162507	Macys inc	Other / Unknown	[56, 63]	N/A
2018	39000	Massachusetts Department of Revenue	Other / Unknown	[92]	FALSE
2018	276057	Med associates inc	Other / Unknown	[43, 56]	FALSE
2018	120000000	NameTests	Unintended disclosure	[23]	FALSE
2017	316000	Patient Home Monitoring/AWS	Unintended disclosure	[110]	**TRUE**
2014	20428	Rady Children's Hospital	Unintended disclosure	[5]	FALSE
2018	18785	Rail Europe North America inc	Other / Unknown	[56, 112]	FALSE
2018	18133	Rea.deeming beauty inc	Other / Unknown	[56, 91]	FALSE
2015	50000	Regional Income Tax Agency	Unintended disclosure	[50]	FALSE
2015	1400000	Register.com	Hacking or malware	[98]	N/A
2018	11000000	SaverSpy	Unintended disclosure	[26]	FALSE
2017	650000	Shelby County Tennessee	Unintended disclosure	[51]	FALSE
2018	135000	St. Peter's Surgery & Endoscopy Center	Hacking or malware	[76]	FALSE
2018	17618	Systeme software inc	Other / Unknown	[54, 56]	FALSE
2018	3030920	Task rabbit inc	Other / Unknown	[20, 56]	N/A
2014	48439	The Office of Personnel Management (KeyPoint Government Solutions)	Hacking or malware	[107]	N/A
2018	27000000	Ticketfly	Hacking or malware	[83]	FALSE
2014	35000	Tri-City Medical Center	Unintended disclosure	[102]	FALSE
2015	165000	Uncle Maddio's Pizza Joint	Unintended disclosure	[34]	FALSE
2018	150000000	Under Armour	Hacking or malware	[114]	FALSE
2014	3650000	United States Postal Service	Hacking or malware	[88]	FALSE
2017	819977	vBulletin	Hacking or malware	[117]	FALSE
2016	27393	Walmart #2	Unintended disclosure	[39]	FALSE
2018	1000000	Xgimp / MaxiPDF and Docswork	Unintended disclosure	[42]	FALSE

4.3 Interview sessions

The experience of the participants varied. Some of the participants had previously worked as IT managers for larger corporations which gave them extensive insight in

aspects concerning on-premise security. All participants were experienced in cloud security technologies and various cloud service provider technologies. Their answers are listed in the following subsections.

4.3.1 Statement 1 (Answers)

4.3.1.1 Interviewee 1

"A clear trend is that more attacks are directed toward endpoints and user identity. We are also seeing more supply-chain attacks. One incident that stands out is the Log4J incident. Because this incident is tied to the Java log tools it is used both in the cloud and on-premises environments so the attack is not tied to a specific environment."

"There are a couple of reasons why on-premises environments are over-represented when it comes to successful cyberattacks in your incident review."

"One reason is that on-premises environments have more problems when it comes to zero-trust. Another reason is that a cloud environment generally has robust protection when it comes to infrastructure security compared to a typical on-premises environment. A third reason is that the cloud in general can utilize more resources than a typical on-premise environment has. For example, the staff and level of competence within the staff are higher inside cloud service providers. A fourth reason is that you get a centralized overview of your entire environment inside a cloud environment. These consoles often help users to choose best practices and provide patching options on the software that is running in the cloud, for example in PaaS instance."

"However, when it comes to software vulnerabilities and other design flaws, a cloud environment will be just as vulnerable as an on-premise environment and vice versa. For example, if you run a website that is vulnerable to SQL injection, an attack will succeed no matter where the website is hosted."

4.3.1.2 Interviewee 2

"I do not have a specific incident in mind where a cloud feature was exploited, however, cloud platforms are often exploited to conduct phishing campaigns. One example is that malicious actors can use either Outlook online or Gmail to send malicious e-mails to users which might be trusted by a company spam filter because these domains are legitimate."

4.3.1.3 Interviewee 3

"I do not have a specific incident that exploited a cloud feature. The recent ENISA Threat landscape report states that both incidents and breaches were more common among external cloud assets compared to services running on-premises in 2021 [45]. The report also mentions that a possible reason for this rise in cloud incidents and breaches is the pandemic and that many companies migrated to the cloud in haste. Therefore many companies missed implementing their cloud services according to best practices which made them exposed to malicious actors." The report the interviewee referred to can be found on ENISA's website.

"A cloud attack vector is larger than a typical attack against an on-premises environment and this is one reason cybercriminals prefer to breach a cloud environment over an on-premises alternative."

"Attackers are currently leveraging medium vulnerabilities the most because research shows that those are left open for a longer period in comparison to higher risk score vulnerabilities."

4.3.1.4 Interviewee 4

"I do not have a specific incident in mind that exploited a cloud feature, however, a typical data breach is caused by a lack of access control. One example is that you have an exposed database that was configured incorrectly. Many types of misconfigurations can happen that could cause serious incidents." These comments correlate to what is mentioned in the ENISA Threat landscape report as well as the CSA Cloud Security Risk report [29, 45].

4.3.2 Statement 2 (Answers)

4.3.2.1 Interviewee 1

"I agree that misconfigurations should be the most common cause of incidents and data breaches, but the term "misconfiguration" is very broad."

"Multi-tenancy bolsters the severity of misconfigurations in the cloud. But one should remember that misconfigurations are also a big problem in on-premise environments. It is easier to make the right decision in an environment where the majority of services are located in the cloud because a hybrid (cloud and on-premise combined) or on-premise environment often contains a large variety of components that you have to keep track of yourself to make them secure."

"The risk of someone gaining access to another tenant because of a misconfiguration done by the cloud provider is very small (depending on the provider of course). One has to weigh the risk of this happening with the benefits that a cloud migration could provide."

"Unfortunately, I have little information when it comes to solutions for multi-tenancy problems as these solutions are often related to hardware which I have less experience with. However, I know that cloud providers put great effort into solving this. Cloud providers also put a lot of resources into other physical security aspects that on-premise environments rarely have."

4.3.2.2 Interviewee 2

"I agree that misconfigurations should be the primary cause of incidents that happen inside cloud environments."

"I have never experienced an incident that was caused by a problem regarding multi-tenancy factors, so I would guess that this problem is less likely to happen compared to misconfigurations. The problem with misconfigurations is that there is a lack of knowledge amongst those who administrate cloud services. A hybrid environment is the most dangerous one because you need to know how to secure

the services running both in the cloud and on-premises. This increases the risk of misconfigurations to happen which can, or rather will, lead to incidents."

4.3.2.3 Interviewee 3

"Yes, misconfigurations are the most common cause of incidents to happen. For example, if you do not protect the APIs correctly, this could cause one tenant to access another tenant's environment."

"Misconfigurations are more problematic than multi-tenancy. Not everything is linked to misconfigurations, many that rush migration to the cloud end up making an insecure architecture design. This problem is also extensively seen today, especially since when the pandemic started and many companies wanted to rush their services up into the cloud for better availability."

4.3.2.4 Interviewee 4

"Definitely, the majority of incidents are caused by misconfigurations."

"Misconfigurations and multi-tenancy are linked. If you deploy something in a public cloud, another one that gets access to that cloud could then access your resources under the right circumstances. These kinds of multi-tenancy problems usually happen because of a misconfiguration, either by the cloud consumer or by the cloud provider, and are rarely a "bug" problem inside the cloud environment."

"The vast majority of misconfigurations are done by the cloud consumers themselves rather than the cloud provider. If a cloud provider makes a misconfiguration, it will have consequences on the multi-tenancy level while a misconfiguration done by a cloud consumer will only affect the cloud consumer that applies the configuration. A misconfiguration done by a cloud provider is a rare form of vulnerability while a misconfiguration done by a cloud consumer is more common."

4.3.3 Statement 3 (Answers)

4.3.3.1 Interviewee 1

"There is much psychology in this. It feels more secure if you can see exactly where your data is stored. It is also a question of competence. Cloud service providers have in the past left out information about their customers when states have asked for it, for example, the USA. What is less known is that this is done in extremely rare circumstances and when it does happen the customer is notified about it and will know exactly what was shared with the agency that asked for the information."

"In normal circumstances, the cloud providers do not have access to their customer's data but they do have a backdoor that can be initiated if it is necessary. These circumstances also include times when a cloud provider needs to access the customer environment to help in a support case. This depends of course on what cloud service provider you have and as a customer, you should be wary about cloud service providers with prices/conditions that are too good to be true."

"CSPs are obligated to recommend to users how they can set up their services in a secure manner, but it is the customers' choice if those recommendations are

followed or not. If we consider a PaaS implementation the customer will however be forced into best practice settings for the underlying operating system."

"There is an exaggerated belief that MSPs are very good at what they do. If you outsource to an MSP you should consider the likelihood that this MSP has less competence in regards to cybersecurity in comparison to an established CSP. We now circle back to factors related to psychology/competence rather than rational reasoning. A customer feels more secure with an on-premise MSP because the technologies are familiar to the customers and they often have direct communication lines with responsible within that MSP".

4.3.3.2 Interviewee 2

"My experience in the IT security field does not align with the result of the study. I think that the result is more of a reflection that there are concerns about giving your data to someone else rather than that someone being a public CSP. A good analogy is that people tend to be more scared of traveling by aircraft than by car even though car accidents are more common. When you travel in an aircraft you do not have any control of your surroundings and this lack of control transitions into fear."

"I think the result of the study would be pretty much the same if it targeted customers that use on-premise MSPs because the issue is again not that people do not trust clouds to manage their data, but rather that the data is managed by someone else. When you are past this outsource threshold and you are willing to let someone else handle your data, the step from on-premise MSPs to public cloud storage is in my experience not that significant."

4.3.3.3 Interviewee 3

"This should be the biggest concern. The cloud provider has access to whatever you upload to the cloud. In many cases, public cloud consumers don't even know in which country their data is stored. In a public cloud, the customer loses control of their data. One big problem is that data protection laws are different in the US compared to the EU. For an on-premise environment, there is no question about where the data resides."

"There are also attacks such as hypervisor high-jacking that leverage the nature of virtualization where an attacker could access data on other tenants at the same virtualization layer. This is a bigger problem for public cloud environments because anyone can open an account or instance there including potential attackers."

4.3.3.4 Interviewee 4

"Cloud consumers are worried about their data because they don't know where it is located. When you create resources in the cloud you can select a specific region, for example, East US or North Europe, where the data will be stored but you don't know for example which country or which particular data center the data is located. The fact that you don't have this transparency is one of the reasons customers are worried about storing their data in the cloud."

4.3.4 Statement 4 (Answers)

4.3.4.1 Interviewee 1

"I think that more than 18% would be uncertain if they have experienced any cybersecurity incident during the last year or not. It all comes down to how many resources you have that can be used to discover intrusions in the cloud. If you are using a PaaS solution the responsibility of detecting these intrusions is transferred to the CSP which means that you will see less intrusion activity yourself."

"The alternative is having an employee whose only task is to discover intrusions in the environment. This employee will however not be enough to detect certain intrusions because they will have limited resources. However, since they have the sole responsibility for the environment they might think that they can say if there has or has not been any intrusion in the environment during the last year. This also depends on the fact that you trust yourself more than you trust a third party which in this case is a CSP."

"When it comes to transparency, the cloud can in many ways outperform on-premise environments. One example is that cloud technology allows you to get a very detailed picture of what has happened during an incident. The main reason for this is that there is more consistency within the cloud environment compared to typical on-premise environments. In other words, the transparency is often better in the cloud when it comes to cybersecurity but not as good if you want to view what processes are running on your PaaS instance."

4.3.4.2 Interviewee 2

"The result of the study is not specific to the cloud. My experience is that you have less insight into your on-premises environment because when you migrate your environment to the cloud, you have a detailed view of what you migrated. If you have an on-premise environment, there are many components that either have been forgotten/misplaced or that you did not know even existed in your environment (shadow IT). The exception is that employees can sometimes buy e.g. Amazon or Azure services without company approval. This means that the insight into one's environment can also be poor in the cloud. You will however have a better chance of keeping track of what has access to the company environment in the cloud compared to in an on-premises environment."

"There exists good solutions today that provide insight into a cloud environment. Cloud solutions often include centralized consoles that consolidate information about the environment. Trust is a vital factor when outsourcing applications and systems. If one chooses to migrate a solution to a PaaS environment, you trust that the CSP handles the security patches and monitors the underlying operating system to detect possible intrusions and does so much better than you would do yourself."

"In regards to transparency, there were, in the past, functionality that was only available if you had an on-premise environment. Today the cloud has evolved to a point where this is no longer true and the cloud can even be more transparent than a typical on-premise environment. Both for yourself but also for the CSP."

4.3.4.3 Interviewee 3

"The transparency into PaaS instances is different depending on the provider. Usually, a cloud provider does not inform cloud consumers of incidents even though they should according to ISO 270001. This shortcoming could contribute to the result of this study."

"The percentage could be lower in an on-premises environment but the ground issue is that incidents are not shared with everyone, even within the same company. To reduce this number, a norm of sharing incident data should be introduced."

4.3.4.4 Interviewee 4

"If you don't know if an incident has occurred it is likely because the cloud consumer has not configured proper logging for their applications rather than lack of transparency from the cloud provider side. Again we have an issue tied to misconfiguration. It is also possible that some of the ones that answered the survey will be aware of incidents after they participated. It is not uncommon that incidents are discovered much later when the damage has already been done, this is the result of not having enough resources and/or tools for detecting incidents in real-time."

"Because cloud consumers have no access to some of the stacks inside a cloud it could cause some concern for some public cloud users were handling their security is important. If you would use a PaaS instance, written reports will be the only source of information regarding security incidents in the cloud infrastructure. Unfortunately, these reports don't give a very detailed view of the incidents (why the compromise happened for example) and are more legal than technical."

4.3.5 Statement 5 (Answers)

4.3.5.1 Interviewee 1

"A decisive factor for the success of recent ransomware (for example during the pandemic) is that employees sit at home where they do not have direct access to the central environment. This is a problem if a new security patch needs to be installed on their PCs that are distributed through a company patch server inside the company network. This makes teleworkers vulnerable to ransomware attacks and other intrusions."

"My own experience tells me that cloud environments tend to be more resilient against ransomware than on-premise environments, primarily because patch management becomes easier using a cloud alternative."

4.3.5.2 Interviewee 2

"During the last 10 years, on-premise environments have been more exposed to ransomware. This will not change any time soon. In cloud environments, you often have robust resilience that lets you restore environments in a relatively easy way compared to an on-premise alternative. It is important to note that if you do not have any backups it does not matter which environment you use".

4.3.5.3 Interviewee 3

"On-premises will suffer more damage from a ransomware attack than a cloud alternative. Backup alternatives are more difficult to set up in an on-premises environment, especially a backup that works. With that said, a backup done in the cloud should be tested and verified as well, but it is easier to test a cloud solution backup/restore than doing the same on an on-premises solution."

4.3.5.4 Interviewee 4

"Ransomware could affect anything, it does not matter if the data is in the cloud or on-premises. As long as you have a working backup, the ransomware is pointless. In the cloud, you have a better chance of restoring your systems because there are better tools for backup and restore there."

"Concerning disaster recovery, you are not limited by hardware in the cloud or the architecture. This makes a cloud alternative superior to an on-premise solution. Even here cloud misconfigurations are common but as long as you make the correct choices as a cloud consumer you have a high resilience against ransomware."

4.3.6 Statement 6 (Answers)

4.3.6.1 Interviewee 1

"Yes, attacks against cloud solutions will continue and get worse in the future because that is where the data will be located. Hackers choose the easy route but also turn their attention toward interesting objectives. CSPs put a lot of resources into making their platforms the "hard route".

"In the future, if the majority of all environments will be based around cloud technology, that is where cybercriminals will focus their time and resources."

4.3.6.2 Interviewee 2

"Cloud environments will be targeted in the future in a more frequent way than they are today. This is only natural because many services and data will be hosted/based on cloud technology."

"Cybercriminals are often familiar and know about on-premise technology which can be one factor to why on-premise environments might be more represented in studies such as the one you did. It is hard to say how the relationship between the environments will look like in 10 years, but today cloud environments are more secure."

4.3.6.3 Interviewee 3

"The threat landscape has and will change in the future. The attacks against cloud providers will get bigger and scarier in the future. It is much about the risks and what you gain/lose when you migrate to a cloud environment. More standards are coming out regarding cloud security from organizations such as ENISA to help cloud providers be more secure."

"Social engineering is still a big part of most incidents. This has and will probably be used in the future to a large extent in compared to built-in zero-day vulnerabilities in the cloud environment."

"Even though we have a higher security standard in large cloud providers the attack rate will probably increase because the incentives and economic gain of hacking will not decline."

"The number of nation state-sponsored attacks increase as well and are more numerous than ever before."

4.3.6.4 Interviewee 4

"Cloud incidents will be more common in the future because much of the assets that cybercriminals want will be located in the cloud. The situation is similar to when people started to use banks. Instead of breaking into homes stealing money from one family or company, criminals now break into banks to steal more gold in one heist."

"This could increase the risk of using cloud services in the future because there is a bigger risk for a cloud zero-day exploit which gives an attacker access to all data that resides in a particular cloud. Even though your company was not explicitly targeted, it would still be affected because you were using the same cloud as another company the cybercriminals were interested in."

4.3.7 Statement 7 (Answers)

4.3.7.1 Interviewee 1

"It depends on the intention of the attackers. There is no distinction between untargeted attacks for on-premise and PaaS solutions because automatic scanners will find the vulnerability wherever the application is located as long as it is reachable from the internet. The answer to this question will reflect more on which environment has most of the vulnerable applications rather than what environment is more secure than the other."

4.3.7.2 Interviewee 2

"It is hard to say that a specific environment is more exposed than another one when it comes to untargeted attacks. The exposure will depend on what kind of connection the environment has to the internet."

4.3.7.3 Interviewee 3

"Both environments have the same potential to get hit by an untargeted attack, but the damage will be more extensive to on-premises environments."

4.3.7.4 Interviewee 4

"All environments are exposed to untargeted attacks. In an untargeted attack, anything that is connected to the internet will be exposed. However, if you have a known vulnerability in an operating system in a PaaS solution, that vulnerability

should be patched by the cloud provider or the PaaS platform will not be compliant with GDPR (for example) which could have massive financial consequences for them if an incident were to happen because of it."

4.3.8 Statement 8 (Answers)

4.3.8.1 Interviewee 1

"It does not matter where the services or data is hosted. If a malicious actor wants your information badly enough they will get it in a targeted attack."

"Cybercriminals are well aware of what capacity the major CSPs have and that these CSPs harbor expert knowledge whose only job is to make it as hard as possible for cybercriminals to penetrate their services. Therefore it could be possible that less serious actors ignore a specific target simply because they are after quick money. For more serious actors such as APTs it does not matter and they will eventually succeed when it comes to targeted attacks no matter what environment you are using."

4.3.8.2 Interviewee 2

"The cybercriminals that are after quick money in targeted attacks do not care that much about the environment but try to exploit the people that are using the services that run in that environment through social engineering.

"If the goal of the targeted attack is extortion, e.g. ransomware, on-premise environments are the ones that are most exposed."

"Important information is traditionally kept inside on-premise environments which would probably make them more targeted".

"Nowadays a lot of cybersecurity solutions are cloud-based. These solutions are often used by on-premise systems which makes it hard to make this kind of comparison."

"It is easier to stop an intrusion inside a cloud environment as long as the malicious actor does not leverage a zero-day exploit. If a zero-day exploit is used, the environment does not matter because the attack will with all certainty succeed".

"The majority of cybercriminals can move on if they bump into a service they know they cannot get through easily, but this is less dependent on the environment itself."

4.3.8.3 Interviewee 3

"I think the environment is an important aspect of an attack when you plan it which is usually done in a targeted attack. This is because they use tools that are specific to a certain environment in their attacks."

"However, when you have a targeted attack you don't discriminate between environments or cloud providers. When you have a target, you focus on it no matter who it is. This is different depending on what kind of threat actor is performing the attack and what information they are after. If the reward is not high enough compared to the time they need to spend to get inside, they will be "scared" away to try either another component or maybe another company altogether."

4.3.8.4 Interviewee 4

"It is not easy to have data on how attacks have been carried out."

"It depends on which kind of cybercriminal attacks the company. In a targeted attack, the cybercriminal will attack anything they can attack and do not discriminate or favor different environments. Targeted attacks are most successful while untargeted attacks are not. This is because targeted attacks are carried out by professionals while untargeted attacks are mostly conducted by amateurs."

"When you attack a cloud service you take on a fortress. A cybercriminal could target the fortress to get what they want, but it is usually easier for them to just attack the users of the service instead, for example by using phishing attacks."

4.4 Threat modelling

4.4.1 Architecture overview

The features that should be supported by the environment(s) includes:

- Ability to login to the booking system.
- Ability to add new bookings.
- Ability to manage existing bookings (delete/edit).
- Ability to access the booking system outside of the office in a secure manner.
- Ability to send status to and get anti-virus updates to a centrally managed service.
- Ability to send device status and receive software updates via a centrally managed service.
- Ability to access the internet.

The booking system itself has the following features:

- Ability to authenticate a user using a third party AAA solution.
- Ability to store booking information in a SQL database.
- Ability to have information backed up periodically.
- Ability to alert IT staff if services fail.

4.4.2 Decomposition of environment (On-premises)

Figure 4.1 shows the flow of data inside the on-premises environment and figure 4.2 represents the same for the cloud environment.

The decomposition of the two environments was done separately because the data flows differed. The architecture was the focus of the decompositions to show the data flow that supported the features of the environment.

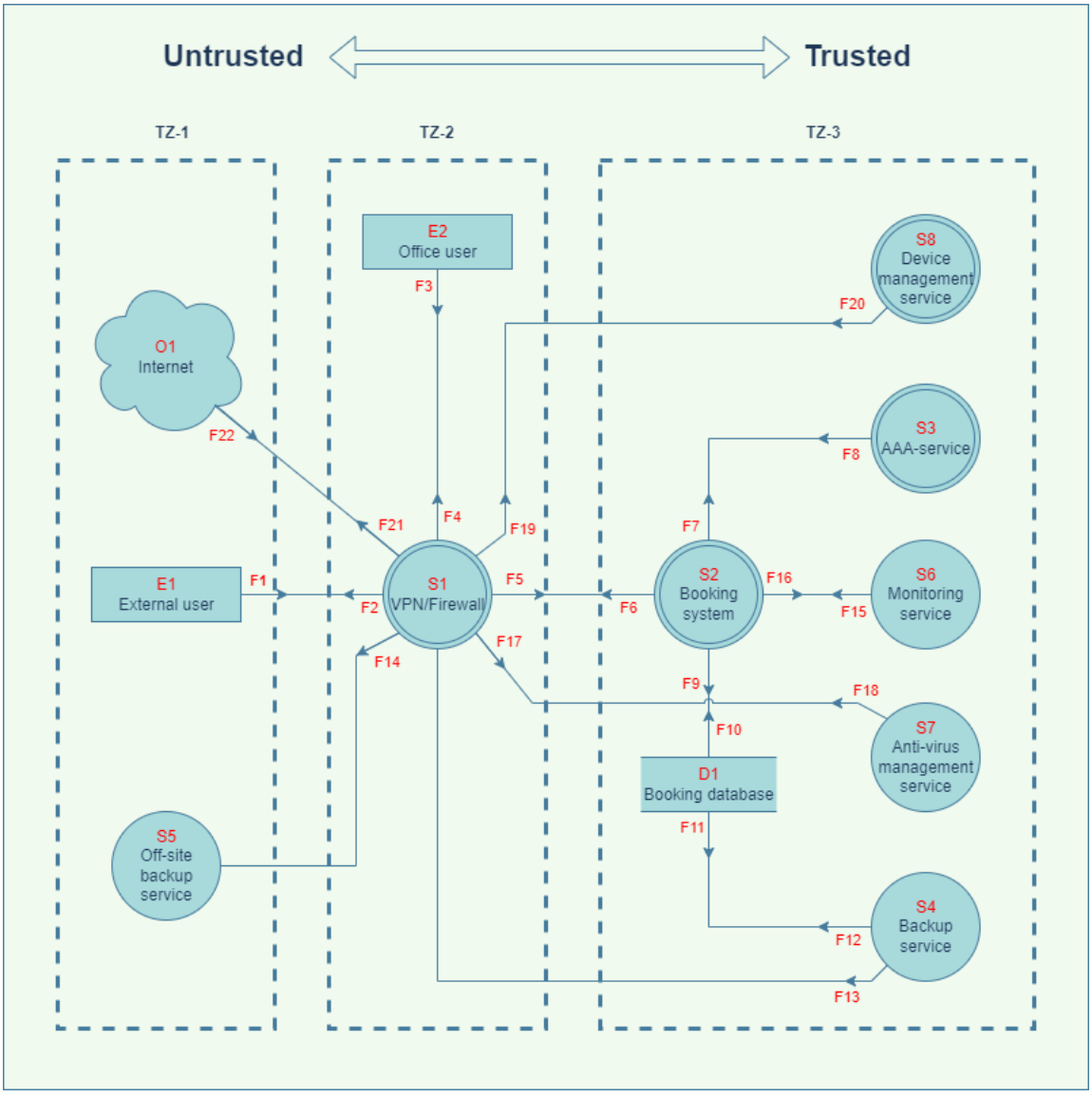

Figure 4.1: The data flow diagram of the on-premises environment.

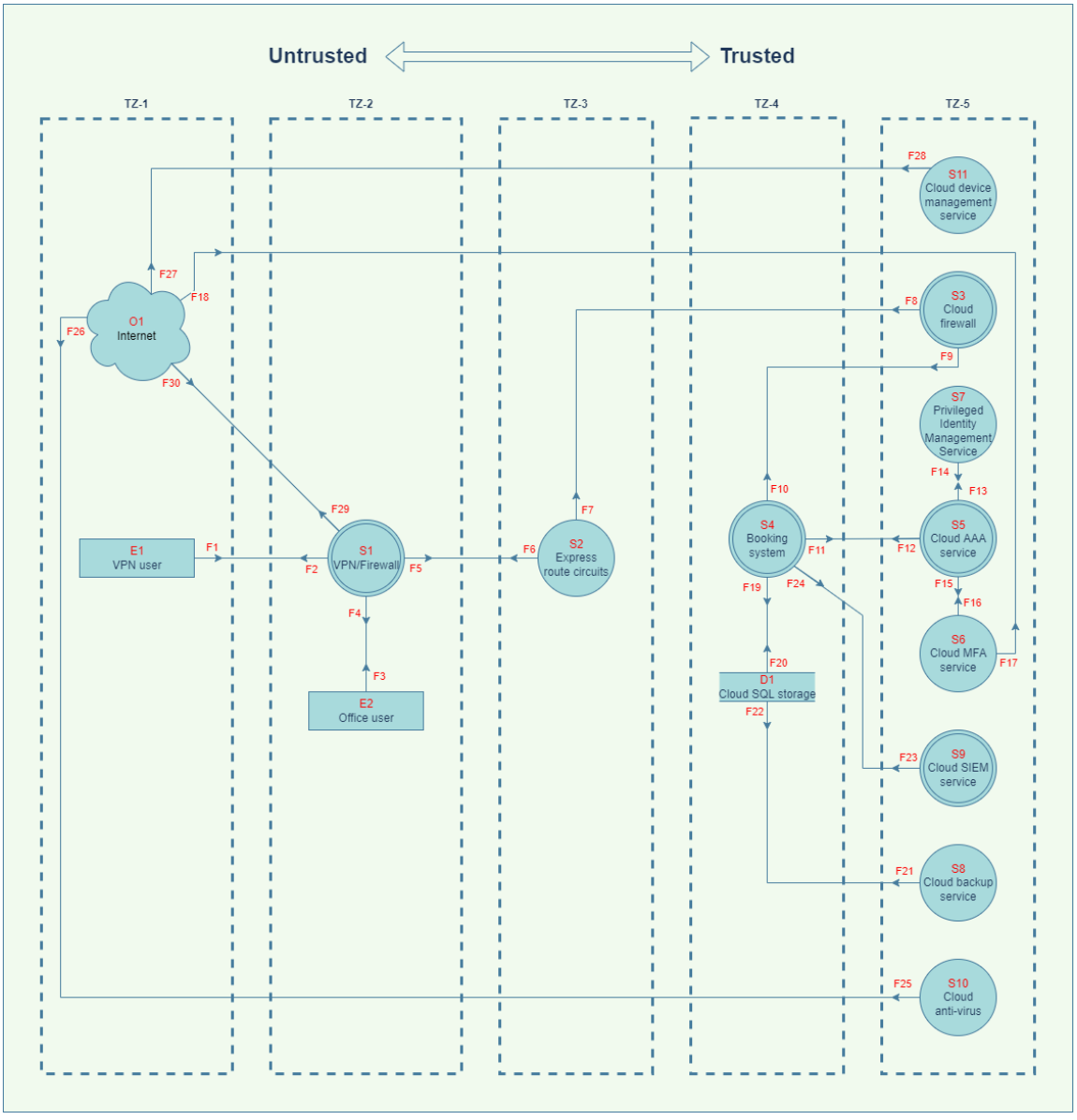

Figure 4.2: The data flow diagram of the cloud environment.

4.4.2.1 Login process steps

1. *E1* and/or *E2* initiates a login request (*F1* & *F3*).
2. *S1* forwards the request to *S2* (*F5*).
3. *S2* forwards the request to *S3* (*F7*).
4. *S3* validates the provided login credentials and sends back a response to *S2* (*F8*).
5. *S2* sends a login cookie to the user (*F6*).
6. *S1* relays the cookie to either *E1* or *E2* (*F2* & *F4*).

4.4.2.2 Booking management steps

1. *E1* and/or *E2* initiates an authenticated request to *S1* (*F1* & *F3*).
2. *S1* forwards the request to *S2* (*F5*).
3. *S2* queries *S3* to find out if the user is authorized to make the request (*F7*).
4. *S3* gives a response to *S2* (*F8*).
5. *S2* queries the required information from *D1* (*F9*).
6. *D1* responds to *S2* with the requested information (*F10*).
7. *S2* makes the requested changes to the data (Add, Delete or Modify) and saves the changes inside *D1* (*F9*).
8. *S2* sends a change confirmation to the user client (*F6*).
9. *E1* or *E2* gets the change confirmation/denial (*F2* & *F4*).

4.4.2.3 Booking system backup steps

1. *S4* sends a backup request to *D1* (*F12*).
2. *D1* responds to *S4* with the requested data (*F11*).
3. *S4* sends the data to *S5* through *S1* (*F13*).
4. *S5* receives the backup data from *S1* (*F14*).

4.4.2.4 Booking system monitoring

1. *S6* requests the system status from *S2* (*F15*).
2. *S2* responds with the current system status to *S6* (*F16*).

4.4.2.5 Anti-virus update steps

1. *E1* and/or *E2* sends an update request to *S2* (*F1* & *F3*).
2. *S1* forwards the update request to *S7* (*F17*).
3. *S7* sends the updated anti-virus information to *S1* (*F18*).
4. *E1* and/or *E2* receives the update from *S1* (*F2* & *F4*).

4.4.2.6 Device management steps

1. *E1* and/or *E2* initiates a status update/software update request (*F1* & *F3*).
2. *S1* forwards the request to *S8* (*F19*).
3. *S8* acknowledge the status update and/or sends additional instructions/software/commands to *E1* & *E2* (*F20*).
4. *S1* forwards the update data to *E1* and/or *E2* (*F2* & *F4*).

4.4.2.7 Internet access steps

1. *E1* and/or *E2* initiates a connection to *S1* (*F1* & *F3*).
2. *S1* forwards the connection to *O1* (*F21*).
3. *O1* returns the requested information (*F22*).
4. *S1* relays the information to either *E1* or *E2* (*F2* & *F4*).

4.4.3 Decomposition of environment (Cloud)

4.4.3.1 Login process steps

1. *E1* and/or *E2* initiates a login request (*F1* & *F3*).
2. *S1* forwards the request to *S2* (*F5*).
3. *S2* forwards the request over a private WAN link to *S3* (*F7*).
4. *S3* forwards the request to *S4* (*F9*).
5. *S4* sends an authentication request to *S5* (*F11*).
6. *S5* validates the provided login credentials and initiates a MFA authentication request to *S6* (*F15*).
7. The MFA request is relayed to the user through *O1* (*F17*).
8. *S1* receives the MFA request from *O1* (*F30*).
9. *E1* and/or *E2* receives the MFA request from *S1* (*F2* & *F4*).

10. *E1* and/or *E2* approve or deny the MFA request by signing and sending it back through *S1* (*F1* & *F3*).
11. *S1* relays the signed response to *O1* (*F29*.
12. *S6* gets the response from *O1* (*F18*).
13. *S6* sends the MFA result to *S5* (*F16*).
14. *S5* sends the authentication result back to *S4* (*F12*).
15. *S4* sends an error message or authentication token back to the user (*F10*).
16. *S3* forwards the token to *S2* (*F8*).
17. *S2* forwards the token to *S1* (*F6*).
18. *E1* and/or *E2* receives their authentication token from *S1* (*F2* & *F4*).

4.4.3.2 Booking management steps

1. *E1* and/or *E2* sends an authenticated request to *S1* (*F1* & *F3*).
2. *S1* forwards the request to *S2* (*F5*).
3. *S2* forwards the request over a private WAN link to *S3* (*F7*).
4. *S3* forwards the request to *S4* (*F9*).
5. *S4* queries *S5* to find out if the user is authorized to make the request (*F11*).
6. If the action requires elevated privileges the request is forwarded to *S7* (*F13*).
7. *S7* either confirms or deny the privileged request and send the response back to *S5* (*F14*).
8. *S5* gives a response to *S4* (*F12*).
9. *S4* queries the required information from *D1* (*F19*).
10. *D1* responds to *S4* with the requested information (*F20*).
11. *S4* makes the requested changes to the data (Add, Delete or Modify) and saves the changes inside *D1* (*F19*).
12. *S4* sends a change confirmation to the user client (*F10*).
13. *S3* relays the notice through S2 (*F8*).
14. *S2* relays the notice to *S1* (*F6*).
15. *E1* or *E2* gets the change confirmation/denial (*F2* & *F4*).

4.4.3.3 Booking system backup steps

1. *S8* sends a backup request to *D1* (*F21*).
2. *D1* responds to *S8* with the requested data (*F22*).

4.4.3.4 Booking system monitoring

1. *S9* requests the system status from *S4* (*F23*).
2. *S4* responds with the current system status to *S9* (*F24*).

4.4.3.5 Anti-virus update steps

1. *E1* and/or *E2* sends an update request to *S1* (*F1* & *F3*).
2. *S1* forwards the request to *O1* (*F29*).
3. *S10* picks up the update request coming from *O1* (*F26*).
4. *S10* sends back updated information (*F25*).
5. *S1* receives the information from *O1* (*F30*).
6. *E1* and/or *E2* receives the update from *S1* (*F2* & *F4*).

4.4.3.6 Device management steps

1. *E1* and/or *E2* initiates a status update/software update request (*F1* & *F3*).
2. *S1* forwards the request to *O1* (*F29*).
3. *S11* picks up the request coming from *O1* (*F27*).
4. *S11* sends back updated information (*F28*).
5. *S1* receives the information from *O1* (*F30*).
6. *E1* and/or *E2* receives the update from *S1* (*F2* & *F4*).

4.4.3.7 Internet access steps

1. *E1* and/or *E2* initiates a connection to *S1* (*F1* & *F3*).
2. *S1* forwards the connection to *O1* (*F29*).
3. *O1* the requested information is returned (*F30*).
4. *S1* relays the information the either *E1* or *E2* (*F2* & *F4*).

4.4.4 STRIDE analysis

Figure 4.3 shows the STRIDE analysis of the different components of both environments. The analysis was done collectively for all similar components because they have similar threats. For example, all services could experience denial of service and elevation of privileges. A benefit of this approach is that the threats stated in this study will be more generalized. Also, this study aims to evaluate the environments, not the individual components.

4.5 Cybersecurity metrics

A total of 20 cybersecurity metrics were defined and used in the case study. The inspiration for the metrics came from both the data flow diagrams created during the threat analysis and the Continuous Audit Metrics Catalog. The metrics will have a preferred value, either high or low, depending on the logical reasoning behind the metric.

Table 4.3 shows the metrics together with a short description of what each metric is supposed to measure.

Table 4.3: The metrics that were used to compare the two environments.

Metric ID	Metric	Preferred value	Inspiration
CSM1	Number of identity verification steps when logging into the environment.	High	Figure 4.1 & 4.2
CSM2	Percentage of users with MFA enabled (VPN and Booking system).	High	[24]
CSM3	Log rotation time.	High	[24]
CSM4	Number of layer 1 and layer 2 (OSI model) security mechanisms.	High	[24]
CSM5	Number of components that send unencrypted data (e.g. HTTP or non-VPN traffic).	Low	[24]
CSM6	Number of services inside the environment that are exposed to the Internet.	Low	Figure 4.1 & 4.2
CSM7	Percentage of data with confidentiality classifications stored in the environment.	High	[24]
CSM8	The number of data flows in the environment.	Low	Figure 4.1 & 4.2
CSM9	Number of availability assurance mechanisms (redundant systems/DDoS protection etc...).	High	[24]
CSM10	Average service uptime in the environment.	High	[24]
CSM11	Number of accounts that has privileged access in the environment.	Low	[24]
CSM12	The average time to patch a component in the environment after a new patch is available.	Low	[24]
CSM13	The average number of databases hosted on a single datasource.	Low	[24]
CSM14	Percentage of sensitive data-at-rest that is encrypted. in the environment.	High	[24]
CSM15	Number of components that harbor sensitive data.	Low	[24]
CSM16	Number of failed backup-jobs per month.	Low	[24]
CSM17	Average time to conduct vertical scaling of a component.	Low	[24]
CSM18	Number of components that are using certificates to validate connections.	High	[24]
CSM19	Number of trust zones/barriers in the environment.	High	Figure 4.1 & 4.2
CSM20	Number of operating system platforms in the environment.	Low	[24]

The motivation and baseline values for each metric are described below. The

Components	S	T	R	I	D	E
Ex	**T1**: Someone spoofs the identity of the user by stealing the authentication cookie. **T2**: Someone logs in with a leaked/stolen password.	-	**T3**: Someone hides their activity by altering user log data. **T4**: Someone performs actions from a rouge device.	-	-	-
Sx	**T5**: Someone sets up a rouge system, spoofing Sx inside network.	**T6**: Someone alters the traffic going to/from Sx.	**T7**: Someone alters the logs of Sx to hide their activity.	**T8**: Someone intersects communication between Sx and the user/system. **T9**: Someone accesses information from Sx through an exposed service. **T10**: Someone misplaces confidential data making it accessible to unauthorized users.	**T11**: Systems in the environment goes down from overload.	**T12**: Someone gets privileged access to a component.
Dx	-	**T13**: Someone makes changes to the database using non-standard methods.	-	**T14**: Someone accesses the database in a no-standard way to disclose information.	**T15**: Backup-jobs runs during work hours blocking normal database activity.	-
Fx	-	**T16**: Someone modifies data-in-transit in a main-in-the-middle attack.	-	**T17**: Someone views confidential information inside data-in-transit in a man-in-the-middle attack.	**T18**: Someone sends an excessive amount of data to a component.	-

Figure 4.3: The STRIDE analysis of the components that are present in both environments. $Ex = Entity(x)$, $Sx = System(x)$, $Dx = Datasource(x)$, $Fx = Flow(x)$.

baseline values were set in collaboration with the two cloud architects that performed the migration.

- CSM1 - More login steps to access the environment will reduce the risk of a successful brute-force attack. This metric will be represented as a numeric counter. No baseline value could be determined for this metric. The metric should not be too low because it will lower the number of steps a cybercriminal has to go through to highjack an account. It should not be too high because it will make the login process cumbersome for the actual users that need to log in. This measurement is a classic example of the CIA pyramid where you must balance availability towards integrity.

- CSM2 - More MFA-enabled users will reduce the risk of a user account compromise. This metric will be represented as a percentage counter because a percentage will represent the state of MFA usage even though the number of users increases between environments. The baseline value for this metric was set to 100%.

- CSM3 - If the rotation time is low there will be a high chance of missing any important activity. This metric will be represented by the number of hours because 1 hour is a reasonable amount of time for a log rotation to occur. When discussing this metric together with the two cloud architects, it was determined that there are few cases where logs are needed further back than 30 days. Therefore the baseline value for this metric is set to 30 days.

- CSM4 - Many security mechanisms in layers 1 and/or 2 of the OSI model will reduce the risk of someone spoofing a component inside the environment. This will be represented by a numeric counter. The baseline value for this metric was set to 1 because it is unknown how many layer 2 security mechanisms there are in an average environment.

- CSM5 - Many components sending clear text data will increase the risk of confidential data being intercepted and exposed. This will be represented by a percentage counter because the value had to be estimated by the SMEs. The baseline value for this metric is 0%.

- CSM6 - Many services exposed to the internet will increase the risk of an intrusion happening inside the environment. This will be represented by a numeric counter. The baseline of this value is 1 because to access the internet, one would need at least one firewall towards the internet.

- CSM7 - If a majority of the data inside the environment is classified, the risk of confidential data being accessed by unauthorized users will increase. This will be represented as a percentage counter because the value had to be estimated by the SMEs. No baseline value was defined for this metric.

- CSM8 - Fewer data flows in the environment will decrease the risk of data-in-transit interception. This will be represented as a numeric counter. No baseline value was defined for this metric.

- CSM9 - A higher level of availability assurance will reduce the risk of components failing inside the environment. This will be represented as a numeric counter. It is hard to set a baseline for this metric.
- CSM10 - If service uptime is high inside the environment the availability of those services will be higher. This will be represented as a percentage counter because that will be compatible with any defined service level agreement. A baseline for this metric is 99%.
- CSM11 - Fewer accounts with privileged access inside the environment will lower the risk of someone gaining privileged access to the environment through unconventional means. This will be represented as a numeric counter because the number of privileged accounts should not be that many. It is hard to set a baseline value for this metric because it is not favorable to share accounts amongst administrators either. It depends on how many employees there are that need this access.
- CSM12 - By patching systems quickly there is less risk that a vulnerability will be exploited by a malicious actor inside the environment. This metric will be represented in hours because it is not reasonable to think that a company can manually patch a vulnerable component in less than 1 hour after a vulnerability has been disclosed. The baseline value for this metric is 1-5 days.
- CSM13 - If fewer applications are using the same database server the risk of someone gaining access to the database through unconventional means will be lower. This will be represented as a numeric counter. The baseline for this value is a 1:1 ratio. It means that each database server should map toward one application.
- CSM14 - If much of the sensitive data is encrypted the risk of the data being leaked in a data breach will be lower. This will be represented as a percentage counter because the amount has to be estimated by the SMEs. The baseline value for this metric was set to 100%, however, the cloud architects were clear that they have never seen an environment that would have a 100% coverage in this metric.
- CSM15 - Having more components that harbor data-at-rest will increase the risk of data being accessed by unauthorized individuals. This will be represented as a numeric counter. It is hard to set a baseline value on this metric because it is not desirable to store important information on multiple nodes. At the same time, it can also be a tough decision to store all confidential data on one component.
- CSM16 - If many backup jobs fail, the risk of databases being unavailable during normal work hours will increase. This will be represented as a percentage counter because the number has to be estimated by the SMEs. The baseline value for this metric is set to 0%.
- CSM17 - If it takes a lot of time to scale a component the risk of an overload and DoS will be higher. This will be represented by minutes as it is reasonable

to assume that this is the smallest amount of time one could ask for to perform this kind of task. The baseline value for this metric was set to 1 minute because this is the least amount of time that one would require a trained technician to do the virtual scaling.

- CSM18 - If many systems in the environment are using certificates to validate connections the risk of a connection inside the environment being intercepted will be lower. This will be represented by a percentage counter because the number will have to be estimated by the SMEs. The baseline value for this metric was set to 100%.

- CSM19 - By separating the environment into multiple trust zones the risk of malicious data-in-transit will reach a trusted component will be lower. This will be represented as a numeric counter. There is no baseline value for this metric.

- CSM20 - By having less platform diversity in your environment you will decrease the patching overhead and the risk of having a lingering vulnerability. This will be represented as a numeric counter. There is no baseline for this metric.

A mapping of each cybersecurity metric and the threats they could measure is visible in table 4.4. Two metrics represented four threats, four metrics represented three threats, five metrics represented two threats and nine metrics represented one threat each. This becomes clearer in figure 4.4 that displays the reverse relationships found in table 4.4.

Table 4.4: The mapping between threats defined in the STRIDE threat analysis and their corresponding metric IDs.

Threat ID	**Metric ID**
T1	CSM1 & CSM19
T2	CSM2
T3	CSM3
T4	CSM4
T5	CSM4 & CSM18
T6	CSM18
T7	CSM3 & CSM11
T8	CSM6 & CSM8 & CSM18
T9	CSM6
T10	CSM7
T11	CSM8 & CSM9 & CSM10 & CSM17 & CSM19
T12	CSM11 & CSM12 & CSM19 & CSM20
T13	CSM11 & CSM12 & CSM13 & CSM19
T14	CSM11 & CSM12 & CSM13 & CSM14 & CSM15
T15	CSM16
T16	CSM4 & CSM8
T17	CSM5
T18	CSM9 & CSM17

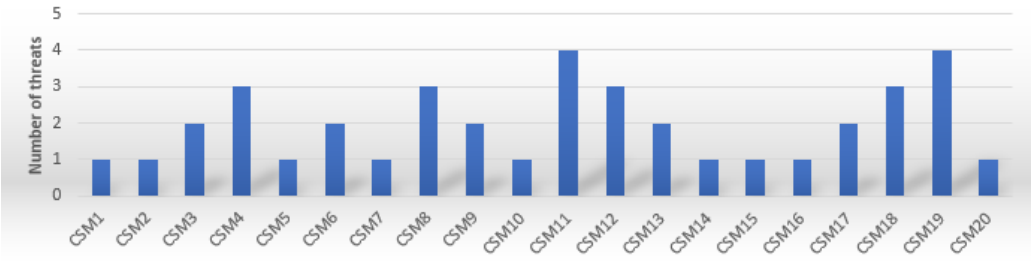

Figure 4.4: The number of threats from the STRIDE analysis that each metric represents.

4.5.1 Measurements

Some metrics were not possible to measure to an exact number. However, the cloud architects that performed the migration could estimate these metrics. Therefore the value of these metrics would be higher or lower in one of the environments compared to the other. To symbolize this, the values for these metrics were set to either a plus sign (+) or a minus sign (−). A plus sign would mean that the metric is higher than in the other environment. The minus sign means the metric is lower. It was a necessary adjustment, given the time allocated to this study.

Table 4.5 shows the values for each measurement. The cloud environment was the preferred choice in 12 (60%) of the metrics used in this study. The on-premise environment was the preferred choice in two (10%) metrics. There were six ties where both environments generated the same measurements. It means that the tie measurements represented 30% of all the metrics in the case study.

Table 4.6 lists results when using the weighted measurements. The weighted score of the cloud environment resulted in 28.5 points from a total score of 39 while the on-premise environment scored 10.5 points. The cloud environment scored $\approx 73\%$ of the available points. Therefore the score of the on-premise environment amounts to $\approx 27\%$ of all available points.

Table 4.5: The measured values for each cybersecurity metric defined in table 4.3

Metric ID	**On-Premises**	**Cloud**	**Preferred choice**
CSM1	6	18	Cloud
CSM2	−	100%	Cloud
CSM3	24	720	Cloud
CSM4	1	1	-
CSM5	80%	0%	Cloud
CSM6	1	4	On-Premises
CSM7	0	0	-
CSM8	22	30	On-Premises
CSM9	1	2	Cloud
CSM10	−	99.95%	Cloud
CSM11	+	−	Cloud
CSM12	30	30	-
CSM13	+	−	Cloud
CSM14	100%	100%	-
CSM15	2	2	-
CSM16	+	−	Cloud
CSM17	1	1	-
CSM18	−	+	Cloud
CSM19	3	5	Cloud
CSM20	1	0	Cloud

Table 4.6: The measurement results for the cloud environment.

CSM_m	p_m	w_m	w_{score}
CSM_1	1	1	1
CSM_2	1	1	1
CSM_3	1	2	2
CSM_4	0.5	3	1.5
CSM_5	1	1	1
CSM_6	0	2	0
CSM_7	0.5	1	0.5
CSM_8	0	3	0
CSM_9	1	2	2
CSM_{10}	1	1	1
CSM_{11}	1	4	4
CSM_{12}	0.5	3	1.5
CSM_{13}	1	2	2
CSM_{14}	0.5	1	0.5
CSM_{15}	0.5	1	0.5
CSM_{16}	1	1	1
CSM_{17}	0.5	2	1
CSM_{18}	1	3	3
CSM_{19}	1	4	4
CSM_{20}	1	1	1
SUM			28.5

Table 4.7: The measurement results for the on-premises environment.

CSM_m	p_m	w_m	w_{score}
CSM_1	0	1	0
CSM_2	0	1	0
CSM_3	0	2	0
CSM_4	0.5	3	1.5
CSM_5	0	1	0
CSM_6	1	2	2
CSM_7	0.5	1	0.5
CSM_8	1	3	3
CSM_9	0	2	0
CSM_{10}	0	1	0
CSM_{11}	0	4	0
CSM_{12}	0.5	3	1.5
CSM_{13}	0	2	0
CSM_{14}	0.5	1	0.5
CSM_{15}	0.5	1	0.5
CSM_{16}	0	1	0
CSM_{17}	0.5	2	1
CSM_{18}	0	3	0
CSM_{19}	0	4	0
CSM_{20}	0	1	0
SUM			10.5

Chapter 5

Analysis and discussion

5.1 Incident history and interviews

Some key points from the interviews are listed below. These were created based on the shared views of the answers given during the interviews.

- None of the participants could recall an incident that leveraged a multi-tenancy exploit.
- All agree that misconfigurations are the main concern regarding cloud security.
- Multi-tenancy can be linked to misconfigurations made by the CSP.
- Consumers are worried about how their data is handled in the cloud. However, data is generally more secure when stored in the cloud if done correctly.
- There is much psychology involved when a consumer chooses between an on-premise and cloud environment.
- There are many forms of environmental transparency. Knowing what is inside the environment is the one that matters.
- In some cases, cloud providers do not properly inform their cloud consumers about an incident. However, it is also up to the cloud consumer to configure proper logging of their service if they use PaaS.
- Cloud environments are more resilient to ransomware.
- Cloud incidents will be more common in the future. There will also be more zero-day exploits aimed at the cloud.
- Cloud- and on-premise environments are not necessarily selectively targeted by advanced cybercriminals.
- There is no specific distinction between on-premise and cloud environments concerning untargeted attacks. Misconfigurations on both ends are common. If a misconfigured service faces the internet, it can quickly be detected and exploited.

The literature- and dataset review results show that only four incidents were caused by cloud exploitation, which point to cloud exploitations are relatively rare. However, most of the reviewed incidents could have happened on either platform. In most cases, the platform just happened to be on-premise. One example is that it does not matter where one would run a vulnerable web application. It is especially true for cloud environments where web applications run on a PaaS solution. Several interviewees pointed out that even though something vulnerable is running in the cloud, it has more options in terms of incident response tools. An incident is bound to happen sooner or later. It is just a matter of time. It is why one decisive aspect regarding security is how the environment supports incident response.

5.1.1 Defense capabilities

All the interviewees agreed that a cloud platform would be the preferred choice for a consumer who is worried about getting hit by ransomware. A ransomware victim needs to have a tested and verified backup solution. It is often trivial in cloud environments while being more cumbersome on-premise. Therefore this is the primary factor that makes an environment based on cloud technology superior to an on-premise environment in terms of resilience against ransomware. Interestingly, the only cloud-related incident identified during the literature review (Elekta) was a ransomware attack. During this incident, the SaaS platform Elekta hosts in their cloud had to be taken down for some time while dealing with the attack. It shows that clouds are still vulnerable to ransomware, considering this event happened in 2021. Despite this, it is also important to note that many ransomware worms spread through disclosed operating system vulnerabilities. Because PaaS and SaaS platforms are patched by the CSP based on a SLA, this becomes less problematic in cloud environments. However, as one of the interviewees mentioned and proven in the Elekta incident, ransomware could affect anything. There is no guarantee that you are safe from the effects of a ransomware attack. It all comes down to how well the cloud consumer has prepared themselves against that scenario. The key argument is there are more options available when using a cloud environment. The capabilities to handle cybersecurity incidents could be one factor that affected the results of the review studies conducted in this thesis.

5.1.2 Incident transparency

It is important to note that the sources used in the reviews only consisted of disclosed incidents of successful breaches. Even though cloud environments are the subject of multiple attacks, the majority never cause any critical harm to their victim. Another reason why there are so few studied cloud incidents in the past is that the CSPs share little or no information about the incidents that occur in their environments. It brings up the issue of transparency between the CSP and their cloud consumers. However, there are many forms of environmental transparency. If a cloud consumer wants to know what processes are running on the operating system of a PaaS instance, it might not be possible. If this is important, the cloud consumer might want to reconsider migrating to a cloud environment. But this should not be important to cloud consumers unless they have the necessary capabilities to use such information.

There should be enough resources inside the company to act on an incident. If this is not the case, there is no need for this kind of environment transparency. Environment transparency is also related to asset management. In this case, most interviewees agreed that a cloud consumer would be better off using a cloud platform. This result might be unexpected because, in an on-premise environment, you own all the equipment and software that runs inside that environment. Therefore you can monitor these resources in greater detail inside an on-premise environment. However, the interviewees argue that one can find out what is in a corporate cloud environment from a central management interface. In an on-premise environment, one might have to perform network scans to see what is there. When considering a PaaS solution, one will still have the capability to see what is going on inside the application, given that the cloud consumer has configured the logs correctly.

It is also important to note that the ENISA threat report mentions that there has been a spike in non-malicious cloud-related incidents recently. It could have contributed to the result of the reviews because these incidents might not be reported to the cloud consumer and therefore never end up in a public dataset or research paper. However, the ENISA threat also mentions that clouds were the most frequent type of environment in data breaches in 2021 [45]. One explanation for why the result from the literature- and dataset review differ from the results of the ENISA threat report is that most of the incidents related to cloud exploitation happened in recent years. Therefore the majority of these incidents might not have been disclosed yet. It is likely more scientific reports and datasets in the future will contain more examples of incidents caused by cloud-related exploits.

5.1.3 Lack of competence

Several interviewees mentioned cloud consumers lack competence in managing their services in the cloud. The ENISA report brings up the same problem. One of the interviewees points out that CSPs are obligated to provide best practice recommendations to their cloud consumers. It means that either the best practice guidelines from the CSPs are complex and hard to understand, or many cloud consumers ignore information regarding best practices. Many cloud consumers do adhere to best practices which means that the second scenario is more likely. It became evident in the middle of the global pandemic when all companies and organizations needed to adjust their services for teleworkers. Both the interviewees and the ENISA threat report agree that mistakes made on the consumer side during these migrations enabled cybercriminals to access services with unintended settings. Therefore this recent trend does not confirm that services hosted in cloud environments are fundamentally less secure than those hosted on on-premises. It is an effect of cloud consumers not adhering to best practices. This problem is not unique to the cloud. It also includes consumers that use on-premise environments where this has been a problem for a long time. One of the interviewees stated that cloud platforms give users better tools to implement best security practices. If this is true, it should enable many cloud consumers to mitigate a considerable amount of exploits. However, it looks like these incidents still occur.

The ENISA report also mentions that cybercriminals are starting to show exceptional knowledge of cloud environments. One of the interviewees pointed out that

cybercriminals traditionally know more about on-premise technology than cloud technology. It is another factor that could have contributed to the result of this study. However, it is dangerous to rely on cybercriminals sticking to what they know.

5.1.4 Causes for cybersecurity incidents

It was discovered during the study that misconfigurations are a big problem for cloud environments. All of the interviewees agreed on this. What separates a misconfiguration inside a cloud environment from a similar error in an on-premise environment is that the consequences could be more severe. A misconfiguration made by the CSP could affect all tenants of the cloud environment. The interviewees mention that misconfigurations done by the CSP are the leading cause of issues involving multi-tenancy. Zero-day exploits that enable a malicious user to access multiple tenants in a cloud environment are rare. At least they are rarely exploited in the wild. Cloud service providers put a lot of resources into solving this issue. In fact, none of the participants had experienced an incident where a cybercriminal compromised multiple tenants inside a cloud environment this way. The results from the reviews made in this study also point this out. However, multi-tenancy should still be regarded as a relevant problem when choosing an environment because public CSPs allows anyone to create an account. It includes cybercriminals. The possibility of sharing cloud services with cybercriminals could be why an on-premise environment would feel more secure than a cloud alternative. One interesting aspect that one of the interviewees mentioned was that these hybrid environments are the ones that one should avoid. A hybrid environment is a mixture of different technologies that connect an on-premise environment to the cloud. These hybrid setups often require high privilege services in both environments which creates a new layer of threats. A hybrid environment also raises the risk of misconfigurations because it adds a new level of complexity that can be hard to maintain and protect from malicious intent.

More companies may have had their services running on-premises than in the cloud. If this is true, there may be skewed data in this thesis. However, one security aspect speaks for the result of this study; patching. Patching an on-premise environment is tedious work. It is either done by a MSP or the company themselves. By using a cloud alternative such as PaaS or SaaS, the patching overhead for the operating system is relieved from the companies that use the services. In the case of SaaS platforms, the cloud consumer does not have to worry about application patching either. Zero-day exploits were not the leading cause for most of the incidents in this thesis. Instead, disclosed vulnerabilities such as EternalBlue were the leading cause for most incidents in this study. A cloud platform can be the preferred choice if a consumer wants to avoid similar incidents. Therefore it is hard to say that the result of the literature study is simply a coincidence.

5.1.5 Physiological factors of environments

One other factor that speaks for the outcome of the review studies is that CSPs have more capital to invest into their security posture. Several interviewees mentioned this and circled back to this argument to answer multiple questions. However, public

cloud providers have a larger attack surface than a traditional on-premise environment. It means that CSPs have to invest more capital, attract talented employees and stay at the forefront of the cybersecurity industry. One of the interviewees that have worked as an IT manager for many years also mentioned that there is an exaggerated belief that MSPs are more competent in managing an IT environment than they really are. One of the interviewees mentioned there is much psychology involved when a consumer chooses their outsourcing partner. Especially when it comes to choosing between a MSP or a CSP. A consumer might feel more welcome and appreciated if they sign a contract with a small MSP. However, because this decision might not be the most rational one, it could impact the security aspects.

The interviewees agree that consumers should consider data management when choosing a platform. The interviewees mention that it can be vital for a cloud consumer to know the exact location of the data centers that host their services. It can be problematic because public CSPs only allow cloud consumers to choose a specific region to store their data, not a country or a data center. One example is that different countries have different laws concerning what they can and cannot do with data stored in that country. Compliance aside, the interviewees agree that storing data in the cloud is secure and can be preferred if the alternative is an on-premise environment. By storing data in the cloud, a consumer will have high availability of that data and reliable backup/restore capabilities. The analogy that it is safer to travel by plane even though you feel safer in a car puts this into perspective. By placing data inside a fortress instead of hiding it in your backyard, it will be more protected.

5.1.6 Cybercriminal preference

When it comes to targeted and untargeted attacks, there is little difference between on-premise and PaaS solutions. The difference between the environments is when the malicious actor target disclosed operating system vulnerabilities in an untargeted attack. The Wannacry worm is a prime example of where unpatched on-premise systems still can be hit today if they have not yet installed the patch for the Eternalblue exploit.

The interviewees were not convinced a specific environment could "scare" away a malicious actor. However, this depends on the threat actor and their objective. A cybercriminal looking for quick money would probably move to their next target if they found an environment that would take a long time to breach. An APT on the other hand would not give up until they obtained their objective. The interviewees also mentioned that cybercriminals often go after the users of a specific system instead of the system themselves. One interviewee argues that social engineering is the easiest way of accessing an environment because it allows a cybercriminal to circumvent all protection.

5.1.7 Future developments

All the interviewees agree that more cloud-related incidents will happen in the future. Two of them described what is now happening as a transition from petty theft from isolated environments to supply-chain attacks that affect many consumers at once.

As they mentioned, the similarities between this and banking are uncanny. Instead of stealing money from one person, a criminal target a bank to steal money from more people. Therefore, just as in the early days of banking, it would not be surprising if more criminals would target public CSPs instead of lone organizations using on-premise technologies. APTs would be excluded from this prediction because they are more focused on stealing information from a specific target. Even though this could increase the risk of using cloud services in the future, it is vital to remember that even if there is a lot of money to be made by robbing banks, robbing banks is not the primary choice today for criminals if they want to make money.

5.2 Case study

From the results, it is clear that the cloud environment was considered the most secure one. After discussions with the two cloud architects that performed the migration, it was clear that the results from the measurements are not always black and white.

5.2.1 Validity of measurements

Because CSM1 only affects one of the threats identified during the threat model analysis, it might not be the most vital metric. However, in combination with CSM2, it can be one of the most important metrics because it also considers aspects of threats regarding social engineering. If an employee receives a spam e-mail and gets their password stolen, the company can still be safe if they have implemented robust login steps such as MFA. CSM2 is important as a standalone metric because it provides another dimension in keeping the environment safe from account highjacking. If the value is not 100%, the accounts that do not use MFA become potential targets. A value of 99% is enough to jeopardize the whole environment. It is especially true if that 1% is one or more privileged account(s). Therefore, additional dimensions should be defined to increase the accuracy of this metric. This problem does not invalidate the result of this study, but it should be considered in the future.

Not much information could be found regarding CSM4. Layer-2 security is a vital aspect of environment security, so the metric itself is not irrelevant. However, few layer-2 security mechanisms exist in a cloud environment known to a cloud consumer. For CSM4 to be a decisive metric when comparing on-premise and cloud environments, the cloud consumer has to consult the CSP. Unfortunately, there was no time to do so in this thesis.

The CSM5 metric was a complement to CSM18. Certificates form secure communication channels in an environment, but their primary use is to identify the server/service as authentic. All components within the cloud relied on HTTPS for communication. Therefore the measurement became 0% for the cloud environment. The cloud architects that performed the migration estimated that 80% of the components inside the on-premise environment was using certificates.

Because most cloud services were exposed to the internet, the on-premise environment became the preferred choice in CSM6. In a discussion with the cloud architects, it was determined that this metric could be misleading. Some of these cloud services

were segmented from the other services used by the customer. Unfortunately, this was hard to illustrate during the threat model process. A more detailed DFD should mitigate this problem.

CSM7 was interesting because neither cloud architect has seen a customer using confidentiality classifications on all their files. However, they agree that the metric is vital and sound. This metric is hard to satisfy because most organizations have only recently started to work with data classification.

CSM8 was also interesting because it measures the complexity of the environment. The two cloud architects agree that a complex environment is hard to secure. However, there are many forms of environmental complexity. This metric could only measure one of them. For example, it may be more complex to manage services in an on-premise environment because you have no centralized console where you can monitor the status of the environment as you would have in a cloud environment.

Measuring the number of availability assurance mechanisms proved to be a difficult task. It means that CSM9 should be reevaluated. The metric is sound in the way that it tries to measure how well an environment can stand up against a DDoS attack. However, just because both environments have DDoS protection does not mean that both of them perform equally. Therefore, the capacity of these mechanisms should be measured instead of how many of them there are. Such a measurement would require information that is not easy to obtain.

The values for CSM10 were based on the SLA levels provided by the CSP of the cloud environment. The on-premise environment did not have an SLA which is a problem in itself. Therefore, no service up-time was guaranteed before the migration. It was enough to flip the favor of this measure to the cloud environment. Unfortunately, the subject in the case study did not keep records of how often their services were disrupted in each environment. It was why the measurement had to be based on the SLAs.

One of the metrics that measured most of the threats was CSM11. Therefore it was one of the most vital metrics in this study. However, the PIM cloud feature contains this problem even if there are a high number of privileged accounts. It introduces a new element into this measurement which further speaks for a cloud environment being the preferred choice.

One constraint with CSM12 is that, before applying a patch, the consumer needs to verify the patch works as expected. It is a time-consuming procedure that often requires a test environment and days of testing. Therefore it is not reasonable to assume that patching a component can be done sooner than 1-5 days. Because the booking system was migrated to a PaaS instance, application level patches still had to be applied by the consumer. It is why both environments share the same score in this metric. However, the metric does not measure how many systems the consumer needs to patch. The CSP applies the patches for the operating system of the PaaS instance. It means that the patching overhead for the consumer was reduced post-migration. The consumer can then focus on keeping the application itself up to date. However, it is still important to measure the time it takes for an organization to patch its systems. Therefore this metric should be considered when performing an environment comparison.

The cloud architects mentioned that many aspects could jeopardize an IT environment regarding databases. Such claims validate the relevancy of CSM13. One

example is two databases hosted on the same database server for two different applications. If one of these applications requires a specific old version of the database, it could put the other database at risk. Another aspect to consider when comparing a cloud and on-premise environment is that databases run on isolated instances in the cloud. It creates conditions where the metric reaches the baseline value right away.

Some aspects need to be considered regarding the result of CSM14. The backup solution used in the on-premise environment needed a separate license for encrypting data while this functionality was native in the cloud environment. Therefore there is an additional cost to using this feature in the on-premise environment. Because the cost of keeping the data secure is not something that is considered, both environments get the same score in this metric.

When it comes to CSM15, both environments had the same amount of components that harbored sensitive data. These components were the AAA service and the booking service. However, other components could store sensitive data as well. It depends on the definition of sensitive data. In this case study, personally identifiable information was the definition of sensitive data. If the value of this metric is above 1, the environment should have a 100% coverage on CSM14.

The cloud architects did not know how many backup jobs failed per month but estimated CSM16 to be less in the cloud environment. The argument is that backup functionality is built into the cloud environment and fully supports other cloud services. In the case of on-premise environments, there is an overhead regarding the operating system that runs the backup service. If there is a problem with the operating system, the backup jobs will also fail, disturbing the everyday actions of users in the environment.

It is hard to measure CSM17 because it depends on the human factor. It is vital to show how the different environments react to a sudden increase in activity. Even though the result is a tie, the cloud architects mention that cloud environments offer many solutions for automatic scaling. Automatic scaling decreases the downtime of an environment when the needs of that environment change. This metric was closely related to CSM10.

During the study, it became clear that there are other uses for CSM18 than just identifying how many components use secure connections in the environment. Old devices and services are not likely to use secure communication such as HTTPS. Because of this, CSM18 also identifies how many outdated services there are in an environment. It could be a vital complement to CSM12. Just because a system does not have a patch available does not mean that the system is secure. Therefore CSM18 could also identify components in an environment that have passed their due date.

From the threat analysis it became evident that the number of trust zones in each environment was different. CSM19 shows that the cloud environment provides another level of protection from malicious traffic since there is a second firewall inside the cloud. The metric signifies which environment has the most capabilities for a defense in depth approach. In this case, it was the cloud environment. What is also interesting is that even though the cloud environment has more trust barriers that can catch malicious traffic, many services in the cloud communicate with the users through the internet. It means cloud services circumvent the additional trust barriers when communicating with the users. In this sense, the cloud- and on-premise

environments have the same number of trust zones. But in the context of the booking system, the additional trust zones apply.

CSM20 was also an interesting metric. It was set to zero for the cloud environment because there are no operating systems for the cloud consumer to manage. It is closely related to the patching overhead inside the environment (CSM12). It would have been interesting to measure the diversity in the cloud backbone. Unfortunately, this was not possible in this study because there was no direct contact with a cloud technician of the CSP that hosted the cloud environment. It is still important to note the results from the study reviews and interviews showed that multi-tenancy and cloud backbone exploits are rare. Therefore, measuring the environment diversity from the consumers' perspective will have a more profound impact on the security of the environment.

5.2.2 Validity of the methodology

Because the comparison methodology used in this case study relied on a novel approach to measuring risks within an environment, the results should be discussed. The methodology bases its assessment on the quantitative aspects. Therefore, it ignores that some risks or threats are more serious than others. Its results should be regarded as an alternative way of measuring risks inside an environment.

The weighted score did help balance the measurements because it showed that the on-premise environment was the preferred choice in some important metrics. Therefore it helped making the result more fair where the metric that measures the most threats are in high regard.

5.3 Threats to validity

5.3.1 External validity

Because not all cyberattacks are disclosed to the public, the literature and dataset review does not represent all forms of cybersecurity incident/breaches. However, the focus of this study was on disclosed cyberattacks. Therefore, undisclosed cyberattacks were out of scope for this study.

In this study, the threat model methodology used an example on-premises environment that might have been less hardened than a typical on-premises environment. Therefore, it is hard to say if the methodology works for all types of environments. This problem could have been mitigated by doing multiple case studies. However, there was limited time and resources to finish this thesis, so there was only room for one case study. The limited number of studied cases is not a problem because the comparison methodology applies to any environment. The facts that the samples for the comparison methodology were real environments also contribute to the general validity of this thesis.

Another important detail is that the dataset used in this study only contained incidents concerning US-based companies. It could have affected the outcome of the dataset review because it does not take any other companies into account. Therefore the result from the dataset review itself cannot be fully generalized. However, it is

still a decent complement to the literature review because it gives a broad view of the incidents inside a country that has experienced the most cybersecurity incidents in the world during the last two decades [105].

There is also a problem that the proposed comparison methodology will be export-dependent because of the way a STRIDE analysis works. It might also affect the general applicability of the methodology as it will perform better or worse depending on the practitioner. The threats defined during the STRIDE analysis are only as thorough as the practitioner that performs the analysis. Therefore, the methodology is limited to cybersecurity professionals and might be incomplete if anyone uses it that has no prior knowledge in the cybersecurity field. However, there is no professional requirement to interpret the result of the methodology. It will still be possible for a CEO or board member with no previous cybersecurity experience to interpret the result as they are simply numbers that are either high or low.

5.3.2 Internal validity

Regarding the literature review, cybercriminals may favor on-premise environments for a different reason than their inherent security flaws discussed in this thesis. One example could be there is more confidential information stored in on-premise environments, and they are therefore more targeted. However, many incidents studied in this thesis could have been avoided if the victim had used a cloud alternative. Therefore it is reasonable to assume the result of the literature- and dataset review is sound.

One of the interviewees pushed the issue that CSPs does not always disclose incidents to their cloud consumers, despite that CSPs are obligated to do so according to ISO 270001. It means that cloud providers could keep some incidents from their customers. It will result in fewer cloud-related incident disclosures. If this is true, then it would be a decisive factor that comes into play concerning the reviews in this study.

It is also important to note that there might have been bias in the answer gained from the participants as well. The interviews were designed to be open to mitigate this. By having open interviews, the interviewer can provide counterarguments and argue against the answers to find any obvious biases.

5.3.3 Construction validity

There have been previous attempts to extract information from incident databases. M.-S. Pang and H. Tanriverd mention in their study that the information inside publicly available datasets is insufficient to determine what caused the cybersecurity incident to happen [86]. This is why information about the incidents had to be researched manually through different sources such as peer-reviewed articles, blogs, and news articles with reliable sources.

5.3.4 Conclusion validity

The case study never revealed how frequent or severe each threat would be. It means that even though one environment excels by having a better value of a particular

metric that measures a particular threat, it is left to the practitioner to decide on which of the threats are most important to the environment. Therefore an expert opinion is still needed to apply the threat model methodology if the severity of the threats has to be considered. However, formula 3.3 provides a solution for this in regards to measuring the importance of the metrics instead.

Chapter 6

Conclusions and Future Work

6.1 RQ1

To what extent have cybercriminals exploited on-premises environments compared to cloud environments between 2008 and 2021? - From a historic perspective, on-premise environments have been more exposed to cyberattacks than cloud environments. The results from the literature and dataset review in conjunction with the interviews point out that there have been few incidents and campaigns concerning cloud exploitation. However, recent reports, such as the ENISA threat report, have pointed out cybersecurity incidents against cloud infrastructures have increased. Even though more cloud-related incidents (especially supply-chain attacks) are expected in the future, this sudden spike in cloud-related incidents can be explained by the sudden need for teleworker support during the COVID19 pandemic combined with careless migrations.

6.2 RQ2

Which cybersecurity metrics are appropriate to use when comparing on-premises and cloud (Paas/SaaS) environments? - 20 cybersecurity metrics were defined during the study. Some metrics were based on the DFDs that were created in the STRIDE threat analysis. Other metrics were inspired by external sources such as the literature review analysis and the continuous audit metrics published by the CSA.

The following metrics were considered to be appropriate for comparing two IT envinroments to each other in regards to security:

- CSM1
- CSM2
- CSM3
- CSM5
- CSM6
- CSM7
- CSM8

- CSM10
- CSM11
- CSM12
- CSM13
- CSM14
- CSM15
- CSM16
- CSM18
- CSM19
- CSM20

The metrics that were not as helpful when performing the comparison are listed below:

- CSM4
- CSM9
- CSM17

The intent of these metrics still serves a purpose. It is the measurements themselves that should be evaluated.

6.3 RQ3

How are these cybersecurity metrics affected in an infrastructure migration from a on-premise environment to a cloud environment? - The study shows that most of the metrics were useful when determining the state of security in an IT environment. The case study concluded the cloud environment was the preferred choice. During the thesis, it became evident that it is hard to make accurate measurements based on quantitative data alone. The weighted measurements balanced the metrics to each other. However, because these metrics only consider quantitative aspects, they ignore the severity of the threat they measure. A severity level for the threats could solve the problem. These security levels would be a measurement based on qualitative data. The method proposed in this thesis considers one vital aspect of measuring security in an environment. However, a qualitative assessment is still needed to give a complete picture of the state of security inside an IT environment.

6.4 Concluding remarks

The literature- and dataset review show that on-premise environments have been the primary target of cybercriminals because this is what they know best. The majority of incidents studied in this thesis happened between 2010 to 2018. Experts interviewed during this thesis and recent threat reports suggest that cloud environments have recently been a target for cybercriminals.

The competence of cybercriminals in terms of cloud exploitation is on the rise. Therefore cloud-related incidents will be more frequent in the future. Cybercriminals need to adapt because it is in the cloud where future data will be stored. Because of this, the cloud will be an attractive target for future cybercriminals. It might be something to consider as a cloud consumer.

The primary reason for the recent increase in cloud exploitation is the number of misconfigurations done when companies had to migrate their environments during the pandemic. It means that the sudden rise in incidents could be temporary. The results from the reviews are more relevant as they describe the relationship between the environments in their natural state. Because misconfigurations of cloud services are the primary cause of cloud-related incidents, the consumers also need to increase their competence.

Cybersecurity metrics are not always black and white. The method used in this study did try to divert from qualitative assertions and determine the security of an environment solely based on quantitative data. The conclusion from using the methodology is that it is hard to assess the security of an IT environment by only using quantitative data. The reason is that there are nuances in the measurements that can be hard to express in numbers. Therefore, a measurement in favor of one environment could be due to coinciding factors.

The methodology presented in this study is a decent first step when determining the security posture of an IT environment. Based on the results of this thesis, this method could be used by a company or organization that is considering a migration from on-premise to the cloud. It is vital to mention that a decision cannot be based entirely on this methodology. However, it will show how the actual security posture of the environment will change post-migration. Therefore the proposed comparison methodology could help companies make decisions that favor the security of their environment.

6.5 Future works

The external validity of the comparison methodology could increase by conducting more case studies. One interesting case study would be a hybrid environment migration to see if the results would agree with the expert opinions of this study that hybrid environments are a less secure alternative.

The metrics that were not as helpful in this study as the rest should also still be considered in future studies. Extending the number of metrics and dividing them into different categories would also be an intriguing continuation of this study.

A hybrid comparison methodology could also be a subject for future works considering that the current methodology does not consider the severity of the threats.

This hybrid approach could be based on quantitative measurements while taking advantage of qualitative aspects. The accuracy of the methodology could be increased by letting the practitioner decide the severity of the threat that has been found in the threat model process. It would also be possible to introduce MITRE ATT&CK scores as qualitative weights. Both these suggestions would mean that formula 3.3 has to be adjusted.

References

[1] G. Abbiati, S. Ranise, A. Schizzerotto, and A. Siena, "Merging datasets of cybersecurity incidents for fun and insight," *Frontiers in Big Data*, vol. 3, 2021. [Online]. Available: https://doi.org/10.3389/fdata.2020.521132

[2] Acronis, "The nhs cyber attack: how and why it happened, and who did it," https://www.acronis.com/en-gb/articles/nhs-cyber-attack/, (Accessed on 04/04/2022).

[3] I. Agrafiotis, J. R. C. Nurse, M. Goldsmith, S. Creese, and D. Upton, "A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate," *Journal of Cybersecurity*, vol. 4, no. 1, 10 2018. [Online]. Available: https://doi.org/10.1093/cybsec/tyy006

[4] Y. Ahmed, S. Naqvi, and M. Josephs, "Cybersecurity metrics for enhanced protection of healthcare it systems," in *2019 13th International Symposium on Medical Information and Communication Technology (ISMICT)*, 2019, pp. 1–9. [Online]. Available: https://doi.org/10.1109/ISMICT.2019.8744003

[5] S. Alder, "Rady children's hospital reports 14,000-record hipaa breach," https://www.hipaajournal.com/rady-childrens-hospital-reports-14000-record-hipaa-breach/, 06 2014, (Accessed on 03/30/2022).

[6] ——, "Lsu health laptop theft exposes phi of at least 5,000 minors," https://www.hipaajournal.com/lsu-health-laptop-theft-exposes-phi-of-5000-minors-8104/, 09 2015, (Accessed on 03/30/2022).

[7] ——, "Cyberattackers demand $3.6m ransom from hollywood hospital," https://www.hipaajournal.com/cyberattackers-demand-3-6m-ransom-from-hollywood-hospital-8313/, 02 2016, (Accessed on 04/04/2022).

[8] ——, "Radiation treatments disrupted after cyberattack on software vendor," https://www.hipaajournal.com/healthcare-providers-postpone-radiation-treatments-cyberattack-elekta/, 04 2021, (Accessed on 04/04/2022).

[9] R. Alkhadra, J. Abuzaid, M. AlShammari, and N. Mohammad, "Solar winds hack: In-depth analysis and countermeasures," in *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2021, pp. 1–7. [Online]. Available: https://doi.org/10.1109/ICCCNT51525.2021.9579611

[10] L. Allodi and F. Massacci, "Security events and vulnerability data for cybersecurity risk estimation," *Risk analysis*, vol. 37, no. 8, pp. 1606–1627, 2017. [Online]. Available: https://doi-org/10.1111/risa.12864

[11] A. K. AlSalamah, "Security risk management in online system," in *2017 IEEE ACIT-CSI*. IEEE, 2017, pp. 119–124. [Online]. Available: https://doi.org/10.1109/ACIT-CSII-BCD.2017.59

[12] V. V. Arutyunov, "Cloud computing: Its history of development, modern state, and future considerations," *Scientific and Technical Information Processing*, vol. 39, no. 3, pp. 173–178, 2012. [Online]. Available: https://doi.org/10.3103/S0147688212030082

[13] T. August, M. F. Niculescu, and H. Shin, "Cloud implications on software network structure and security risks," *Information Systems Research*, vol. 25, no. 3, pp. 489–510, 2014. [Online]. Available: http://www.jstor.org/stable/24700307

[14] AWS, "Shared responsibility model - amazon web services (aws)," https://aws.amazon.com/compliance/shared-responsibility-model/, (Accessed on 05/02/2022).

[15] C. Ballod, "Blue beacon breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/blue-beacon-20180406.pdf, 04 2018, (Accessed on 03/29/2022).

[16] M. Berndtsson, J. Hansson, B. Olsson, and B. Lundell, *Thesis projects: A Guide for Students in Computer Science and Information Systems*, 2nd ed. Springer, 2008.

[17] J. Beyer, "Adam shostack on threat modeling," *IEEE Software*, vol. 37, no. 6, pp. 110–112, 2020. [Online]. Available: https://doi.org/10.1109/MS.2020.3017406

[18] K. S. Bhosale, M. Nenova, and G. Iliev, "A study of cyber attacks: In the healthcare sector," in *2021 Sixth Junior Conference on Lighting (Lighting)*, 2021, pp. 1–6. [Online]. Available: http://doi.org/10.1109/Lighting49406.2021.9598947

[19] K. bok Lee and J. in Lim, "The reality and response of cyber threats to critical infrastructure: A case study of the cyber-terror attack on the korea hydro & nuclear power co. ltd." *KSII Transactions on Internet and Information Systems*, vol. 10, no. 2, pp. 857–880, 2016. [Online]. Available: http://dx.doi.org/10.3837/tiis.2016.02.023

[20] S. Brown-Philpot, "Sample notice_23.pdf," https://oag.ca.gov/system/files/Sample%20Notice_23.pdf, 05 2018, (Accessed on 03/30/2022).

[21] R. Buyya, S. N. Srirama, G. Casale, R. Calheiros, Y. Simmhan, B. Varghese, E. Gelenbe, B. Javadi, L. M. Vaquero, M. A. S. Netto, A. N. Toosi, M. A. Rodriguez, I. M. Llorente, S. D. C. D. Vimercati, P. Samarati, D. Milojicic, C. Varela, R. Bahsoon, M. D. D. Assuncao, O. Rana, W. Zhou, H. Jin, W. Gentzsch, A. Y. Zomaya, and H. Shen, "A manifesto

for future generation cloud computing: Research directions for the next decade," *ACM Comput. Surv.*, vol. 51, p. 1–38, 2018. [Online]. Available: https://doi-org/10.1145/3241737

[22] J. Buzzio-Garcia, V. Salazar-Vilchez, J. Moreno-Torres, and O. Leon-Estofanero, "Review of cybersecurity in latinamerica during the covid-19 pandemic. a brief overview," in *2021 IEEE ETCM*. IEEE, 2021, pp. 1–5. [Online]. Available: https://doi.org/10.1109/ETCM53643.2021.9590693

[23] I. D. Ceukelaire, "This popular facebook app publicly exposed your data for years | by inti de ceukelaire | medium," https://medium.com/@intideceukelaire/this-popular-facebook-app-publicly-exposed-your-data-for-years-12483418eff8, 06 2018, (Accessed on 03/30/2022).

[24] J. L. Christopherson, W. Fabritius, R. Krishnamurthy, D. Catteddu, K. Murphy, A. Pannetrat, C. Pedigo, M. Platt, M. Pritikin, A. Scarfe, and C. Victoria, "The continuous audit metrics catalog | csa," https://cloudsecurityalliance.org/artifacts/the-continuous-audit-metrics-catalog/, (Accessed on 11/21/2021).

[25] C. Cimpanu, "Data from two muslim dating sites, two others dumped online," https://news.softpedia.com/news/data-from-two-muslim-dating-sites-two-others-dumped-online-506356.shtml, 07 2016, (Accessed on 03/30/2022).

[26] ——, "Mongodb server leaks 11 million user records from e-marketing service | zdnet," https://www.zdnet.com/article/mongodb-server-leaks-11-million-user-records-from-e-marketing-service/, 09 2018, (Accessed on 03/30/2022).

[27] W. G. Cochran, *Experimentation in software engineering*. John Wiley & Sons, 1977. [Online]. Available: https://archive.org/details/Cochran1977SamplingTechniques_201703/page/n89/mode/2up

[28] R. Courtney, "Fresno state data breach exposes personal information of 15,000 people - abc30 fresno," https://abc30.com/fresno-state-data-breach-exposes-personal-information-of-15000-people/3182146/, 03 2018, (Accessed on 03/30/2022).

[29] CSA, "State of cloud security risk, compliance, and misconfigurations | csa," https://cloudsecurityalliance.org/artifacts/state-of-cloud-security-risk-compliance/, (Accessed on 02/25/2022).

[30] ——, "Star registry | csa," https://cloudsecurityalliance.org/star/, 2009, (Accessed on 11/21/2021).

[31] J. Davis, "Long island provider exposes data of 42,000 patients in misconfigured database | healthcare it news," https://www.healthcareitnews.com/news/long-island-provider-exposes-data-42000-patients-misconfigured-database, 03 2018, (Accessed on 03/30/2022).

[32] N. Dissanayake, A. Jayatilaka, M. Zahedi, and M. A. Babar, "Software security patch management - a systematic literature review of challenges, approaches,

tools and practices," *Information and software technology*, vol. 144, p. 106771, 2022. [Online]. Available: https://doi.org/10.1016/j.infsof.2021.106771

[33] DISSENT, "Ca: Mailing error in newfoundland and labrador reported to privacy commissioner," https://www.databreaches.net/ca-mailing-error-in-newfoundland-and-labrador-reported-to-privacy-commissioner/, 07 2015, (Accessed on 03/30/2022).

[34] ——, "Database leak exposes uncle maddio's employees' and customers' info – or did it? (updated)," https://www.databreaches.net/database-leak-exposes-uncle-maddios-employees-and-customers-info/, 12 2015, (Accessed on 03/30/2022).

[35] ——, "Two apps with health info found leaking: researcher. part 1: ifit," https://www.databreaches.net/two-apps-with-health-info-found-leaking-researcher-part-1-ifit/, 12 2015, (Accessed on 03/30/2022).

[36] ——, "Dozens of clinics, thousands of patients impacted by third-party data leak," https://www.databreaches.net/dozens-of-clinics-thousands-of-patients-impacted-by-third-party-data-leak/, 09 2016, (Accessed on 03/30/2022).

[37] ——, "Eye associates of pinellas notifying 87,000 patients of bizmatics breach (update2)," https://www.databreaches.net/eye-associates-of-pinellas-notifying-87000-patients-of-hackingit-incident/, 05 2016, (Accessed on 03/30/2022).

[38] ——, "Henry county residents' information is exposed in hacking," https://www.databreaches.net/henry-county-residents-information-is-exposed-in-hacking/, 12 2016, (Accessed on 03/30/2022).

[39] ——, "Walmart vendor error exposed limited patient information," https://www.databreaches.net/wal-mart-vendor-error-exposed-limited-patient-information/, 06 2016, (Accessed on 03/30/2022).

[40] ——, "Aflac says agent emails were hacked, exposing personal information of clients," https://www.databreaches.net/aflac-says-agent-emails-were-hacked-exposing-personal-information-of-clients/, 05 2018, (Accessed on 03/30/2022).

[41] ——, "Fasthealth notifies patients of 2017 incident involving their data," https://www.databreaches.net/fasthealth-notifies-patients-of-2017-incident-involving-their-data/, 02 2018, (Accessed on 03/30/2022).

[42] ——, "Homeapplicationsxgimp & maxipdf apps leak thousands of private photos and docs online xgimp & maxipdf apps leak thousands of private photos and docs online," https://www.databreaches.net/homeapplicationsxgimp-maxipdf-apps-leak-thousands-of-private-photos-and-docs-online-xgimp
02 2018, (Accessed on 03/30/2022).

[43] F. Donovan, "Med associates healthcare breach exposes data of 270,000 patients," https://healthitsecurity.com/news/270000-put-at-risk-by-med-associates-healthcare-data-breach, 06 2018, (Accessed on 03/29/2022).

[44] J. M. Durán and J. Martínez, "Software supply chain attacks, a threat to global cybersecurity: Solarwinds' case study," *International Journal of Safety and Security Engineering*, vol. 11, no. 5, pp. 537–545, 2021. [Online]. Available: https://doi.org/10.18280/ijsse.110505

[45] ENISA, "Enisa threat landscape 2021 — enisa," https://www.enisa.europa.eu/publications/enisa-threat-landscape-2021, 10 2021, (Accessed on 04/01/2022).

[46] FBI, "Update on sony investigation — fbi," https://www.fbi.gov/news/pressrel/press-releases/update-on-sony-investigation, 12 2014, (Accessed on 04/04/2022).

[47] G. Ferguson, "La fashion enterprise breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/la-fashion-20180823.pdf, 08 2018, (Accessed on 03/29/2022).

[48] M. D. Firoozjaei, N. Mahmoudyar, Y. Baseri, and A. A. Ghorbani, "An evaluation framework for industrial control system cyber incidents," *International Journal of Critical Infrastructure Protection*, vol. 36, p. 100487, 2022. [Online]. Available: https://doi.org/10.1016/j.ijcip.2021.100487

[49] N. Fish, "Goodyear investigating a possible bill payment system compromise," https://eu.azcentral.com/story/news/local/southwest-valley-breaking/2018/05/08/goodyear-investigating-possible-compromise-bill-payment-system/593403002/, 05 2018, (Accessed on 03/30/2022).

[50] B. Geiselman, "Rita loses personal info for 50,000 people, offers free credit monitoring - cleveland.com," https://www.cleveland.com/metro/2016/01/rita_loses_personal_info_for_5.html, 01 2016, (Accessed on 03/30/2022).

[51] D. Gilmour, "Hackers discover 650,000 voter records on voting machine sold on ebay," https://www.dailydot.com/debug/hackers-650000-voter-records-voting-machine-ebay/, 08 2017, (Accessed on 03/30/2022).

[52] J. Gray, "Ashley madison data breach | advanced persistent security," https://advancedpersistentsecurity.net/ashley-madison-data-breach/, 09 2015, (Accessed on 04/04/2022).

[53] J. Green, "Cpt group inc security incident - office of the vermont attorney general," https://ago.vermont.gov/blog/2018/04/27/cpt-group-inc-security-incident/, 04 2018, (Accessed on 03/30/2022).

[54] P. Haney, "Systeme software inc notice of data breach to consumers - office of the vermont attorney general," https://ago.vermont.gov/blog/2018/06/07/systeme-software-inc-notice-of-data-breach-to-consumers/, 06 2018, (Accessed on 03/30/2022).

[55] S. Hodges, "Ca customer notification letter - final_0.pdf," https://oag.ca.gov/system/files/CA%20customer%20notification%20letter%20-%20final_0.pdf, 03 2018, (Accessed on 03/30/2022).

[56] IndianaAttorneyGeneral, "Attorney general: Id theft prevention: Security breaches," https://www.in.gov/attorneygeneral/consumer-protection-division/id-theft-prevention/security-breaches/, (Accessed on 03/30/2022).

[57] ITGovernanceUSA, "Health insurance portability and accountability act | it governance usa," https://www.itgovernanceusa.com/hipaa, (Accessed on 04/13/2022).

[58] J. Jang-Jaccard and S. Nepal, "A survey of emerging threats in cybersecurity," *Journal of Computer and System Sciences*, vol. 80, no. 5, pp. 973–993, 2014, special Issue on Dependable and Secure Computing. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0022000014000178

[59] W. Jansen and T. Grance, "Guidelines on security and privacy in public cloud computing (sp.800-144) | nist," https://doi.org/10.6028/NIST.SP.800-144, 12 2011, (Accessed on 11/13/2021).

[60] C. Joyce, F. L. Roman, B. Miller, J. Jeffries, and R. C. Miller, "Emerging cybersecurity threats in radiation oncology," *Advances in radiation oncology*, vol. 6, no. 6, pp. 100 796–100 796, 2021. [Online]. Available: https://doi.org/10.1016/j.adro.2021.100796

[61] K. Kandasamy, S. Srinivas, K. Achuthan, and V. P. Rangan, "Digital healthcare - cyberattacks in asian organizations: An analysis of vulnerabilities, risks, nist perspectives, and recommendations," *IEEE Access*, vol. 10, pp. 12 345–12 364, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3145372

[62] P. Kaplan, "Privacy incident involving dhs oig case management system (update) | homeland security," https://www.dhs.gov/news/2018/01/18/privacy-incident-involving-dhs-oig-case-management-system-update, 01 2018, (Accessed on 03/30/2022).

[63] S. Keane, "Macy's data breach exposed customer data - cnet," https://www.cnet.com/news/privacy/macys-data-breach-may-have-seen-customer-info-stolen/, 07 2018, (Accessed on 03/29/2022).

[64] S. Khandelwal, "324,000 financial records with cvv numbers stolen from a payment gateway," https://thehackernews.com/2016/09/bluesnap-payment-gateway-hack.html, 11 2016, (Accessed on 03/30/2022).

[65] KI, "Structured literature reviews – a guide for students | karolinska institutet university library," https://kib.ki.se/en/search-evaluate/systematic-reviews/structured-literature-reviews-guide-students, 12 2021, (Accessed on 01/07/2022).

[66] ——, "Systematic reviews | karolinska institutet university library," https://kib.ki.se/en/search-evaluate/systematic-reviews, 02 2022, (Accessed on 04/11/2022).

[67] B. Krebs, "Breached credit union comes out of its shell – krebs on security," https://krebsonsecurity.com/2016/02/breached-credit-union-comes-out-of-its-shell/, 02 2016, (Accessed on 03/30/2022).

[68] ——, "Human resources firm complyright breached – krebs on security," https://krebsonsecurity.com/2018/07/human-resources-firm-complyright-breached/, 07 2018, (Accessed on 03/29/2022).

[69] C. S. Kruse, B. Frederick, T. Jacobson, and D. K. Monticone, "Cybersecurity in healthcare: A systematic review of modern threats and trends," *Technology and Health Care*, vol. 25, pp. 1–10, 2017. [Online]. Available: https://doi.org/10.3233/THC-161263

[70] O. Leon, "California state university fresno notice of data breach to consumers - office of the vermont attorney general," https://ago.vermont.gov/blog/2018/03/06/california-state-university-fresno-notice-of-data-breach-to-consumers/, 03 2018, (Accessed on 03/29/2022).

[71] P. M. and M. C., "Teaching case security breach at target," *Journal of Information Systems Education*, vol. 29, no. 1, p. 11 – 20, 2018. [Online]. Available: http://jise.org/Volume29/n1/JISEv29n1p11.html

[72] G. M. Makrakis, C. Kolias, G. Kambourakis, C. Rieger, and J. Benjamin, "Industrial and critical infrastructure security: Technical analysis of real-life security incidents," *IEEE Access*, vol. 9, pp. 165 295–165 325, 2021. [Online]. Available: https://doi.org/10.1109/ACCESS.2021.3133348

[73] P. Mell and T. Grance, "The nist definition of cloud computing (sp.800-145) | nist," https://doi.org/10.6028/NIST.SP.800-145, 09 2011, (Accessed on 01/17/2022).

[74] A. Montemaggio, S. Iannucci, T. Bhowmik, and J. Hamilton, "Designing a methodological framework for the empirical evaluation of self-protecting systems," in *2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion (ACSOS-C)*, 2020, pp. 218–223. [Online]. Available: https://doi.org/10.1109/ACSOS-C51401.2020.00059

[75] J. Murdock, "Leaked database with 2.9 million records exposes 'every voter in the state of louisiana'," https://www.ibtimes.co.uk/every-voter-louisianas-details-exposed-massive-election-blunder-1583006, 09 2016, (Accessed on 03/30/2022).

[76] B. Nearing, "Hackers hit patient records at st. peter's center," https://www.timesunion.com/business/article/Hackers-hit-patient-records-at-St-Peter-s-center-12723418.php, 03 2018, (Accessed on 03/30/2022).

[77] C. J. Nelson, N. H. Lester-Coll, P. C. Li, H. Gagne, C. J. Anker, M. A. Deeley, and H. J. Wallace, "Development of rapid response plan for radiation oncology in response to cyberattack," *Advances in radiation*

oncology, vol. 6, no. 1, pp. 100 613–100 613, 2021. [Online]. Available: https://doi.org/10.1016/j.adro.2020.11.001

[78] S. Nifakos, K. Chandramouli, C. K. Nikolaou, P. Papachristou, S. Koch, E. Panaousis, and S. Bonacina, "Influence of human factors on cyber security within healthcare organisations: A systematic review," *Sensors*, vol. 21, no. 15, 2021. [Online]. Available: https://doi.org/10.3390/s21155119

[79] J. Novet, "Pentagon asks amazon, google, microsoft, oracle for cloud bids," https://www.cnbc.com/2021/11/19/pentagon-asks-amazon-google-microsoft-oracle-for-cloud-bids.html, (Accessed on 03/04/2022).

[80] NumFOCUS, "pandas - python data analysis library," https://pandas.pydata.org/about/, (Accessed on 02/18/2022).

[81] P. H. O'Neill, "Baton rouge police database 'hacked' in retaliation for killing of alton sterling - the daily dot," https://www.dailydot.com/irl/alton-sterling-baton-rouge-website-hack/, 07 2016, (Accessed on 03/30/2022).

[82] Openstack, "Security/ossa-metrics - openstack," https://wiki.openstack.org/wiki/Security/OSSA-Metrics#Calibration, (Accessed on 04/13/2022).

[83] C. Osborne, "Ticketfly cyberattack exposed data belonging to 27 million accounts | zdnet," https://www.zdnet.com/article/ticketfly-cyberattack-exposed-data-belonging-to-27-million-accounts/, 06 2018, (Accessed on 03/30/2022).

[84] M. Ouedraogo and H. Mouratidis, "Selecting a cloud service provider in the age of cybercrime," *Computers & Security*, vol. 38, pp. 3–13, 2013. [Online]. Available: https://doi.org/10.1016/j.cose.2013.01.007

[85] P. Paganini, "[24]7.ai payment card breach affected major firms, including best buy, after delta air lines and sears holdingssecurity affairs," https://securityaffairs.co/wordpress/71109/data-breach/247-ai-security-breach.html, 04 2018, (Accessed on 03/30/2022).

[86] M.-S. Pang and H. Tanriverdi, "Strategic roles of it modernization and cloud migration in reducing cybersecurity risks of organizations: The case of u.s. federal government," *The journal of strategic information systems*, vol. 31, no. 1, 2022.

[87] A. Pattanayak and M. Kirkland, "Current cyber security challenges in ics," in *2018 IEEE ICII*. IEEE, 2018, pp. 202–207. [Online]. Available: https://doi.org/10.1109/ICII.2018.00013

[88] E. Perez, "Massive postal service breach hits employees - cnnpolitics," https://edition.cnn.com/2014/11/10/politics/postal-service-security-breach/, 11 2014, (Accessed on 03/30/2022).

[89] S. Piasecki, L. Urquhart, and P. D. McAuley, "Defence against the dark artefacts: Smart home cybercrimes and cybersecurity standards," *Computer Law & Security Review*, vol. 42, p. 105542, 2021. [Online]. Available: https://doi.org/10.1016/j.clsr.2021.105542

[90] L. F. Plá, N. Shashidhar, and C. Varol, "On-premises versus secaas security models," in *2020 8th International Symposium on Digital Forensics and Security (ISDFS)*, 2020, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ISDFS49300.2020.9116453

[91] J. Prendergast, "rea.deeming beauty breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/readeeming-beauty-20180105.pdf, 01 2018, (Accessed on 03/30/2022).

[92] PYMNTS, "Massachusetts dept of rev exposes business info. | pymnts.com," https://www.pymnts.com/news/security-and-risk/2018/massachusetts-department-of-revenue-data-breach/, 02 2018, (Accessed on 03/29/2022).

[93] B. Radke, "Centris federal credit union breach notice - iowaattorneygeneral.gov," https://www.iowaattorneygeneral.gov/media/cms/030118_Centris_Federal_Credit_Union_C682106B3D20B.pdf, 03 2018, (Accessed on 03/29/2022).

[94] A. RAZA, "#escortsoffline campaign gains steam - stolen data from 79 escort websites," https://www.hackread.com/stolen-data-from-79-escort-websites/, 01 2016, (Accessed on 03/30/2022).

[95] Reuters, "Thousands of guests' data may have been hacked at starwood, marriott, hyatt hotels," https://www.nbcnews.com/tech/tech-news/thousands-guests-data-may-have-been-hacked-starwood-marriott-hyatt-n630811, 08 2016, (Accessed on 03/30/2022).

[96] E. Roseman, "Tip leads to notification of security lapse: Roseman | the star," https://www.thestar.com/business/personal_finance/2014/09/09/tip_leads_to_notification_of_security_lapse_roseman.html, 09 2014, (Accessed on 03/30/2022).

[97] L. Seungjin, A. Abdullah, and N. Jhanjhi, "A review on honeypot-based botnet detection models for smart factory," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 6, 2020. [Online]. Available: http://dx.doi.org/10.14569/IJACSA.2020.0110654

[98] E. Shim, "Register.com cyber hack may involve chinese military - upi.com," https://www.upi.com/Top_News/World-News/2015/03/18/FBI-looking-into-Chinese-military-involvement-in-cyber-hack-of-US-company/2531426688682/, 03 2015, (Accessed on 03/30/2022).

[99] A. Shostack, "Experiences threat modeling at microsoft," https://adam.shostack.org/modsec08/Shostack-ModSec08-Experiences-Threat-Modeling-At-Microsoft.pdf, (Accessed on 04/13/2022).

[100] ——, *Threat Modeling: Designing for Security*, 1st ed. Wiley, 2014.

[101] C. Sienko, "The breach of anthem health - the largest healthcare breach in history - infosec resources," https://resources.infosecinstitute.com/topic/the-breach-of-anthem-health-the-largest-healthcare-breach-in-history/, 06 2016, (Accessed on 04/04/2022).

[102] P. Sisson, "Patient data breach at tri-city - the san diego union-tribune," https://www.sandiegouniontribune.com/news/health/sdut-tri-city-data-breach-2014aug15-story.html, 08 2014, (Accessed on 03/30/2022).

[103] Sony, "Sony group portal - news releases - sony online entertainment announces theft of data from its systems," https://www.sony.com/en/SonyInfo/News/Press/201105/11-0503E/, 05 2011, (Accessed on 04/04/2022).

[104] L. Sotto, "Bronson nutritionals breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/bronson-nutritionals-20180319.pdf, 03 2018, (Accessed on 03/29/2022).

[105] Specops, "The countries experiencing the most 'significant' cyber-attacks - specops software," https://specopssoft.com/blog/countries-experiencing-significant-cyber-attacks/, 07 2020, (Accessed on 04/07/2022).

[106] V. Steckler, "Avast forum offline due to attack," https://blog.avast.com/2014/05/26/avast-forum-offline-due-to-attack/, 05 2014, (Accessed on 03/30/2022).

[107] A. Sternstein, "Unclear whether opm hackers accessed feds' names - nextgov," https://www.nextgov.com/cybersecurity/2014/07/its-unclear-whether-opm-hackers-accessed-feds-names/88783/, 07 2014, (Accessed on 03/30/2022).

[108] P. Summer, "Delta airlines breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/delta-air-lines-20180411.pdf, 04 2018, (Accessed on 03/29/2022).

[109] J. Surbiryala and C. Rong, "Cloud computing: History and overview," in *2019 IEEE Cloud Summit*, 2019, pp. 1–7. [Online]. Available: https://doi.org/10.1109/CloudSummit47114.2019.00007

[110] E. Sweeney, "More than 316,000 patient blood tests exposed in breach linked to home monitoring company | fierce healthcare," https://www.fiercehealthcare.com/privacy-security/data-breach-medical-records-blood-tests-patient-home-monitoring-kromtech-security, 10 2017, (Accessed on 03/30/2022).

[111] H. Tabrizchi and M. Kuchaki Rafsanjani, "A survey on security challenges in cloud computing: issues, threats, and solutions," *The Journal of Supercomputing*, vol. 76, no. 12, pp. 9493–9532, 2020. [Online]. Available: https://doi.org/10.1007/s11227-020-03213-1

[112] P. Tavares, "Rail europe north america data breach," https://seguranca-informatica.pt/rail-europe-north-america-data-breach/#.YkQ9wShByUk, 05 2018, (Accessed on 03/30/2022).

[113] T. Troy, "Residents' information is exposed in hacking | the blade," https://www.toledoblade.com/Politics/2016/12/07/Residents-information-is-exposed-in-Henry-County-hacking.html, 12 2016, (Accessed on 03/30/2022).

[114] UnderArmourInc, "Under armour notifies myfitnesspal users of data security issue | under armour, inc." http://investor.underarmour.com/news-releases/news-release-details/under-armour-notifies-myfitnesspal-users-data-security-issue?ReleaseID=1062368, 03 2018, (Accessed on 03/30/2022).

[115] J. R. Vacca, *Cloud computing security: Foundations and challenges.* CRC Press Taylor & Francis Group, 2017, ch. Section 1: Chapter 2.1.

[116] J. Valdetero, "Health equity inc breach notice - nh.gov," https://www.doj.nh.gov/consumer/security-breaches/documents/healthequity-20181116.pdf, 11 2018, (Accessed on 03/30/2022).

[117] G. Vatu, "vbulletin hack exposes 820,000 accounts from 126 forums," https://news.softpedia.com/news/vbulletin-hack-exposes-820-000-accounts-from-126-forums-513416.shtml, 02 2017, (Accessed on 03/30/2022).

[118] WAQAS, "Evony gaming company website hacked; 33m gamer accounts stolen," https://www.hackread.com/evony-gaming-company-website-hacked/, 10 2016, (Accessed on 03/30/2022).

[119] WBURNewsroom, "Bps to change student id cards after information was lost | wbur news," https://www.wbur.org/news/2013/08/12/bps-student-id-cards, 08 2013, (Accessed on 03/30/2022).

[120] C. M. Williams, R. Chaturvedi, and K. Chakravarthy, "Cybersecurity risks in a pandemic," *Journal of medical Internet research*, vol. 22, no. 9, pp. e23 692–e23 692, 2020. [Online]. Available: https://doi.org/10.2196/23692

[121] C. Wohlin, P. Runeson, M. Höst, M. C. Ohlsson, B. Regnell, and A. Wesslen, *Experimentation in software engineering.* Springer Heidelberg, 2012.

[122] A. Zieger, F. Freiling, and K.-P. Kossakowski, "The beta-time-to-compromise metric for practical cyber security risk estimation," in *International Conference on IT Security Incident Management and IT Forensics, IMF.* IEEE, 2018, pp. 115–133. [Online]. Available: https://doi.org/10.1109/IMF.2018.00017

Appendix A

Supplemental Information

A.1 Data processing and sampling

```
import pandas as pd
from sklearn.utils import resample

MIN_YEAR = 2008
MIN_RECORDS = 10000
SAMPLE_SIZE = 20
SOURCE_CSV_FILE = './datasets/merged-cleaned.csv'
EXPORT_CSV_FILE = 'datasets/samples-merged-cleaned.csv'

def filter_data(data: pd.DataFrame) -> pd.DataFrame:
    # Filter out any rows that contains NaN (both Cause and URL have
    ↪  these):
    data = data.dropna()

    # Filter out any rows of incidents that happened before 2008:
    data = data[data['Year'] >= MIN_YEAR]

    # Filter out any rows where the amount of records affected by the
    ↪  breach is lower than MIN_RECORDS:
    data = data[data['Records'] > MIN_RECORDS]

    # Filter out 'Insider job' causes because they have no relevance
    ↪  to the study:
    data = data[data['Cause'] != 'Inside job']

    # Filter out 'Paper Data' causes because they have no relevance
    ↪  to the study:
    data = data[data['Cause'] != 'Paper Data']

    # Filter out any records with URLs that do not contain any
    ↪  valuable information:
    # (ocrportal does not contain any additional information about
    ↪  any incident = filter out)
```

```
    data =
    ↪  data[data['URL'].str.contains('https://ocrportal.hhs.gov/ocr/breach')
    ↪  == False]

    # Remove eventual duplicates based on the four columnns 'Year',
    ↪  'Records', 'Industry' and 'Cause':
    data = data.drop_duplicates(subset=['Year', 'Records'],
    ↪  keep='last')

    return data

def sample_data(data: pd.DataFrame) -> list:
    # Extract all the different Causes:
    causes = set(data['Cause'])
    cause_data_split = [data[data['Cause'] == cause] for cause in
    ↪  causes]

    # Extract samples from each "Cause" category that will be used in
    ↪  the study:
    samples = []
    for cause_category in cause_data_split:
        samples.append(resample(cause_category, random_state=42,
        ↪  replace=False, n_samples=SAMPLE_SIZE))

    return samples

def export_samples(samples: list) -> None:
    # Export the samples to a new CSV file:
    samples[0].to_csv(EXPORT_CSV_FILE, index=False)
    for index in range(1, len(samples)):
        samples[index].to_csv(EXPORT_CSV_FILE, mode='a', index=False,
        ↪  header=False)

    return

data = pd.read_csv(SOURCE_CSV_FILE, delimiter=';')
# Filter and sample the data. Then export the samples to a new CSV
↪  file:
data = filter_data(data)
samples = sample_data(data)
export_samples(samples)
```

Faculty of Computing, Blekinge Institute of Technology, 371 79 Karlskrona, Sweden