Deployment of Next Generation Intrusion Detection Systems against Internal Threats in a Medium-sized Enterprise

Master of Science in Technology thesis
University of Turku
Department of Future Technologies
Security of Networked Systems
October 2020
Filippo Piconese

Supervisors:
Antti Hakkala
Seppo Virtanen
Bruno Crispo

UNIVERSITY OF TURKU
Department of Future Technologies

FILIPPO PICONESE: Deployment of Next Generation Intrusion Detection Systems against Internal Threats in a Medium-sized Enterprise

Master of Science in Technology thesis, 75 p., 0 app. p.
Security of Networked Systems
October 2020

In this increasingly digital age, companies struggle to understand the origin of cyberattacks. Malicious actions can come from both the outside and the inside the business, so it is necessary to adopt tools that can reduce cyber risks by identifying the anomalies when the first symptoms appear. This thesis deals with the topic of internal attacks and explains how to use innovative Intrusion Detection Systems to protect the IT infrastructure of Medium-sized Enterprises. These types of technologies try to solve issues like poor visibility of network traffic, long response times to security breaches, and the use of inefficient access control mechanisms. In this research, multiple types of internal threats, the different categories of Intrusion Detection Systems and an in-depth analysis of the state-of-the-art IDSs developed during the last few years have been detailed. After that, there will be a brief explanation of the effectiveness of IDSs in both testing and production environments. All the reported phases took place within a company network, starting from the positioning of the IDS, moving on to its configuration and ending with the production environment. There is an analysis of the company expectations, together with an explanation of the different IDSs characteristics. This research shows data about potential attacks, mitigated and resolved threats, as well as network changes made thanks to the information gathered while using a cutting edge IDS. Moreover, the characteristics that a medium-sized company must have in order to be adequately protected by a new generation IDS have been generalized. In the same way, the functionalities that an IDS must possess in order to achieve the set objectives were reported. IDSs are incredibly adaptable to different environments, such as companies of different sectors and sizes, and can be tuned to achieve better results. At the end of this document are reported the potential future developments that should be addressed to improve IDS technologies further.

Keywords: Cyber Security, Network Security, Intrusion Detection System, Intrusion Prevention System, Network Traffic Analysis, User and Entity Behaviour Analytics, Machine Learning, Artificial Intelligence, Internal Threats

Contents

Chapter 1

Introduction

This research aims to convey how to seek and resolve weaknesses within midsize businesses using behaviour-based IDSs. By positioning these tools correctly, it is possible to analyze the traffic flowing within the network and correlate the events in order to reinforce the security measures. In general, IDSs can effectively protect the corporate network. However, network traffic changes rapidly, and attacks are becoming more sophisticated. IDSs based on rules or signatures cannot keep pace with attackers. Therefore, they must be combined with IDSs which detect indicators of compromise (IOCs) and identify potential threats. This thesis wants to explain the capabilities of each type of IDS clearly. To understand how to position them correctly within a corporate network and to grasp why behavioural IDSs are essential to improve the company's security position.

To successfully secure the network is necessary to know the different types of internal threats, which have been explained in Chapter 2. Besides the attack types, in that chapter, the vulnerabilities, the available defensive strategies, and the trends related to insiders and internal threats have been reported. Subsequently, Chapter 3 explains the IDS categories, some use-cases and the network placement strategies. The first two chapters point to clarify the starting problem and illustrate possible solutions to it. Chapter 4 introduces two general approaches of behavioural Intrusion Detection Systems (IDSs), one based on deep learning and the other one based on data mining techniques. The third advanced IDS is

a commercial product which encompasses multiple algorithms based on Supervised and Unsupervised Machine Learning, Artificial Intelligence and statistical models. These advanced systems represent one of the most effective solutions to identify and stop internal threats. The just-mentioned type of threats may come from a downloaded malware, an insider, a naive user or an incorrect network configuration. Also, internal threats can come from private cloud since the latter is part of the internal network. Then, Chapter 5 introduces the company in which the use case has been implemented and details the company expectations before the Darktrace deployment. The active and applied part of the thesis is summarized in Chapter 6, which contains in-depth technology tests. The content focuses on the implementation, optimization and verification of the technology. In Chapter 7, the data collected during the testing period have been analyzed. It includes the most exciting facts such as security incidents, compliance issues and network configuration mistakes. Finally, the concluding chapter, other than summarize the main results of this thesis, reflects on the applicability of this technology in different business environments both in terms of size and product sector.

To conclude the introduction, this thesis is conceived to answer the following points:

- How to identify real known and unknown threats that arise within the corporate network.
- Understand how to propose countermeasures able to reduce the number of internal security holes employing innovative IDSs.
- Observe and analyze the positive and negative aspects evaluated after a period of use of the intrusion detection technology in order to generalize the specific case, and apply the same technique to other cases.

Chapter 2

Internal threats

In the past few years organizations have experienced an evolution of their network given by the increasing utilization of personal devices, cloud computing, joint ventures, temporary workers and outsourcing arrangements [1]. The perimeter of the corporate network has extended to cloud services and beyond. All of these changes have led to insidious business security failure points called internal threats.

Internal threats include incorrect network configurations, insufficient security policies, incorrect implementation of access control mechanisms and, not least, the presence of insiders, which is potentially the most dangerous one. A possible definition of malicious insider threat is provided by "The CERT Guide to Insider Threats" [2]: "A malicious insider threat may be either a current or former employee, contractor, or business partner who has or had authorized access to an organization's network, system, or data and intentionally exceeded or misused that access in a manner that negatively affected the confidentiality, integrity, or availability of the organization's information or information systems".

In this chapter, the main concepts, numbers, statistics and examples have been reported. They are essential to understand the importance of internal threats in modern cybersecurity environments. Section 2.1 contains the most common threats and vulnerabilities coming from inside the company network. Then, 2.2 describes the best practices

to defend the internal perimeter against these issues, and the most popular shortcomings. Lastly, 2.3 reports an overview of the trends seen in the last decades and the predictions for the future.

2.1 Threat types and vulnerabilities

There are different types of insider threats which concern modern society [3]. Each of them should be carefully taken into account and weighted to balance the security measures employed as a countermeasure. Specifically, these activities can be categorized as:

- Unintentional, where an employee behaviour is as much naive as dangerous for the organization. In this case, the malevolent action is not made on purpose, but accidentally. Nonetheless, this situation can lead to great money or reputation losses.

- Fraud, like in the previous case with the difference that the actor is aware of what is happening and he intends to gain money selling important company's know-how information and data.

- Intellectual property theft, which concerns the company's patents, trademarks, trade secrets and copyrights that an insider intends to steal and sell to other parties. This issue can be very costly since the value of a company is greatly given by these properties, instead of the physical ones.

- IT sabotage, consisting of ceasing or disruption of technical solutions used by the organization, is made by employees with a technical position and in possession of operational rights. Few lines of malicious code can be devastating, and this is easier and more likely to happen when the modifications are written by someone with extensive knowledge of the program.

- Espionage, where the intent is to understand the target secrets, plans and peculiarities to gain competitive advantages. Competitor firms or governments usually make

this attack.

Most of the times, the just mentioned threats, plus the ones at the network level, are exploited through very subtle vulnerabilities. These security holes may be tough to patch since can also involve humans. Furthermore, in many cases, it is difficult to distinguish an internal threat from legitimate activity.

An example can be a disgruntled employee who tries to steal sensitive information about an ongoing patent procedure and sell it to another company as a revenge to have been fired from the company. Monitoring this type of behaviours is necessary to guarantee a comprehensive security posture. Affiliates security standards are other common weak points. For instance, when an enterprise exchange sensitive information with smaller businesses to increase the number of proposed services or enhance the current ones. In this case, the affiliate becomes a significant company's internal branch, which provides relevant solutions to their customers. However, it reveals to be a weakness under the security perspective. Suppose an attacker compromises the affiliate. In that case, he gets a starting foothold also inside the enterprise infrastructure, and from there, he will spread to other facilities to steal as many secrets as possible.

There can be multiple consequences due to these threats, like financial losses, organization disruption, loss of reputation and impacts on the company's employees who might suffer high pressure caused by the press and the customers. To face and defeat all these problems, well-prepared remediation plans and security controls, which are introduced in the next section, are essential and should be carefully studied and tested.

2.2 Defensive security measures

Some remediation guidelines should be appropriately designed and followed to overcome security issues. The first general advice is to have well structured and adequate security measures in place. It would help with all the phases related to internal threats, from

the prevention to the remediation. Then, a best-practice is to assign an executive-level leader to every project so that there is a reference person ready to put into practice the security policies and decisions taken from the Chief Information Security Officer (CISO). The latter must be a forward-looking person who knows the company processes and can interact with all the departments.

Besides the just mentioned conditions, a company should strike up an Insider Threat Program (InTP). To be effective, an InTP must comprehend stakeholders coming from different sources. For instance, the departments of cybersecurity, information assurance, H&R and law must contribute to the InTP project since it allows to separate the technical measures, needed to secure the machines, from the sociological activities, useful to understand the reason behind human actions. Besides, like reported in [4], there is a tool called Cybersecurity Framework (CSF) conceived by the National Institute of Standards and Technology's (NIST) to help organizations with the implementation of an effective InTP. CSF's ultimate goal is to simplify the measurement and mitigation procedures, reducing cyber risk. Briefly, it consists of five iterative phases: identification of data at risk, protections of the latter, detection of malicious intents, defensive response and, finally, the post-facto recovery. After these steps, the company should be able to identify and correct the weaknesses present in the plan.

To have a more detailed overview of the remediation methods, and to be more effective at decreasing the overall business cyber risk, the following tips should be followed:

- Document and enforce security policies since are required to assure the proper usage and protection of company systems, network and data. These rules are useful to specify what an employee is allowed to do and, at the same time, detect misuses or infringements. An example is credentials revocation upon discharging an employee. Another one is to change the passwords used to access critical systems when a C-level worker or system administrator is fired.

- To sustain the InTP, it is paramount to adopt some technical measures. These can

be used to prevent data losses, analyze the endpoints, perform Security Information and Event Management (SIEM) activities, control incoming e-mails and monitoring the network to spot rogue devices or data exfiltration. For instance, an innovative approach to detect unexpected changes in the company network's nodes behaviour is called User and Entity Behavior Analytics (UEBA). A UEBA system uses Machine Learning and Artificial Intelligence algorithms and compares the subject historical operations with the events happening in real-time. This method is effective in case of complex attacks since multiple attack vectors are adopted to overcome the traditional security measures, creating patterns which can only be detected by advanced UEBA tools.

- Put in place physical controls to enforce security checks before gaining access to physical assets. The more the location contains devices able to access important information, the more the physical barrier should be strict and hard to overcome illicitly.

- Perform security awareness training to all the company's employees. The internal threats component should be particularly stressed since it concerns both technical and non-technical workers. Then, the company security officer should perform frequent tests to assess the understanding of employees and, in case of fails, submit training exercises. Some examples of assessments are simulated phishing campaigns, vishing, left USB sticks unattended all over the working places and other social engineering scams.

- As the last recommendation, a company must be ready to mitigate an internal threat. Remediation plans must be easy to maintain and ready to be activated. Besides, effective escalation and alert notification systems are needed.

A schema of these remediation guidelines, taken from [3], is reported in Figure 2.1. As reported in [1], to be effective against internal threats is necessary to utilize different

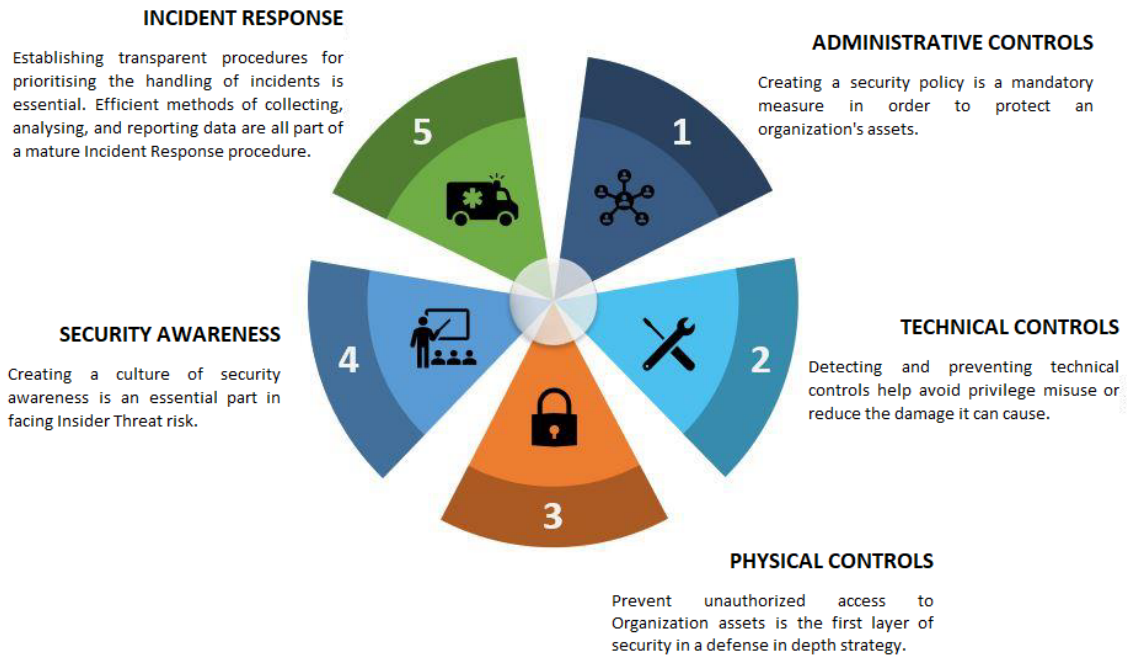

Figure 2.1: Remediation guidelines - image retrieved from [3]

approaches ranging from technical, non-technical and hybridization of the two. These points of view are detailed in the following three sections.

2.2.1 Technical perspective

The combination of several methods and technical means help a company to strengthener its security posture. First of all, the organization policies should be written using standard policy languages so that they can be human-readable, formal, expressive and easy to check for inconsistencies and contradictions. Moreover, a well-written policy should describe the expected workflow as well as the monitoring and enforcing phases. Another step to prevent internal threats consists of adopting strict access control mechanisms. The desired outcome includes user authentication, least privileges authorization and separation of responsibilities among company departments. In real-world scenarios, too strict access controls can limit the workers' productivity, which is not profitable for the company. Therefore, a trade-off between profitability and security is needed. The monitoring phase

is paramount to spot malicious activities. It can be done in different ways, like:

- Detect misuses based on known and pre-defined rules, which is what a common Intrusion Detection System does.
- Detect anomalies, which consists of spot variations to the network entities' normal behaviour.
- Compare multiple events using statistical-based approaches.

These monitoring systems can be placed on the network level, host level or both. Depending on the placement, the visible data changes. Unfortunately, human analysis is frequently needed since false positives and false negatives arise. Last but not least, Data Loss Prevention (DLP) tools [5] as long as system hardening against tampering and manipulation is useful to prevent attacks from both insiders and outsiders. Therefore, these technical methods should be implemented and coherently deployed inside the network infrastructure.

2.2.2 Socio-Technical perspective

A hybrid approach is obtained by combining technical and sociological measures. The aim is to improve effectiveness against internal threats and to explain the policy positive and negative aspects. The first topic is the policy language, which should consider not only the operational needs but also the behaviour and dynamism of the context. For example, a policy should be able to allow infringements in particular situations such as emergencies. Besides, there may be distinctions between technical details of a policy and how it is understood from human-beings. Then there is the profiling phase that precedes the monitoring of anomalies. The characteristics required to rank an insider may be too many to handle. Therefore, a compromise between the technical precision of the policy and the set of allowed behaviours is required.

Analyze the feelings and sociological aspects of employees help to prevent insiders. The psychological indicators should be deeply studied and then translated into technical terms. It is challenging, though, due to both the lack of sample data and human complexity. Suppose there is a potential attack coming from within the network and it is attributable to a worker. In that case, it may be hard for a company to decide whether to stop the attack or let it goes and perform forensics afterwards. A false accusation related to the potential source of the attack, in case it was a natural person, can be counterproductive in many ways. It can negatively influence corporate culture, cause the accused person to resign or seek legal actions. How to intervene must be carefully decided since too strict countermeasures frustrate an honest employee while too soft ones lead the insider to pursuit his activities.

To conclude this topic, in general, proactive and reactive approaches are both needed. Therefore, technical monitoring, along with reports about staff behaviour and non-IT controls, should be combined to create effective detection methods.

2.2.3 Purely human-oriented perspective

Human-oriented approaches aim to find and explain the reasons behind insider intentions. A human being purely conducts them and, as a consequence, vary among individuals. A policy can be considered successfully designed when it strengthens the security posture of a company, without interfering with its workflow. Moreover, that policy should be accepted, well understood, and applied by the staff. The security awareness training lies in this category and should increase the organization security culture while serves as a deterrent to individuals' malicious intentions.

A policy is designed for a specific domain, but an insider could unexpectedly use that domain. Therefore, a policy should be context-dependent and adaptable. Moreover, it is vital to test policies and reduce redundancy. After policy deployment, there is the monitoring phase which helps to understand how the staff perceive it and if it fits the

company's workflow. Besides, it is helpful to promote the culture of reporting anomalous activities made by colleagues other than use purely technical monitoring techniques.

Finally, laws, privacy and ethics are all components to be carefully considered when devising security measures. Furthermore, human characteristics must be observed and used to evaluate their feasibility and effectiveness.

2.3 Statistics and trends

To date, there are two main factors why very little concrete is available about internal threats. Firstly, the attacker has to be identified and caught accordingly to the law. To do this, a company needs concrete elements to incriminate the suspect/culprit. Secondly, the majority of the firms do not share information about an internal lack of security since it could damage their reputation and business. That said, there are some official surveys and studies which report numbers related to this type of threats along with an estimation of revenue losses.

2.3.1 The presence of internal threats in the threat landscape

The Verizon Data Breach Investigations Report [6] is a document drawn up each year in which many details are written about incidents and security breaches. As for the 2019 edition, Verizon analyzed more than 40,000 security incidents including more than 2,000 confirmed breaches. Although the total number of reported incidents in 2019 is lower than in the 2018 edition, where around 53,000 cases were reported, the total amount on revenue losses is higher.

For what concerns internal threats, there are several exciting numbers to take into account. One is that 34% of the total breaches involved internal actors and the curve suggests that it is steadily increasing after the trough in 2015 like reported in Figure 2.2. Other than this, the driving motivations behind these facts are mainly, as expected, related to financial

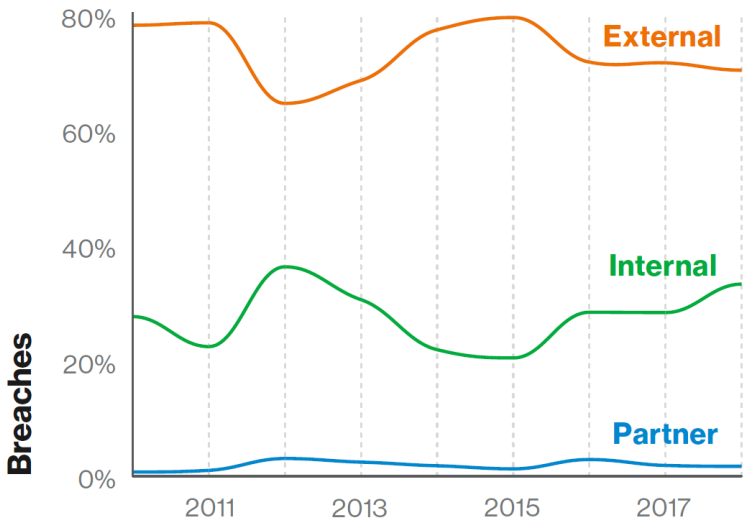

Figure 2.2: Beaches actors over time - image retrieved from [6]

gains and espionage purposes as depicted in Figure 2.3. Digging into the specific threat actors category it is noticeable from Figure 2.4 a decreasing trend of Organized crime in favor of a sky rocketing State-affiliated and System Admins. Last but not least, there are some sectors which have experienced a greater level of internal threats compared to all the other sources. They are, in decreasing order, Healthcare (59%), Educational Services (45%), Information (44%), Financial and Insurance (36%), Manufacturing (30%) and Public Administration (30%). While, on the opposite side, there are three sectors less affected by insiders which are Professional, Technical and Scientific Services (21%), Retail (19%) and Accommodation and Food Services (5%).

Another comprehensive information source is the "Insider Threats 2018 Report" [7]. Among the details reported, this document states that almost 30% of the companies perceive an increment in the threats coming from internal attacks. The most common technologies used to overcome the onset of this problem are Data Loss Prevention (DLP) solutions, data encryption and access control mechanisms. Moreover, to actively detect and block insider attacks, organizations use IDS and IPS technologies, along with log management and SIEM platforms used to converge various log sources into a single plat-

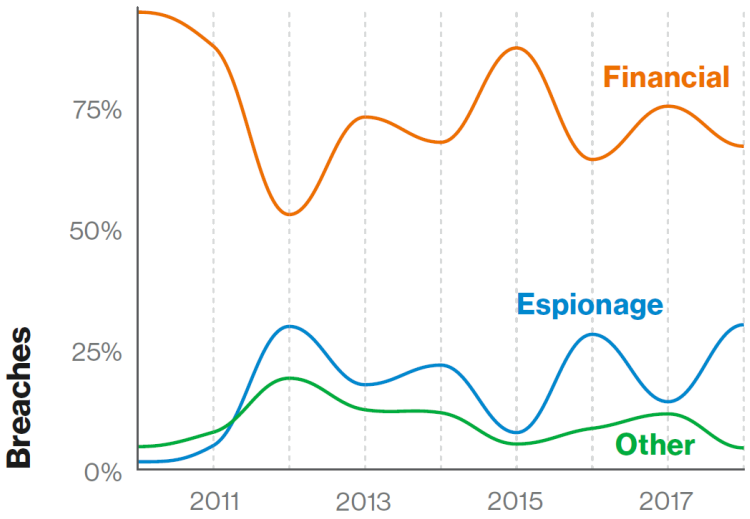

Figure 2.3: Breaches motivations over time - image retrieved from [6]

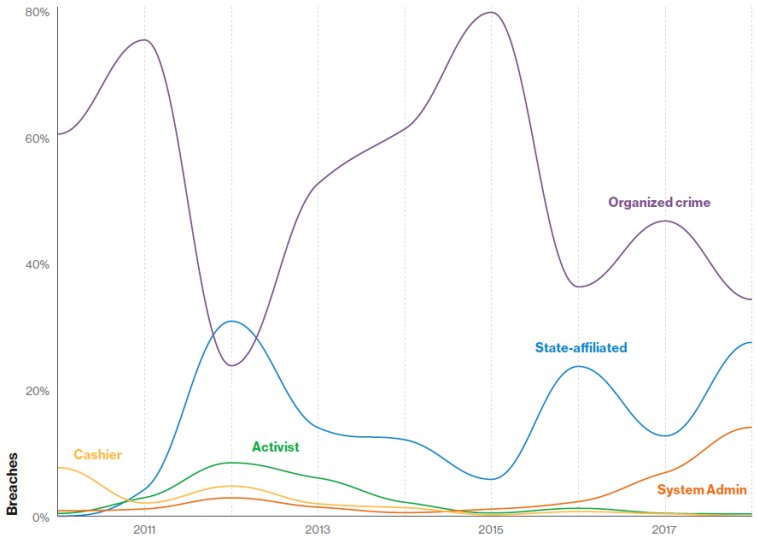

Figure 2.4: Breaches actors categories over time - image retrieved from [6]

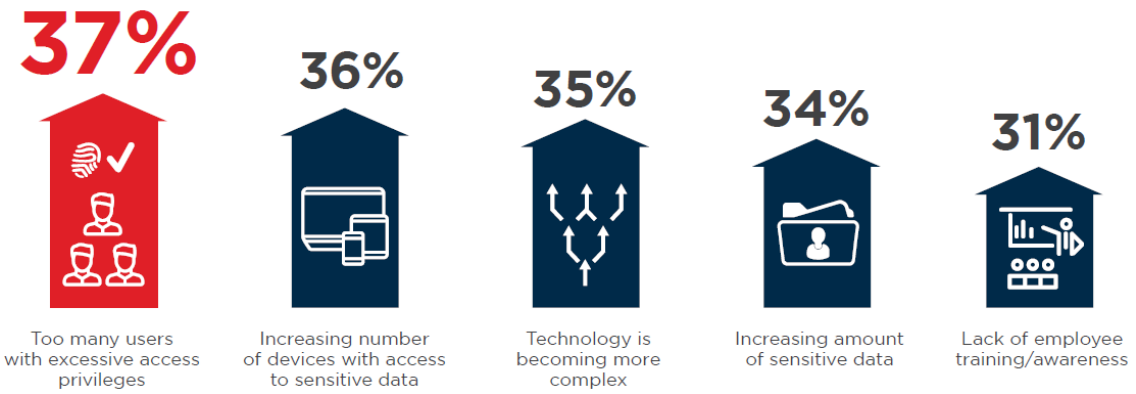

Figure 2.5: Insider attacks facilitators - image retrieved from [7]

form. Besides, the data gathered to produce the statistics indicates that phishing attacks caused, on average, 67% of insider threats. This data is substantially the same as the 2019 report made by Verizon, with the exception that cyber-awareness training is helping to decrease the employees clicking of bogus e-mails. This strategy is more and more adopted, and the goal is to enhance the company employees security culture through a learning process so that to reduce phishing attacks success rate. There are several causes which make internal attacks happening like granting excessive privileges to an employees, an ever-growing number of devices capable of accessing sensitive data, the increasing network structure complexity and a multitude of technologies. All these factors, and more, are depicted in Figure 2.5.

2.3.2 Losses due to internal threats

A valuable study related to the cybercrime costs made by Accenture Security and Ponemon Institute in the "Ninth Annual Cost of Cybercrime Study" [8] outlines the losses caused by internal threats. In that document are reported information including, for example, the forecasted value at risk from cyberattacks, which equals to $5.2t. Besides, it is detailed a breakdown of cybercrime costs starting from the value corresponding to malicious insiders, as reported in Figure 2.6. Looking at the graph, it is clear that in 2018 malicious

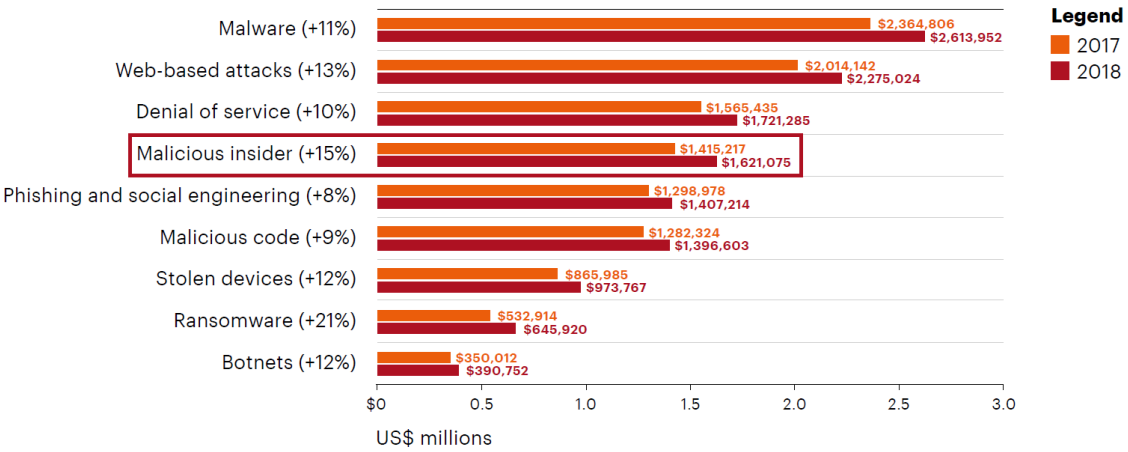

Figure 2.6: Average annual cost of cybercrime by type of attack (2018 total = US$ 13.0 million) - image retrieved from [8]

insiders grew considerably compared to 2017. Therefore, it is fundamental to adopt effective security measures to counteract this type of attacks. Only the ransomware attacks costs have beaten the insiders' growth rate.

To conclude, the report [7] states that 27% of respondents believe that a successful internal attack cost could range between $100,000 and $500,000, while 24% of them believe the cost is higher than $500,000. It is good to reiterate that many of these details are estimations, given the difficulties in finding factual information related to internal attacks, as mentioned at the beginning of the chapter.

Chapter 3

Intrusion detection and prevention systems

Intrusion Detection and Prevention Systems, henceforth called IDS and IPS, are technical measures suitable for dealing with insiders [9], and these technologies can be beneficial. This chapter is wholly dedicated to the introduction and explanation of IDS/IPS functioning, the different variants that exist from a theoretical and practical point of view and some examples of how they can be deployed inside a corporate network. Note that Next-Generation Firewalls are also needed when designing a security solution since they should block all the traffic coming from outside except for the allowed one. As a consequence, an IDS is needed to monitor the "trusted" traffic and detect strange and suspicious behaviours of internal users or devices. Section 3.1 reports the big picture representing IDS solutions starting from their categorisation and continuing with the differentiation of the approaches used to design an IDS solution. Then, sections 3.2 and 3.3 details the various types of IDSs and their combination, along with the related positive and negative aspects. Lastly, section 3.4 indicate a possible deployment strategy.

3.1 IDS overview

The concepts of IDS and IPS mainly differ from the reaction type. Specifically, an IDS is a passive tool which listens to the captured traffic and creates events out of it. Then, it correlates the events to generate alerts about suspicious activities. Lastly, the gathered information is presented to the security team, which performs the investigation phase. In addition to performing the same activities as an IDS, an IPS also has an active part. Specifically, it can react to threats, blocking and containing them, providing valuable time to the security team that will eradicate them from the network. The detection and prevention capabilities are paramount to counteract malicious actions initiated within the business perimeter. The most common attacks discovered by IDSs are:

- DoS or DDoS attacks, which consists of flooding the target system with requests to create slowdowns and disservice.
- Protocol violations, like when a packet is modified, and the result is not coherent with the expected protocol guidelines.
- Policy violations, also in case of custom policies created ad hoc by the company.
- Wrong network configuration, which comprehends, for instance, unexpected communication among subnets and resources accessed by users who should not be allowed to see them.
- Reconnaissance, which is the information gathering process accomplished by an attacker to gain more information about the targeted system.
- Exploits of vulnerabilities present in the technologies protected by the IDS.

The most important aspect to consider when it comes to IDS effectiveness is data visibility. An obvious consideration is that cannot be identified what cannot be seen. Another fundamental feature is the IDS ability to correlate and correctly identify a series of events

that can cause a malicious event. For these and other reasons, there are several IDS categories and approaches to consider when designing internal security for a corporate network.

Taking into account different sources like [10] and [11], there are five types of IDS:

- Network-based IDS (NIDS), which is an IDS connected to a switch to capture its traffic and analyse/evaluate it.
- Host-based IDS (HIDS), where an IDS agent monitors a single host and searches for anomalous system calls, gathers machine logs, identifies files modifications and performs other similar actions.
- Wireless IDS, which consists of a LAN IDS able to examine wireless traffic, identifying the connections made to an access point (AP) by external users who perform dangerous actions. This type of IDS is integrated into an AP or wireless router behind the firewall.
- Perimeter IDS (PIDS), which is placed on the company infrastructure perimeter and utilises optical fibre or electronics to detect intrusion attempts and fire alarms.
- VM-based IDS (VMIDS), in this case, there is an IDS deployed in a virtual environment and uses the VM monitoring feature to collect data.

Note that the most commonly used IDSs are the host-based and network-based ones.

Only one aspect is missing to complete the IDS general view, which is the comparison between the types of detection models used by an IDS to classify the monitored traffic. Substantially, there are two main approaches:

- Signature-based IDS that compares the traffic with a database of signatures representing known attack patterns and creates alerts if a match is found between the analysed traffic and an attack flow. These signatures can be seen as a set of rules

which characterise certain dangerous behaviour. The signature database must always be updated to ensure an acceptable level of security.

- Behaviour-based IDS, also known as User and Entity Behaviour Analytics (UEBA) systems or statistical-based (SBID) systems, consists of defining the "normal" traffic flow that represents the network activities to detect spikes and variances that identify abnormalities. There are multiple algorithms and ways used by these types of IDS to learn entity models, and the more packets are ingested, the more accurate the detection will be. Thanks to this approach, an IDS is potentially able to detect changes in known attacks and 0-day attacks.

In the following sections, there are little comparisons among the differences discussed until now.

3.2 Network-based vs. Host-based IDS

In this section will be highlighted a summary of NIDS and HIDS advantages and disadvantages.

One of the most positive aspects of network-based IDS is the ability to have a complete overview of what is happening in the network. A practical example is the case of a DDoS attack coming from inside the company, maybe due to internal zombie hosts controlled by an attacker which are sending data all together against a specific domain. In this case, the IDS will alert the security team about what is happening and, in case it also has IPS capabilities, it could proactively stop the attack blocking the compromised hosts. Moreover, a NIDS does not require any computational power consumption made from the endpoints and is more resilient to attacks since, usually, the initial foothold of an attacker is an internal host from where he can start spreading into the network. The main disadvantages are poor visibility on encrypted traffic, not being able to monitor user processes and activities within the host and, also, hardware must be added to the corporate

network to allow IDS to receive traffic.

On the other hand, a host-based IDS can collect all the processes and events produced by an endpoint, is convenient for small networks, can also monitor the encrypted traffic and detects internally generated attacks. Unfortunately, this type of IDS requires the installation of an agent inside the monitored device, uses the endpoint processing power to work, and, in case the system is compromised, even the HIDS will no longer be reliable because potentially corrupted.

Decide between NIDS and HIDS is not simple. Therefore, company requirements and expectations should be very carefully considered. The two methods can be combined to achieve a better security posture.

3.3 Signature-based vs Behavior-based IDS

A Signature-based IDS, also known as misuse detection system or knowledge-based detection system, has few advantages which include less false alarms percentage out of the total amount of alerts, quick known threats recognition and easy configuration procedure. There are two relevant shortcomings:

- The impossibility to detect new threats or variants of them, leaving possibilities for very dangerous false negatives when it comes to the production environment.
- The signatures database must be kept up-to-date to keep up with the rapid growth of cyberattacks.

On the opposite side, a Behavior-based IDS, also called anomaly-based detection system, is potentially able to detect threats never seen before. Not only malware but also suspicious users activities, incorrectly configured permissions, forged packets and other network activities which are unexpected compared to the normal traffic flow previously observed by the IDS. One drawback is the relatively high false-positive rate which can decrease the reliability of the solution, with the potential consequence of decreasing

the severity of the alerts perceived by the security administrator. Nonetheless, Machine Learning, Artificial Intelligence and statistical models are becoming more and more accurate in shaping the day by day entities' pattern. This solution is a significant security measure against new malicious threats, even more with some variables tuning and configuration maintenance.

By using together signature-based and behaviour-based IDSs, the rate of false alarms can be significantly reduced, thus achieving greater system accuracy. Signatures lookup allow the system to quickly detect known threats, mark them as malicious, and alert cybersecurity experts. On the other hand, the anomaly-based component detects unknown threats and variants of known ones. Chapter 4 details modern and cutting edge behavioural techniques along with the commercial UEBA system used for this thesis use-case.

3.4 IDS placement scenario

An IDS can be placed in front of the DMZ, inside the corporate LAN or as an endpoint agent. Depending on the placement, the IDS has different defensive purposes and can counteract a specific subset of attacks. To have the needed visibility, an IDS can be attached to the network in two ways:

- Mirroring switches traffic in SPAN or TAP mode so that a copy of every packet is sent to the IDS. In this way, the detection system correlates events and creates alerts, but is not be able to block an attack since the traffic only goes from the network device to the IDS machine.
- Inline, which allows the IDS to both receive the copied packets and block unwanted actions sending commands through an admin interface.

In Figure 3.1 are schematically represented some examples of NIDS and HIDS placement.

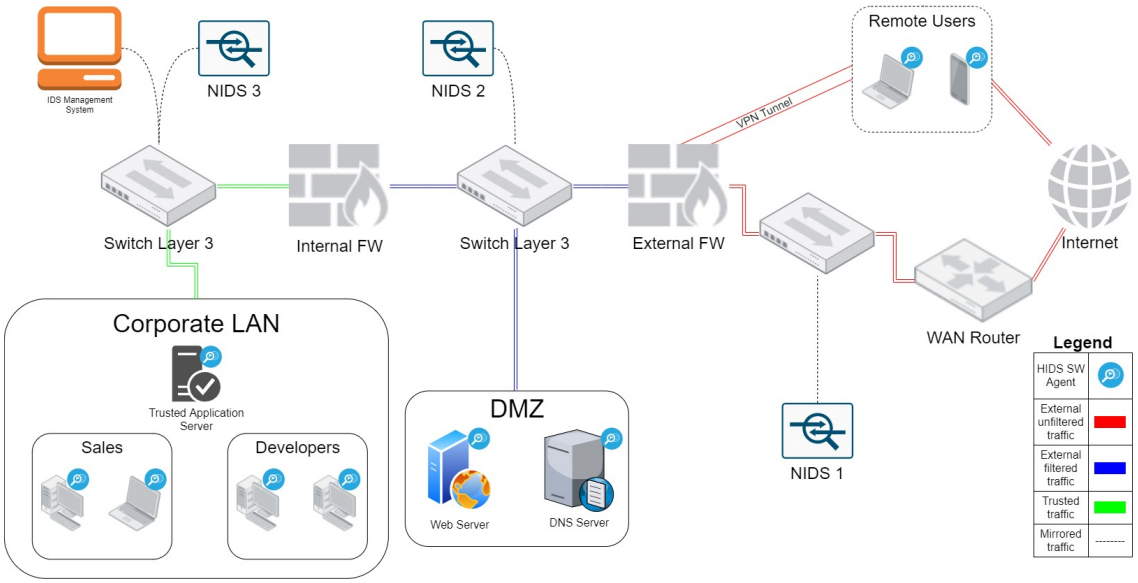

Figure 3.1: IDS placement scenario

The first NIDS is used to produce an idea of the traffic trying to access the company network and to spot port scanning or other attacks related to Layer 4 up to 7. The alerts coming from this NIDS will have a lower priority compared to NIDS 2 and 3 since the received traffic is not filtered at all. Moreover, IPS capabilities should be enabled to produce first gross filtering to reduce the amount of data that enters the network infrastructure, allowing the other security measures to analyse less, but more relevant, packets.

Then, there is the NIDS 2, which should be more sophisticated than NIDS 1 and will protect the DMZ. In case the NIDS is signature-based, the signatures should specifically target the DMZ systems to reduce the number of rules to process, thus requiring smaller bandwidth, less memory and limit the CPU usage. For example, it can be specified the protocols to monitor along with the OSs used by those systems. The same concept can be applied to a behavioural approach, fine-tuning the models which should be more used with the expected traffic.

The last network-based IDS, NIDS 3, is the ultimate defence against malicious actors which try to penetrate the company network and infect the company devices. In this case, it should be configured strict enough to reduce to zero the false negatives, while

minimising the false positives.

Inside the picture are also present host-based IDSs which are installed in every network's endpoint. It is mandatory to guarantee a complete overview of what is happening inside the organisation perimeter. In this case, it is crucial to consider that the HIDS can affect the endpoint performances negatively, which is especially true when the protected endpoint is a smartphone or a tablet which has limited computational resources. Therefore, the IDS agent running into it must be as lightweight as possible. In the article [12] are reported four IDSs suitable for smartphones along with their detection method and algorithms used.

Finally, there is the IDS Management System which collects all the events gathered by the IDS sensors deployed inside the corporate network. It correlates the shreds of evidence and makes them available, in a human-readable and informative format, for the analysts who are in charge of monitoring the company security status.

Chapter 4

State-of-the-art in behavioural intrusion detection

In this chapter, the focus is to explain two modern IDS approaches studied in the literature along with the commercial UEBA system implemented in the use-case scenario, named Darktrace. The first section is based on paper [13] and describes the implementation details and the results of intelligent IDS using Deep Learning approach. The subsequent section reports another IDS technique based on Data Mining detailed in the article [14]. Lastly, the third section gives an overview of Darktrace model and functioning.

This thesis is not drafted for Deep Learning, Data Mining or Machine Learning experts. Therefore, the following sections will not be overly technical or too specific. For more advanced details and testing results, the audience is invited to review the two source papers, [13] and [14].

4.1 Intelligent Deep Learning-based IDS

Paper [13] reports the technical details and the effectiveness of an IDS based on a particular type of Deep Learning approach, which is a Deep Neural Network (DNN), able to detect and categorise unknown and unexpected cyberattacks. The problems that this

research is trying to solve, as well as other researches, based on classic Machine Learning approaches, are trying to do, are above all the difficulty in keeping up with the continuous evolution of the behaviour of network entities and the pace with which attackers invent new exploits. For these and other reasons, many algorithms have been tested and evaluated using multiple data sets to find the most suitable and effective ones to counteract cyber attacks.

Few benchmark malware datasets can be found online, and they show how these IDS solutions perform. The results of the algorithm present in [13] will be better demonstrated later. The DNNs experiments have been run till 1,000 epochs with the learning rate ranging from 0.01 to 0.5. The approach consisted of testing the algorithm with the KDDCup 99 dataset and then apply the most effective DNN model to other datasets like NSL-KDD, UNSW-NB15, Kyoto, WSN-DS and CICIDS 2017 to conduct the benchmark. Moreover, their DNN model used several hidden layers to learn the high dimensional and abstract features of the IDS data representation.

In the paper is reported that after multiple evaluations, the tested DNNs performed better compared to classical Machine Learning classifiers. Moreover, they propose a highly scalable and hybrid DNNs framework called scale-hybrid-IDS-AlertNet to monitor in real-time the host events and network traffic to alert upon a potential malicious event is triggered.

4.1.1 Proposed framework

The main challenge that these researches try to solve is to create an IDS able to deal with the considerable amount of data and traffic created by complex IT infrastructures. Thanks to this quantity of data, a DNN model should be able to extract and differentiate patterns of what is legitimate and what is not, detecting deviations precisely. The other side of the coin is the complexity of effectively handling such data without losses or slowdowns.

The researchers developed the architecture using Big Data techniques, the Apache

Spark cluster processing platform, and the Apache Hadoop Yet Another Resource Negotiator (YARN) for the configuration. Distributed and parallel Machine Learning algorithms and optimisation techniques have been used to build a scalable architecture capable of handling a large amount of traffic. Besides, GPU cores have been employed to speed up and parallelise the analysis of network and host events. Then, the framework possesses two analytic engines, in real and not-real-time, to monitor the flow and generate alerts. Finally, it would be possible to handle even more events if more processing resources were provided to the system, thus increasing its computing power.

For what concerns the processes behaviour classification, Deep Learning algorithms differs from typical Machine Learning ones since the feature engineering, and the feature extraction steps are not needed. Therefore, in this research, text representation methods, associated with NLP, have been used to convert the system calls into feature vectors. Without going much more in detail, the three proposed NLP techniques are Bag-of-Words (BoW), N-grams and Keras Embedding.

The computational model chosen by the researchers is an Artificial Neural Network (ANN). Besides, they employed a Multi-Layer Perceptron (MLP) which is a type of Feed-Forward Neural Network (FFN) that, in turn, is a type of ANN. The way the MLP is formulated for many hidden layers results in a DNN model. This model is more advanced than the classic FFN because it uses a non-linear activation function, called ReLU, which is adopted by each hidden layer to reduce the gradient error and the state of vanishing issues. Moreover, ReLU is faster than other activation functions and helps the MLP model with the training phase with a high volume of hidden layers. Note that the neural network depth is represented by the number of hidden layers, while the width is equal to the maximum neurons. More details about the Loss function, the gradient descend optimisation technique, and the ReLU formulas are reported in [13].

Attack category	Description	Data instances - 10 % data			
		KDDCup 99		NSL-KDD	
		Train	Test	Train	Test
Normal	Normal connection records	97,278	60,593	67,343	9,710
DoS	Attacker aims at making network resources down	391,458	229,853	45,927	7,458
Probe	Obtaining detailed statistics of system and network configuration details	4,107	4,166	11,656	2,422
R2L	Illegal access from remote computer	1,126	16,189	995	2,887
U2R	Obtaining the root or super-user access on a particular computer	52	228	52	67
Total		**494,021**	**311,029**	**125,973**	**22,544**

Figure 4.1: Training and testing connection records from KDDCup 99 and NSL-KDD datasets - image retrieved from [13]

4.1.2 Datasets and statistics

The researches used the publicly available datasets to make their tests. Specifically, the KDDCup 99, NSL-KDD, UNSW-NB15, Kyoto, WSN-DS and CICIDS2017 datasets have been utilised to evaluate the NIDS performances. In contrast, the ADFA-LD and ADFA-WD datasets have been adopted to measure the HIDS outcomes. The researchers declared that such datasets do not sufficiently represent real-world network traffic, and they struggled to achieve good results and scale their solution correctly. Down below are reported the results of the researchers' algorithm related to the used datasets.

The KDDCup 99 dataset was created starting from the 1998 DARPA one processing its tcpdump data. Then, the KDDCup 99 redundant connection records along with the connections 136,489 and 136,497 were removed to create the NLS-KDD dataset. The results obtained by the researchers' solution on these two datasets is reported in Figure 4.1. Another dataset used to evaluate the algorithm is the UNSW-NB15 one. Its data was generated mixing normal and attack behaviors to solve the issues of KDDCup 99 and NLS-KDD datasets. The statistics and results summary table is shown in Figure 4.2. The Kyoto dataset is a mixture of KDDCup 99 and the traffic gathered by the honeypot systems of Kyoto university. The results are depicted in Figure 4.3. Then, there is the

Class	Description	Train	Test
Normal	Normal connection records	56,000	37,000
Fuzzers	Attacks related to spams, html files penetrations and port scans	18,184	6,062
Analysis	Attacks related to port scan, html file penetrations and spam	2,000	677
Backdoors	Backdoors is a mechanism used to access a computer by evading the background existing security.	1,746	583
DoS	Intruder aims at making network resources down and consequently, resources are inaccessible to authorized users	12,264	4,089
Exploits	The security hole of operating system or the application software is understand by an attacker with the aim to exploit vulnerability	33,393	11,132
Generic	Attacks are related to block-cipher	40,000	18,871
Reconnaissance	A target system is observe by an attacker to gather information for vulnerability	10,491	3,496
Shell code	A small part of program termed as payload used in exploitation of software	1,133	378
Worms	Worms replicate themselves and distributed to other system through the computer network	130	44
Total		**93,500**	**28,481**

Figure 4.2: Training and testing connection records of partial UNSW-NB15 dataset - image retrieved from [13]

Class	Training	Testing
Normal	2,384,645	1,405,391
Attack	670,037	158,532
Total	3,054,682	1,563,923

Figure 4.3: Training and testing Kyoto dataset - image retrieved from [13]

Class	Description	Train	Test
Normal	Normal connection records	238,046	102,020
Blackhole	It is a kind of 'DoS' attack where an attacker attacks LEACH protocol and during initial time itself they publicize themselves as a CH	7,033	3,015
Grayhole	It is a kind of 'DoS' attack where an attacker attacks LEACH protocol and during initial time itself they publicize themselves as a CH for other nodes	10,217	4,379
Flooding	Using different ways, an attacker attacks LEACH protocol	2,318	994
Scheduling	Scheduling attack happens during the setup phase of LEACH protocol	4,646	1,992
Total		**262,260**	**112,400**

Figure 4.4: Training and testing WSN-DS dataset - image retrieved from [13]

WSN-DS dataset conceived for wireless sensor networks (WSNs) gathering data from a simulation leveraging low-energy adaptive clustering hierarchy (LEACH) protocol. The details can be observed in Figure 4.4. The last dataset used to evaluate the NIDS solution is the CICIDS2017 one. It is one of the most recent datasets available and contains real-time network traffic along with recent attacks injected during the traffic collection phase. In Figure 4.5 are reported the operational results.

There are two datasets used to evaluate the HIDS which are ADFA-LD (for Linux systems) and ADFA-WD (for Windows systems). Inside the two datasets there are present system calls collected in different situations like normal programs behavior and known vulnerabilities exploitation. The performances of the researchers HIDS are detailed in Figure 4.6.

4.1.3 Results and considerations

The objective of the research was to compare the HIDS and NIDS based on common Machine Learning algorithms with the Deep Neural Networks (DNNs) one. The comparison points are three:

Class	Description	Train	Test
Normal	Normal connection records	60,000	20,000
SSH-Patator	Secure shell - Representation of brute force attack	5,000	897
FTP-Patator	File transfer protocol - Representation of brute force attack	7,000	938
DoS	Intruder aims at making network resources down and consequently, resources are inaccessible to authorized users	6,000	2,000
Web	Attacks are related to web	2,000	180
Bot	Hosts are controlled by bot owners to perform various tasks such as steal data, send spam and others	1,500	466
DDoS	Distributed Denial of Service ('DDoS') is an attempt made to make services down using multiple sources. These are achieved using botnet	6,000	2,000
PortScan	Port scan is used to find the specific port which is open for a particular service. Using this attacker can get information related to sender and receiver's listening information	6,000	2,000
Total		**93,500**	**28,481**

Figure 4.5: Training and testing CICIDS 2017 dataset - image retrieved from [13]

Dataset	ADFA-LD		ADFA-WD	
	Traces	**System calls**	**Traces**	**System calls**
Train	833	308,077	355	13,504,419
Validation	4,372	2,122,085	1827	117,918,735
Attack	746	317,388	5,542	74,202,804
Total	5,951	2,747,550	7,724	205,625,958

Figure 4.6: ADFA-LD and ADFA-WD datasets - image retrieved from [13]

Layers	Type	Output shape	Number of units	Activation function	Parameters
0-1	fully connected	(None, 1,024)	1,024	ReLU	43,008
1-2	Batch Normalization	(None, 1,024)			4,096
2-3	Dropout (0.01)	(None, 1,024)			0
3-4	fully connected	(None, 768)	768	ReLU	7,87,200
4-5	Batch Normalization	(None, 768)			3,072
5-6	Dropout (0.01)	(None, 768)			0
6-7	fully connected	(None, 512)	512	ReLU	3,93,728
7-8	Batch Normalization	(None, 512)			2,048
8-9	Dropout (0.01)	(None, 512)			0
9-10	fully connected	(None, 256)	256	ReLU	1,31,328
10-11	Batch Normalization	(None, 256)			1,024
11-12	Dropout (0.01)	(None, 256)			0
12-13	fully connected	(None, 128)	128	ReLU	32,896
13-14	Batch Normalization	(None, 128)			512
14-15	Dropout (0.01)	(None, 128)			0
15-16	fully connected	KDDCup 99- NSL-KDD- UNSW-NB15- Kyoto- WSN-DS- CICIDS 2017-	Binary, Multi-class: 1, 5- 1, 5- 1, 10- 1- 1, 5 1, 6	$Sigmoid$ for Binary and $Softmax$ for Multi-class classification	

Figure 4.7: DNN model parameters configuration - image retrieved from [13]

1. Decide whether network traffic is genuine or an attack, using all the features.

2. Decide whether network traffic is genuine or an attack, adding the attack category using all the features.

3. Decide whether network traffic is genuine or an attack, adding the attack category using minimal features.

Without going too much into the DNNs' parameters optimisation, in Figure 4.7 is visible the proposed DNN architecture with the already chosen parameters. In the paper [13] the reader can find much more details about the decision process. To summarise the findings and results of the paper, it can be said that multiple classical Machine Learning methods have been compared with their DNNs one based on publicly available NIDS and HIDS datasets. They achieved a training accuracy ranging from 95% and 99% on KDDCup 99 and NSL-KDD datasets, while for UNSW-NB15 and WSN-DS the accuracy

was around 65% to 75%. They also extracted the ROC curve for the other datasets and found out that DNN achieved better results in comparison to Machine Learning classifiers. Specifically, the DNN obtained a higher True Positive Rate (TPR) and lowered False Positive Rate (FPR), close to zero in some occasions. Moreover, the proposed method achieved an overall accuracy of 93.5% with few parameters required. Therefore, this method is computationally inexpensive and can achieve good performances in real-time scenarios and with unseen samples. The DNN model learnt how to categorise "Normal", "DoS" and "Probe" records correctly while it had some troubles with "U2R" and "R2L" records. The reason is that "Probe", "U2R" and "R2L" have common characteristics and, thus, additional features would be required to perform a correct classification. Overall, on KDDCup 99 and CICIDS 2017, both Machine Learning and DNN algorithms performed well, while the outcome is slightly less on NLS-KDD dataset.

To conclude, if an additional module monitored DNS and BGP events in the network, the proposed framework would perform better. The IDS does not provide information on the malware structure and its characteristics. Nonetheless, the researchers showed how good Deep Neural Network techniques are when the application is the IDS field. Further improvements could be achieved if advanced hardware would be used to train a complex DNNs architecture which was not possible for the paper researches due to budget reasons.

4.2 IDS based on data mining techniques

The scope of this section is to illustrate another novel IDS technique based on various Data Mining methods, as explained in detail in [14]. Specifically, the problems to tackle were:

- Perform automatic data classification to help system admins to keep the company surface as much monitored and clean as possible.
- Limit the human interaction in the pre-processing phase keeping high the accuracy

and detection rate.

- Limit the manual phase of traffic labelling, which is very cumbersome, in favour of automatic labelling.

- Mitigate DDoS attacks.

The goal was to solve these problems applying the Data Mining concepts taking as less execution time as possible while providing very relevant alerts. The techniques used to counteract the just mentioned four issues were, respectively:

- Efficient Data Adapted Decision Tree (EDADT) algorithm, which effectively separates attacks from normal traffic guaranteeing a high-quality traffic classification.

- Hybrid IDS model, which is based on SNORT for the signature-based part along with anomaly-based approaches to minimise network admins effort.

- Semi-Supervised Approach, which allows labelling a huge quantity of unlabeled data (unsupervised part) starting from a small set of labelled data (supervised part).

- Varying HOPERAA Algorithm uses varying clock drift to shuffle the ports used by legitimate users, preventing an attacker from accessing them.

4.2.1 Methodology

The general schema containing all the four methodologies proposed by the paper researchers is depicted in Figure 4.8. In the following paragraphs will be briefly explain the four proposed solutions.

In the EDADT algorithm, a hybrid Particle Swarm Optimization (PSO) technique is used to find the optimal solution out of n iterations. The process to obtain it consists of normalising the information gain of each attribute and decision node and calculate the average value, finding the exact efficient features from the training dataset.

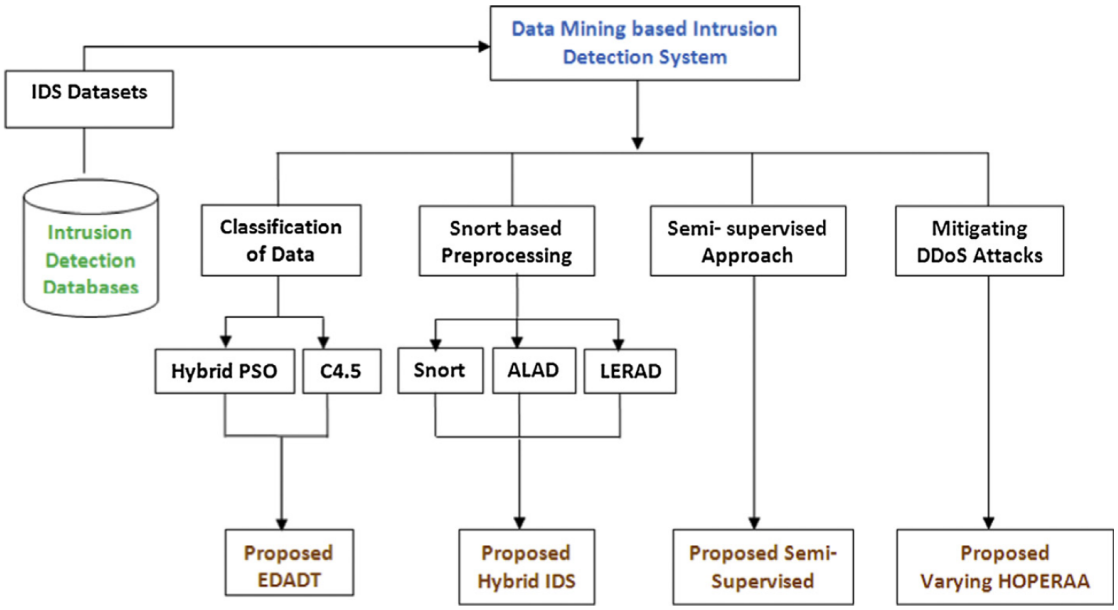

Figure 4.8: Representation of the entire Data Mining IDS framework - image retrieved from [14]

For what concerns the hybrid IDS solution, SNORT has been used to gather the real-time network traffic along with KDD Cup 99 dataset. Note that SNORT is a signature-based IDS. Therefore, other anomaly-based solutions were added to it to discover unknown attacks. These approaches are Packet Header Anomaly Detection, Network Traffic Anomaly Detector, Application Layer Anomaly Detector and Learning Rules for Anomaly Detection.

The Semi-Supervised Approach consists of divide the training and testing data where only a small fraction of the training data is labelled. Then, a Support-Vector Machine (SVM) is used to generate the models starting from the labelled data along with these models, so that the unlabeled data will be automatically labelled. The researchers obtained good results testing this technique with the KDD Cup 99 dataset.

Lastly, the researchers added a variable clock drift method to the previously studied Varying HOPERAA Algorithm to mitigate DDoS attacks. They worked on three main parts:

- The initiation of client-server communication, in which they introduced a pseudo-random function to decide which port to open to prevent attackers from impersonating a legitimate user during this setup phase.
- The data transmission phase, where the researches improved the performance reducing the message delivery latency.
- The Varying HOPERAA Algorithm is used to reduce the growth of waiting intervals among the messages exchanged between client and server. It is done by calculating the clock drift velocity of both client and server and allowing the client to be informed about this parameter, leading it to increase or decrease his clock drift.

4.2.2 KDD Cup 99 dataset overview

The researchers used the KDD Cup 99 dataset to perform their tests and evaluations. In brief, they categorised the attacks as follows:

- Denial of Service attack.
- U2R attack, which consists of legitimate user impersonation aiming to obtain root privileges.
- R2L attack, where a remote malicious subject exploits a vulnerability contained in a host inside the network.
- Probe attack, where the attacker scans the network trying to gather more information about the infrastructure.

All these categories contains 23 attack types as shown in Table 4.1.

4.2.3 Approaches statistics and considerations

The researchers used 60% of KDD Cup 99 dataset for the training phase, while the other 40% for the testing process. The four researched algorithms have been evaluated focusing

Attack Category	Attack Type
Denial of Service	Back, land, neptune, pod, smurf, teardrop
Probes	Satan, ipsweep, nmap, port sweep
Remote to Local	ftp_write, imap, guess_passwd, phf, spy, warezclient, multihop, warezmasterne
User to Root	Buffer_overflow, load module, Perl, root kit

Table 4.1: KDD Cup 99 dataset attack categories and types - table data retrieved from [14]

on accuracy and False Alarm Rate (FAR) values.

The EDADT algorithm was used to classify network data into either normal or attack category. The classification process did not leave any data unclassified, proving to be very efficient. The researchers compared their solution with other Data Mining techniques. They discovered that EDADT reduces the space occupied by the dataset, and it takes less computation and a lower FAR time when it operates in real-time. The performance results of the tested models are reported in Table 4.2. It is shown that EDADT performs better in terms of accuracy, sensitivity and specificity with a very much lower false error rate. The only drawback of the proposed model is the build time represented in Figure 4.9 which is higher compared to C4.5, SVM and C4.5 + ACO models. Overall, experimental results states that the researched model performs better compared to the other tested algorithms. Then, the hybrid IDS used to detect user behaviour anomalies and attack signatures, which has been developed to drastically reduce the human pre-processing interaction, was trained using 320 instances and tested with 180 instances coming from KDD Cup 99 dataset. As said before, the signature-based IDS employed in the research was SNORT, while the researchers used different types of anomaly-based IDSs. They were Application Layer Anomaly Detector (ALAD), Learning Rules for Anomaly Detection (LERAD), Network Traffic Anomaly Detector (NETAD) and Packet

Algorithms	Sensitivity (%)	Specificity (%)	Accuracy (%)	FAR (%)
C4.5	86.57	82.00	93.23	1.56
SVM	83.82	64.29	87.18	3.2
C4.5 + ACO	89.26	85.42	95.06	0.87
SVM + ACO	87.42	67.95	90.82	2.42
C4.5 + PSO	92.51	88.39	95.37	0.72
SVM + PSO	90.06	70.80	91.57	1.94
Proposed EDADT	96.86	92.36	98.12	0.18

Table 4.2: EDADT algorithm performance comparison - table data retrieved from [14]

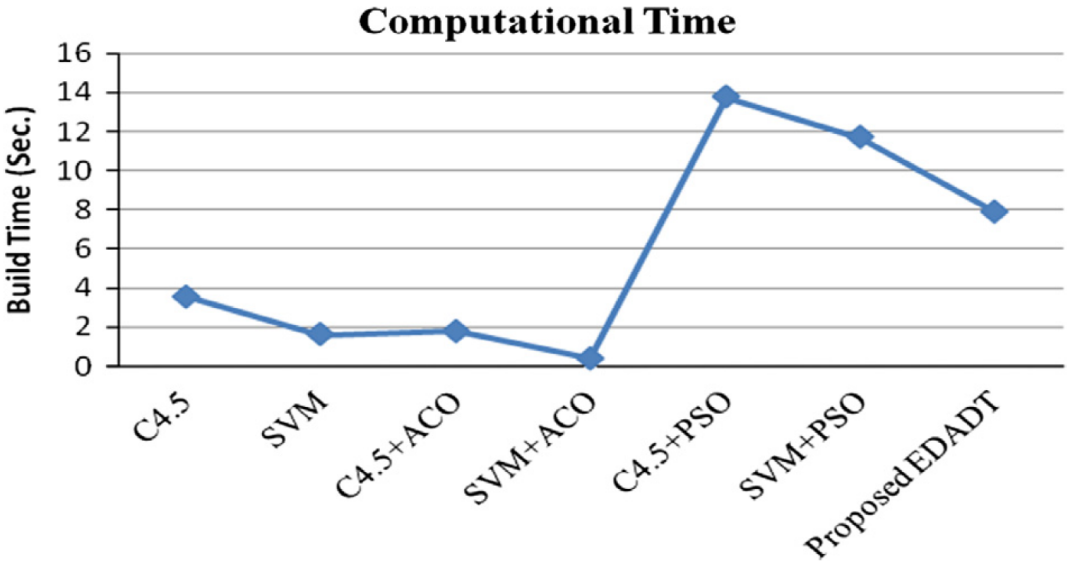

Figure 4.9: EDADT algorithm computational time comparison - image retrieved from [14]

Hybrid IDS Model	Attacks Detected (out of 180)	Detection Rate (%)
SNORT	77	42.78
SNORT + PHAD	105	58.33
SNORT + PHAD + ALAD	124	68.89
Proposed Hybrid IDS (SNORT + ALAD + LERAD)	149	82.78

Table 4.3: Hybrid IDSs performances - table data retrieved from [14]

Header Anomaly Detection (PHAD). The performances of the proposed hybrid models are reported in Table 4.3. It is clear from the numbers that the proposed method is performing better compared to the other versions. SNORT's rule-set is enhanced by LERAD algorithm and the detection rate becomes significantly higher.

Regarding the semi-supervised approach for IDS, the training phase requires a small amount of labelled data and a lot of unlabelled data. While the testing phase is performed on real-time network traffic using only unlabeled data. The attack categories considered for the algorithm evaluation were DoS, Probe, R2L and U2R. Figure 4.10 represents the comparison among the proposed Semi-Supervised Approach and other well-known algorithms like Reduced Support Vector Machine (RSVM), Semi-Supervised clustering algorithm (PCKCM) and Fuzzy Connectedness based Clustering (FCC). The graph shows that the accuracy of the proposed algorithm is 98.88%, just above the FCC one and much more than RSVM and PCKCM. Besides, in Figure 4.11 it is reported that the proposed Semi-Supervised Approach shows a minimal 0.5% of false alarm rate, almost the same of the best-performing algorithm in this category, which is PCKCM.

The last researchers' proposition called Varying Clock Drift Mechanism aimed to mitigate DDoS attacks. The researches compared the proposed algorithm with already existing work. They found out that Varying HOPERAA Algorithm performs better in terms of:

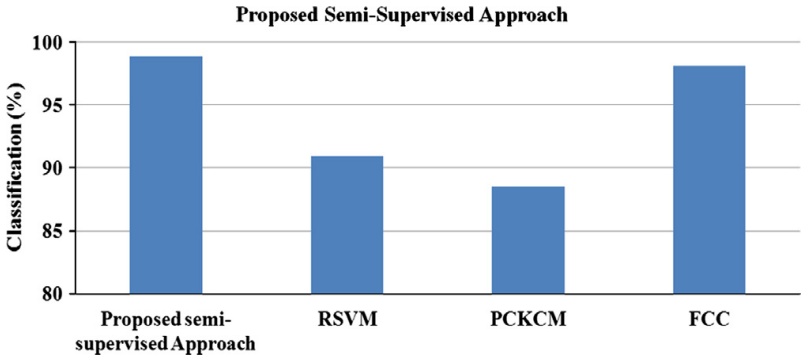

Figure 4.10: Semi-Supervised Approach performances comparison - image retrieved from [14]

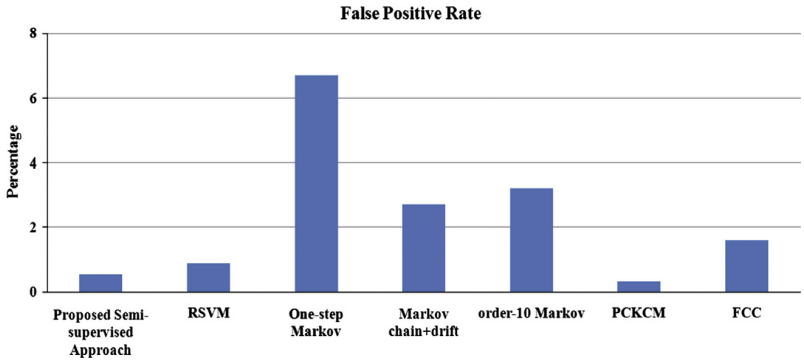

Figure 4.11: Semi-Supervised Approach false positive rate comparison - image retrieved from [14]

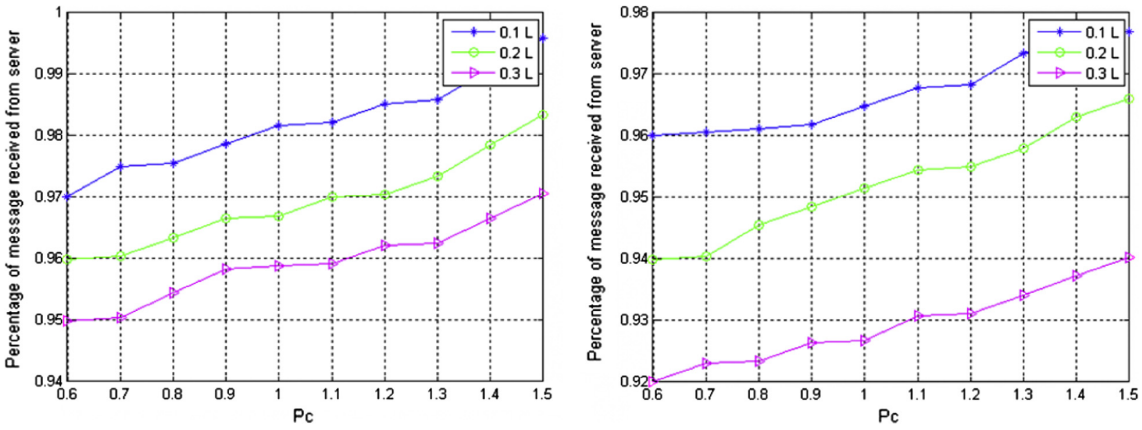

Figure 4.12: Server receiving capacity at 100ms and 40ms rates - image retrieved from [14]

- Contact initiation trails communicated between client and server.
- Reduction of execution intervals needed to estimate the client's clock drift.
- Minimal damage caused by message losses.
- Server's more efficient receiving capability.

Moreover, the message transfer delay and the execution time have been decreased by the researchers' algorithm. The server communicates the clock rate velocity information to the client through the clock drift details. After experimental tests, the receiving capacity at 100 ms of the Varying HOPERAA Algorithm is around 99% and note that it varies very little compared to 40 ms. At the same time, other existing algorithms show the worst deviation between the two rates, as shown in Figure 4.12. Besides, the proposed Varying HOPERAA Algorithm performances in terms of throughput and packet size based on the number of clients is higher compared to already existing algorithms, as illustrated in Figure 4.13.

In conclusion, this research showed how Data Mining techniques could be employed in the IDS field, achieving relevant results in terms of accuracy, efficiency and effectiveness. To develop even more this research, the researchers should provide more compu-

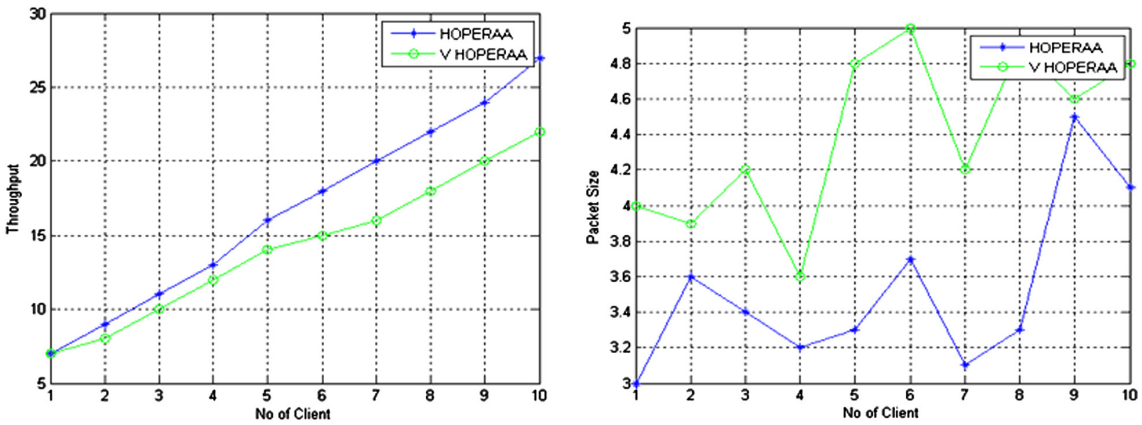

Figure 4.13: Performances of Varying HOPERAA Algorithm versus existing ones concerning throughput, packet size and number of clients - image retrieved from [14]

tational power and better model adjustment information techniques to automate the IDS properly.

4.3 Darktrace overview

This last section concerning the theory behind behavioural IDS reports how Darktrace Enterprise Immune System Antigena Email solutions work. The white papers took as supporting documents to write this deepening are [15] and [16]. Since this is a commercial product, there are no specific test dataset statistics and measures available. Nonetheless, this section shows the discovered information, relating to public domain resources. An overview of Darktrace Cyber AI Platform is shown in Figure 4.14. The IDS/IPS solution conceived by Darktrace is based on Supervised Machine Learning, Unsupervised Machine Learning and Deep Learning. These components present in the Darktrace solution will be separately explained in the following paragraphs. Altogether, these methods build Darktrace's AI technology which generates millions of interrelated mathematical models, for each unique production environment, indicating with some degree of certainty that behaviour is abnormal, significantly avoiding false positives proliferation. These degrees of potential threat level are very useful in establishing the priorities of the blue team's

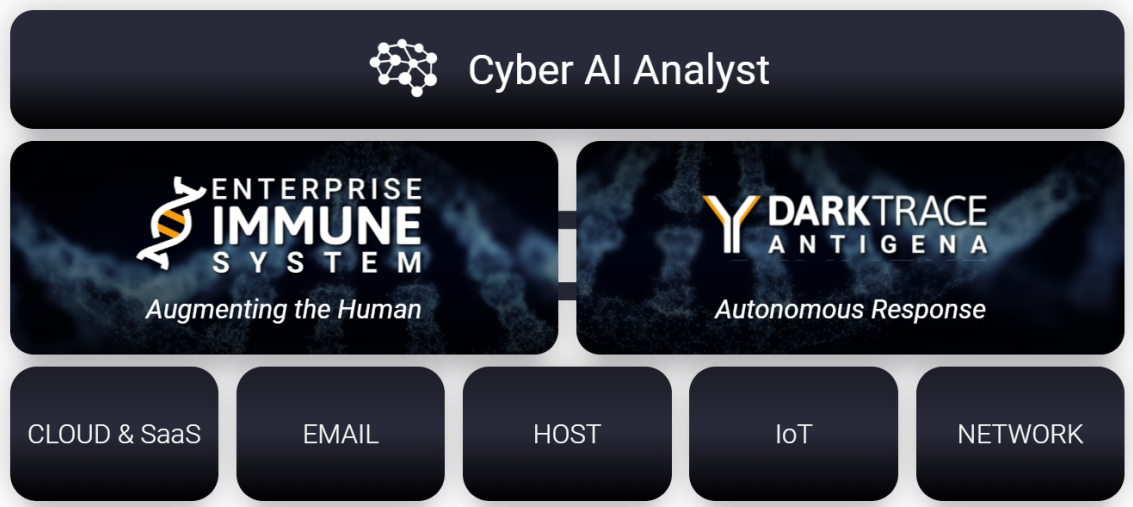

Figure 4.14: Darktrace Cyber AI platform overview - image retrieved from [15]

interventions starting from the most urgent ones.

Darktrace uses the Unsupervised Machine Learning component to discover previously-unknown threats, rare ones or variations of already existing ones and to group devices based on similar behaviours thanks to clustering methods like matrix-based clustering, density-based clustering and hierarchical clustering. In this way, the solution is not bound to outdated training datasets or labelled data, avoiding cumbersome manual pre-processing tasks and dissociate the decision making with historical attacks which may or may not be seen in future. This part is critical because despite labelled data is not required, data patterns and trends are still identified, allowing the Enterprise Immune System to understand whether system processes form new relationships. The scope is to shape the "normal" behaviour of all the entities inside the network seen by Darktrace, like users, devices and groups, and discover deviations from this entities' so-called "pattern of life". Some key points of Darktrace's Machine Learning, validated by thousands of deployments worldwide, are the following:

- The "normal" network traffic learning process is made during the execution phase, without previous knowledge is provided.

- Modern businesses are characterised by complex networks, several unique entities, different sizes and need of flexible and scalable solutions. Darktrace smoothly works in this kind of environments.

- Unusual activities are immediately spotted, inhibiting the attackers' innovative attacks.

- There is continuous monitoring of the "normal" entities' behaviour so that Darktrace, through probabilistic mathematics, can adapt to new business requirements and network changes.

- Human input is not required, and the solution is always up-to-date.

The just mentioned probabilistic mathematics developed by mathematicians from the University of Cambridge is based on Bayesian theory and used to understand the true meaning of unclear data. Darktrace's technology benefits from these studies in terms of new relationships detection, data classification, normal behaviour definition and independence from previous assumptions. Specifically. Iterative matrix methods are utilised to reveal important connectivity structures within the network, like advanced page-ranking algorithms. Moreover, the field of statistical physics has been used to develop models which discover the network's "energy landscape" needed to spot anomalous substructures representing indicators of compromise. Other two approaches used by Darktrace are L1-regularization techniques, based on the lasso method, and the Recursive Bayesian Estimation (RBE). These two methods are used to empower the Enterprise Immune System to understand the associations among the elements inside the network and to adapt the system based on the always-new available information, respectively.

Darktrace uses Deep Learning and Supervised Machine Learning techniques to supplement its AI engine with its cyber analysts' expertise, enhancing the modelling processes. Supervised Machine Learning is used to spot sequences of breaches, unusual patterns or to detect suspicious activities which need a more comprehensive view of the

external conditions to be meaningful. Thanks to this effort, Darktrace AI understands more about the environment characteristics becoming more precise and useful without the need for further human intervention. Deep Learning techniques allow Darktrace to automate actions and tasks needed for the investigation process, which, otherwise, should be performed by humans. This feature simplifies the data harvesting made by the analysts, which is needed to understand the cause of a threat and remediate it.

4.3.1 Darktrace Antigena Network

Once the Enterprise Immune System understands the network devices' pattern of life, Antigena Network module can be activated to empower the machine to neutralise a threat when it is triggered, forcing the device to function only according to its pattern of life. This IPS functionality is essential since it operates at machine speed, avoiding the spread of a threat. At the same time, the infected device/user can still operate under of its pattern of life boundaries, mitigating the threat and blocking only harmful actions, allowing the system administrator to intervene and resolve the issue. This feature has two modes, the human-confirmation one and the autonomous one. In the first case, a human operator must confirm the actions proposed by Antigena, having more control over the IPS capabilities. The second mode allows Antigena to be human-independent, increasing the effectiveness and reducing the response time. In both cases, an Antigena Network action can be reverted if not correct or when the threat has been remedied. Antigena is a surgical tool, able to target only the specific protocols used by a compromised device to perform malicious actions, blocking them. The management interface of the Darktrace appliance must be connected to the company network, and it must be able to reach the network devices to send TCP reset packets.

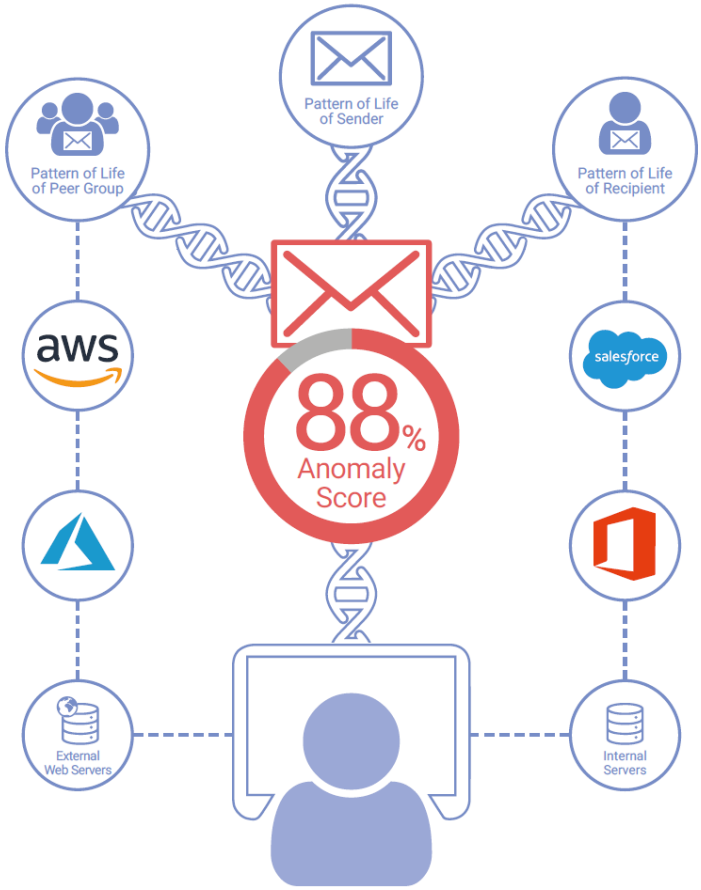

Figure 4.15: Darktrace Antigena Email components schema - image retrieved from [16]

4.3.2 Darktrace Antigena Email

Given that one of the major utilized attack vectors is the e-mail service, Darktrace designed a "new product" which has the same engine of the Enterprise Immune System, but focuses on e-mail threats. In Figure 4.15 is depicted a schema of the macro-elements taken into account by Antigena Email to decide whether to block the incoming e-mail, remove active content like links or attachments or let it through. Some attacks which can be avoided thanks to Antigena email are spear-phishing, payload delivery, supply chain account takeover, social engineering and compromise of credentials. Antigena Email can be added to the Enterprise Immune System to enhance his functionalities extending the monitoring also to the email environment, or it can be used as a standalone system. Note that in the first case, both the systems take advantage of the other one's information,

understanding even better the reason behind a suspicious event thanks to more context available. A simple example of this mutual help is that a user receives an email with a suspicious link and, after a few seconds, his PC starts a network scan, an action never done before. These two behaviours will then be associated, enriching the context and helping the analyst to resolve the breach.

Chapter 5

Use case definition

At this point, it should be clear what are the main threats coming from inside the company network, what is the scope of IDSs in protecting the business from such threats and the technologies and techniques behind behaviour-based IDSs. This chapter defines how to position Darktrace within the network correctly, the use-case scenario, the pain points before the Darktrace implementation and the objectives to achieve at the end of the test period.

5.1 Darktrace placement

This section details the Darktrace placement procedure and deployment examples. To allow Darktrace's Enterprise Immune System to gather meaningful data of the company's network is mandatory to place the physical appliance in a strategic place where it can ingest most, if not all, of the network traffic. There are some examples of information-rich traffic flows like DNS resolution, DHCP traffic, internal services access, traffic between internal devices and others. Briefly, the focus of Darktrace is to analyse this data to understand the "pattern of life" of all the users and systems inside the network, along with the network as a whole, creating unique behavioural models using Machine Learning and Artificial Intelligence algorithms. Therefore, the more data it analyses, the more accurate

the behavioural fingerprint of every network entity will be.

An organisation can deploy Darktrace in several ways to achieve network, cloud platforms, SaaS tools and virtual environments coverage. The data can be ingested using network TAPs, layer 2 SPAN (Port mirroring) or Layer 3 SPAN (VLAN mirroring). Darktrace has four elements that can be installed on the network to capture the traffic:

1. Darktrace Physical Appliances, attached via port mirroring or network TAP to ingest the data in transit which will be analysed to build the "pattern of life" of users, devices and subnets.

2. Darktrace Virtual Sensors, which are lightweight vSensors installed as virtual appliances in an on-premise server and configured to receive the mirrored traffic from the virtual switches. Then, the data is refined and securely sent to the master appliance located in the physical network.

3. Darktrace Sensors, which are lightweight host-based os-sensors, usually installed in cloud endpoints to send the copied data to a local vSensor deployed in the same cloud environment. Then, the vSensor will act like described before.

4. Darktrace SaaS Connectors configured to gather the logs generated by SaaS applications and securely send them to the Darktrace master appliance.

Figure 5.1 represents the first basic placement schema. There is a central network device which routes the traffic from the DHCP and proxy servers to the clients allowing the communication between the two separated subnets. In this case, the Darktrace appliance can be attached to the Layer 3 port mirror of the switch so that it can capture the traffic.

Figure 5.2 depicts another placement example. The Branch Office is a geographically separated from the HQ, and inside it, there is a local DHCP Server and a Darktrace Probe. A site-to-site IPsec VPN tunnel connects the HQ and the Branch Office so that the latter can access the Application Server placed inside the HQ. Both locations have an internet connection. The Darktrace Probe will capture the traffic inside the Branch Office network

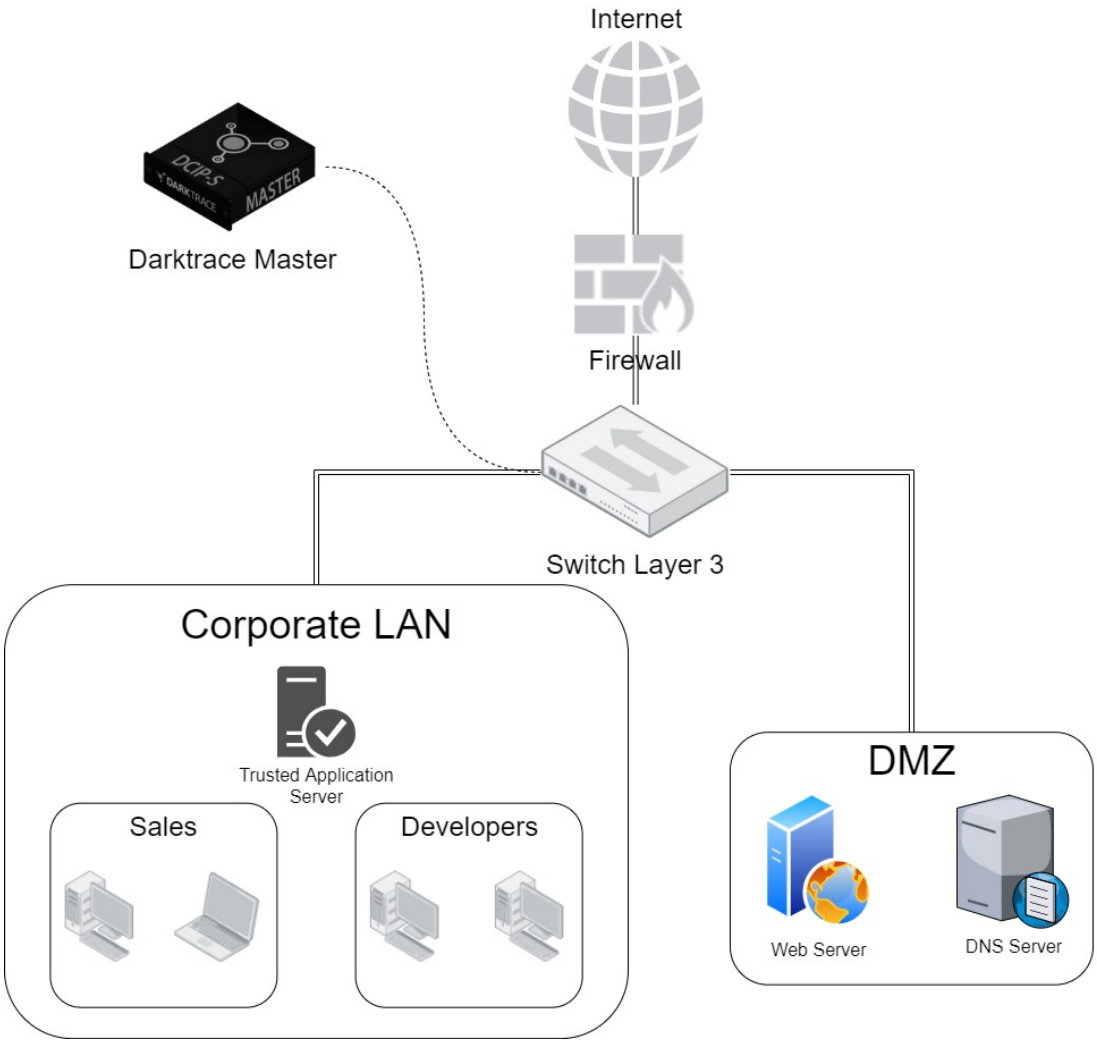

Figure 5.1: Darktrace basic deployment scenario

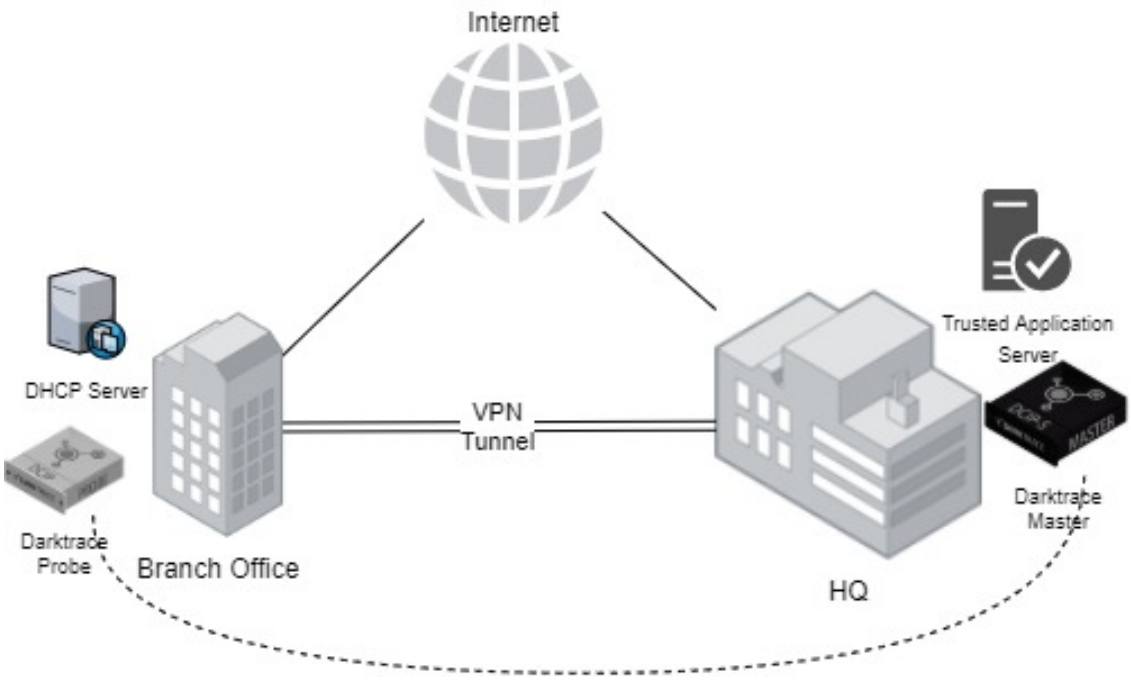

Figure 5.2: Darktrace Master/Probe deployment scenario

and forwards the metadata to the master appliance. Note that on average only the 0.5% of the traffic volume ingested by the Probe is sent over to the master for more in-depth analysis and all the communications are encrypted through IPsec tunnels. The major advantages in this situation are that the user interface is only one, the alerts are provided only by the master appliance along with the behavioural database maintenance, while the data capturing, deep packet inspection and other computational loads are separated. Last but not least, the Probe gives to Darktrace Enterprise Immune System full visibility on what is happening inside the Branch Office. If the Branch office does not install a Darktrace Probe, the behavioural models would be blurred, which means that partial visibility would have led to erroneously classifying traffic coming from the remote office, causing incorrect judgments.

Figure 5.3 shows the last schema, which represents a more advanced Darktrace deployment. It comprehends not only a branch office but also a cloud environment and a SaaS application which need to be viewed by the Enterprise Immune System. Cloud sen-

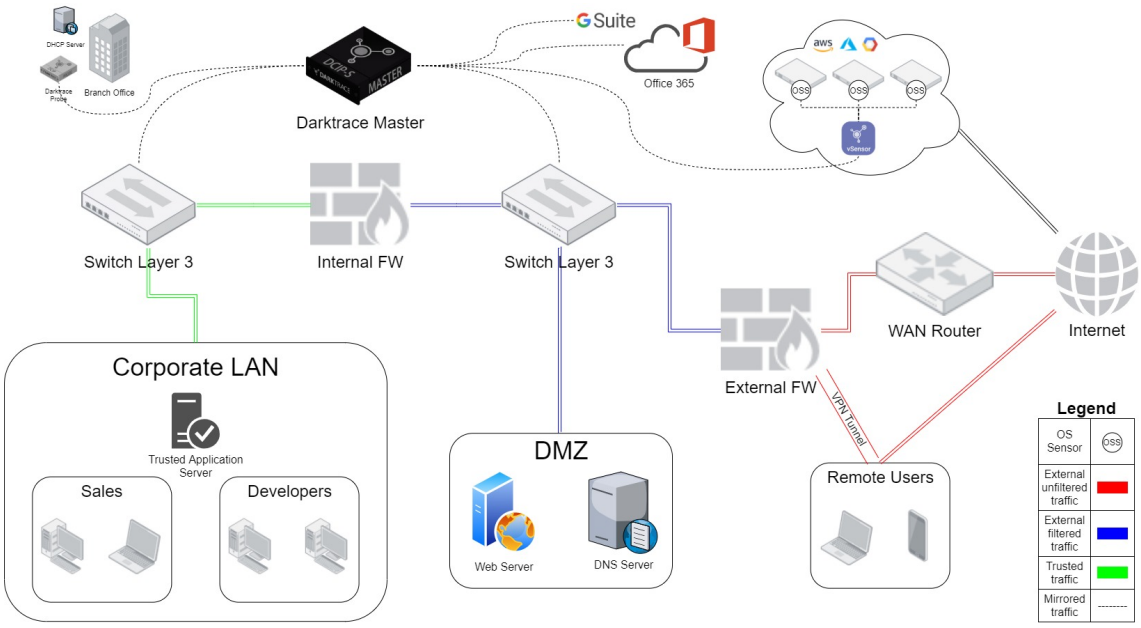

Figure 5.3: Darktrace Advanced Deployment Scenario

sors and OS-Sensors can be deployed and configured to capture the data and send it over HTTPS to the company's internal network to reach the master appliance. These sensors are employed in cloud IaaS such as Amazon Web Services (AWS), Google Cloud Platform (GCP) and Microsoft Azure. To connect a SaaS application with Darktrace, a SaaS Connector must be configured so that Darktrace can communicate to the API of the company's SaaS environment. In this way, Darktrace will be able to retrieve several useful metrics.

Multiple masters and probes appliances could be installed and connected to fulfil complex organisation requirements. The traffic flow should follow the same path explained in these three scenarios and can be done deploying sensors, appliances and other Darktrace components throughout the organisation network.

5.2 Company size and details

The company used as test environment provides air conditioners and refrigerators to other businesses and clients, from large installations in cruise ships to small ones in shops or offices. It has more than 300 employees and by European standards [17] is considered a medium-sized enterprise. Moreover, it has offices in Italy, where the HQ resides, and in the U.S., with an investee company in Finland. Inside his infrastructure, there are many IoT devices powered by uncommon OSs, PLC machines, servers, SaaS applications, phones, laptops and other more common devices which can be found in almost every workplace. There are also remote workers and many business consultants who are always moving around the globe. This company is characterised by a marked heterogeneity that makes it very difficult to manage and protect with legacy tools.

5.3 Shortcomings before Darktrace deployment

Unfortunately, this company never firmly focused on the cybersecurity part, creating a silo of security solutions characterised by separated consoles and logs. As if it was not enough, they only have one security expert in charge of every IT aspect. This situation leads to inadequate monitoring of legacy systems inside the network, so like limited visibility on users actions and prolonged response times in case of critical issues. From these points arose the need for a unified solution, easily manageable by a single person or by a small group of people. To be effective, this technology needed to be able to provide very granular visibility into the health of network components, both within the corporate perimeter and externally. Also, it was necessary to monitor remote workers, SaaS applications and resources installed in cloud environments.

5.4 Darktrace deployment process

The value of Darktrace's Enterprise Immune System lies in his self-learning capabilities. The Cyber AI platform protects the company from cyber threats automatically. Darktrace monitors user and device behaviour both within the network and in SaaS applications such as Microsoft Office 365, Dropbox or GSuite. Moreover, it provides in-depth network visibility, and the user-friendly web interface helps with the breaches triaging phase. The sizing phase is the first step to deploy the Enterprise Immune System. Darktrace intelligence resides in a physical machine, which must have enough computing power to analyse the copied network traffic. few parameters are required to select the correct Darktrace appliance:

- Number of IP addresses and subnets.
- Number of company branches.
- Authentication methods.
- Brand and model of core switches.
- Average and peak throughput.
- Cloud and SaaS applications.

After collecting this information it turned out that the number of estimated IPs was about 500, divided into 40 subnets, the estimated traffic volume was less than 1 Gb/s, and there was Office 365 as an email provider. It has been decided to use a DCIP-S physical appliance which is shown in Figure 5.4. That machine has one out-of-band (OOB) interface, one 1Gbe admin interface and three 1Gbe analysis ports. This type of appliance is ideal for small deployments with a limited number of devices and low generated traffic. The appliance has a peak sustained throughput (meaning 95th percentile of bandwidth ingestion) up to 300 Mbps, can monitor up to 1000 unique internal devices and allows at maximum

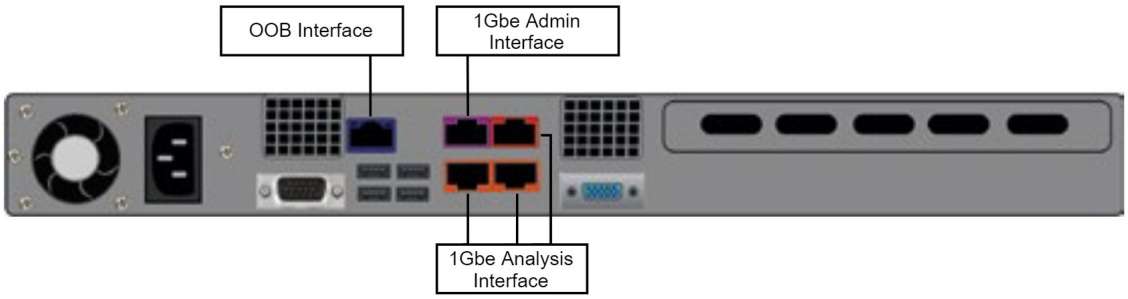

Figure 5.4: DCIP-S physical Darktrace appliance

Component	Port	Direction	Required?
Threat Visualizer and web configuration	443 (TCP)	Inbound	Required
Console application and file transfer via SFTP	22 (TCP)	Inbound	Required
Network Time Protocol	123 (UDP)	Outbound	Required
Syslog ingestion of mapping data	514 (UDP)	Inbound	Optional
DNS querying	53 (TCP & UDP)	Outbound	Optional
Remote management KVM	80, 443, 7578 (TCP)	Inbound	Optional
Call Home	22 (TCP)	Outbound (to specific IP)	Optional

Figure 5.5: Ports to open to configure Darktrace appliance

2,000 connections per minute. The appliance was prepared and shipped to the company HQ. For the configuration part, the company was required to define some firewall rules to allow traffic through specific ports for a set of IPs to allow the appliance to communicate with the Darktrace servers located in Cambridge. These ports are shown in Figure 5.5. a static IP address has been assigned to the appliance, and an NTP server has been configured. The appliance was connected to the SPAN port of the company core switch to ingest network traffic. At that point, the testing period began. Week after week, the data presented by the Enterprise Immune System were analysed. The most interesting results will be explained in the 6 chapter.

Chapter 6

Test analysis

The test period lasted from the end of February to the end of March, for a total of three months. During that timeframe, the Enterprise Immune System learned the behaviours of the systems, devices and users. From that information, it built the models needed to compare the network traffic generated by an entity with its pattern of life. In the following sections, remarkable statistics like the total alerts fired during those three months and the number of discovered devices compared to the forecasted ones have been written. The section below indicates some examples of potential breaches discovered by Darktrace and how they were escalated and remedied. Finally, in the last section, some negative findings will be reported.

6.1 Statistics

From the first implementation day, the Darktrace appliance started to ingest traffic passing through the HQ core switch where all the traffic converges in a north-south fashion. The company expected to have approximately 500 devices within its network. After a few hours of analysis, Darktrace showed that there were over 650 active IP addresses. In the graph represented in Figure 6.1 it can be seen that the number of active devices decreased during the most acute period of the COVID-19 pandemic, from the first days of March

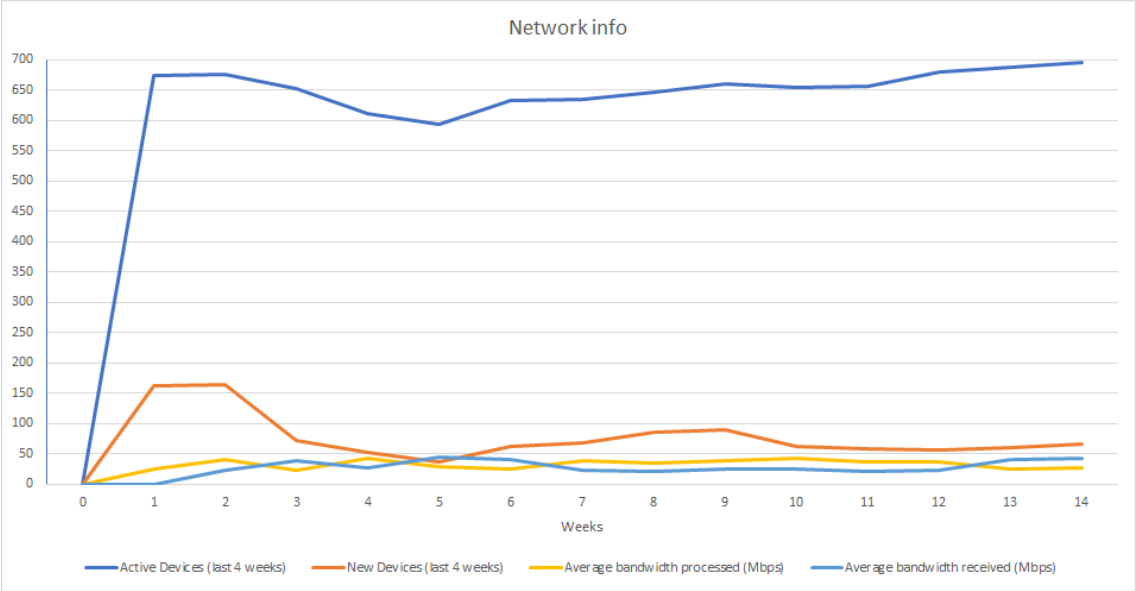

Figure 6.1: Company network statistics

for three weeks, and then rose again reaching little less than 700 devices. Coherently to the total number of devices, there is the line corresponding to the newly discovered ones (meaning the devices never seen in the last four weeks) which increased and decreased accordingly. Furthermore, the average bandwidth received increased during the two weeks of forced remote labour caused by the COVID-19 pandemic, and the average processed bandwidth peaked during that time.

There are several models created by the Enterprise Immune System to categorize the monitored traffic and correlate the events. The specific models are always updated, but they can be summarized in categories like attack models, compliance models and monitoring models. Attack models include events that have been identified as potential attack stages. This division will be reported and explained in a few paragraphs. Then, the monitoring models observe the breached attack models and trace the behavioural characteristics of the subject that triggered the anomaly, to define an in-progress attack or any other violation. Lastly, the compliance models indicate incorrect behaviours which may or may not be dangerous but does not stick to the company policies. Also, the various compliance infringement types will be reported in this section.

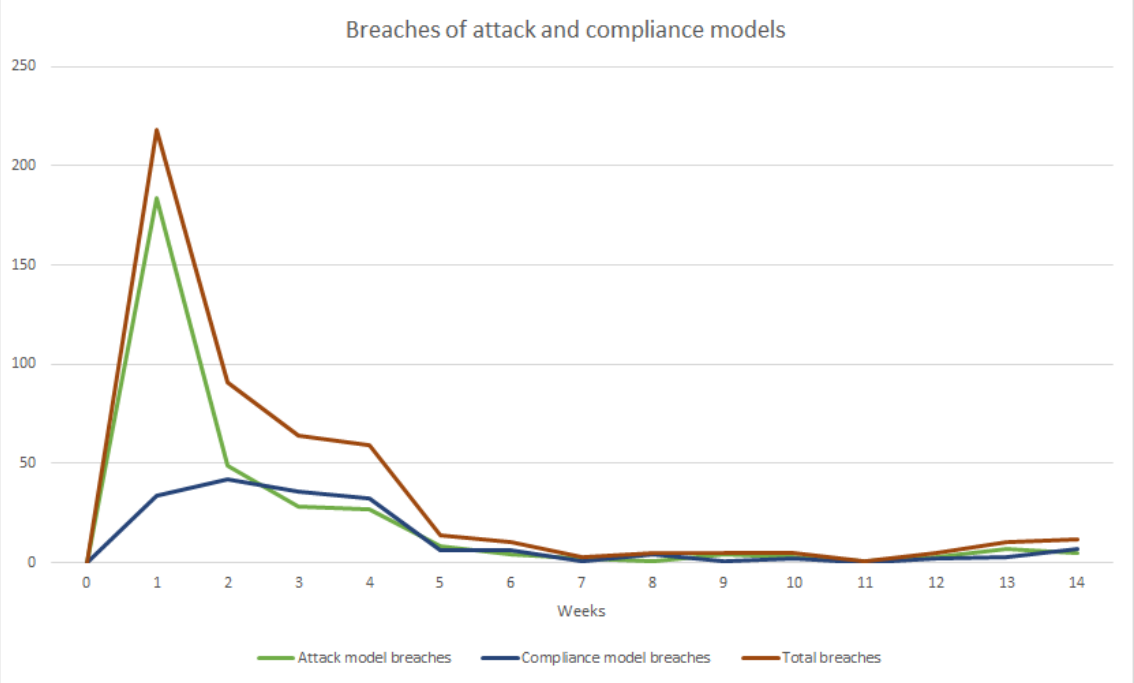

Figure 6.2: Trend of attack and compliance breaches week after week

Figure 6.2 shows the trend of the attack and compliance breaches detected by the Enterprise Immune System during the monitored weeks. When the appliance was installed, the number of alerts was tremendously high because the Machine Learning engine was seeing for the first time that kind of traffic. Already after two weeks, the number of alerts rocket down, to reach reasonable volumes from the fourth week onwards. The graph trend is the very same of other installations that we made throughout these months, and this is the reason why the POV (Proof of Value) period of the Enterprise Immune System lasts four weeks. Note that no manual configuration has been required during the first weeks to low down the number of alerts. Instead, the platform autonomously modelled the normal behaviour of the devices and users and, based on the acquired knowledge, generated the breaches.

As previously announced, there are different categories used by Darktrace to divide the attack and compliance models breaches. In the following lines are explained the classifications needed for the use-case, but there are more in the Darktrace environment.

For what concerns the breaches of attack models it can be found:

- Bruteforce; which means that a device is initiating a large number of connections to a server - this may be consistent with a user attempting to gain access using common credentials.
- Egress; meaning that a device moved a large amount of data to unusual file storage like Dropbox or WeTransfer.
- Exploit; this category is applied to actions which may indicate vulnerability exploitation made by an attacker or malware. For instance, a DNS request to an unusual server or a vulnerable name resolution.
- Internal Recon; the abbreviation stands for reconnaissance, and this category comprehends the events which concern the information gathering process made by a subject within the company network.
- Lateral Movement; in this case, the model includes activities related to strange movements inside the network, which an attacker does to find credentials, schedule tasks or other activities that would lead him to pursue his intent after an initial foothold.
- Tooling; when an action includes the use of tools like TeamViewer, Zip or Gzip, EXE uploads or downloads and other similar events which deviate significantly from the observed "pattern of life".
- Scanning; the events included in this category are related to network scanning actions. An example is the usage of Nmap or other tools designed to identify the devices inside the network.

Along with the just mentioned categories, there are the breaches of compliance models which comprehend:

- Dropbox; since many companies allow only specific cloud storage, like in the case of the tested one. Therefore, all the events related to data upload or download from Dropbox are reported as a compliance violation.

- Pastebin; this is a website where people can share text with other users. Usually, it is used to share programming code. However, an employee could also share sensitive documents via this website. For this reason, the usage of this tool is a potential compliance issue.

- Possible Cleartext Password In URI - external; this is self-explanatory since it is triggered when inside the URI is detected a clear-text password. It is a serious compliance violation since that password could be sniffed by attackers and probably the website is storing it unhashed.

- Possible Unencrypted Password Storage; this template includes Excel files, notes, Word files or other types of files that are not indicated for saving passwords. Darktrace recognizes these files also based on their title, such as "Sam_Password.txt" or "sales_department_credentials.xlsx".

- Remote Management Tool On Server; this model shows the servers that have a remote management tool installed. Two examples are TeamViewer and Remote Desktop Manager.

- SMB Version 1 Usage; Server Message Block (SMB) is an application layer protocol that provides shared access to files and other objects for devices in a network. The last version is SMBv3 and has built-in security features, while SMBv1 is deprecated and is not secure at all since it allows anonymous access by default. Therefore, if this protocol is used, Darktrace will fire a compliance model breach.

- Sensitive Terms in Unusual SMB Connection; this model is related to sensible information sharing using the SMB protocol and seen by the Enterprise Immune Sys-

tem.

- Unusual Data Volume to Dropbox; this model is related to the Dropbox model explained before, but it is triggered only in case there is a large amount of data exchanged with a Dropbox instance. Therefore, this is a more relevant compliance violation compared to the similar one indicated before because the user may be trying to exfiltrate a considerable amount of data from the company network.

- Vulnerable Name Resolution; this model is related to DNS vulnerabilities which lead an attacker to fool the DNS system, redirecting the users to a server controlled by himself instead of the required one.

- WeTransfer; like in the case of Dropbox, in case the company does not have a policy that allows WeTransfer usage once a user uploads or downloads files from that service Darktrace will spot that acts as a compliance infringement.

Now that the models have been explained it is possible to analyze more in-depth what Darktrace found out during these months. The two graphs represented in Figure 6.3 and Figure 6.4 contain all the details about the type of models breached during the weeks. In the first one are represented the attack models, while the second includes the compliance ones. Note that the reported percentages refer to the total number of breaches that occurred in that week. For example, at the end of the first week about 180 attack models were breached and, as it can be noticed in Figure 6.3, 60% of them belonged to the "Internal Recon" category.

Thanks to this high-level view, it is possible to analyze the most common attack models detected by Darktrace, which were Internal Recon, Lateral Movement and Tooling. On the other hand, frequently violated compliance models comprise Dropbox, which was very popular during the first few weeks, Possible Unencrypted Password Storage, which was not resolved during the monitored period and, as last, Vulnerable Name Resolution. In the following two sections are reported the positive and negative findings based on the

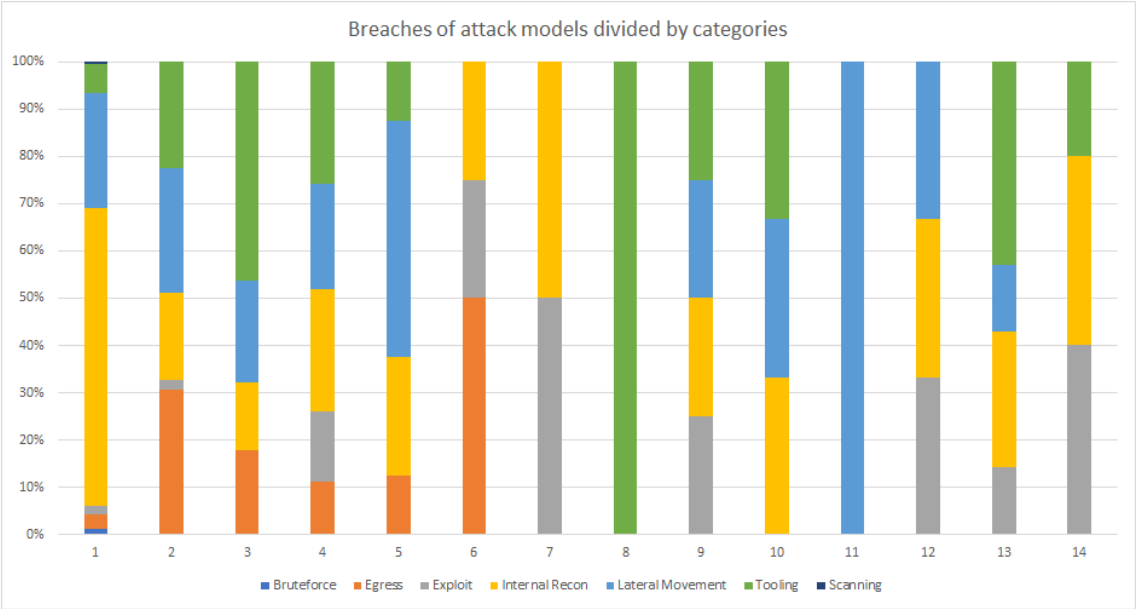

Figure 6.3: Breaches of attack models divided by categories

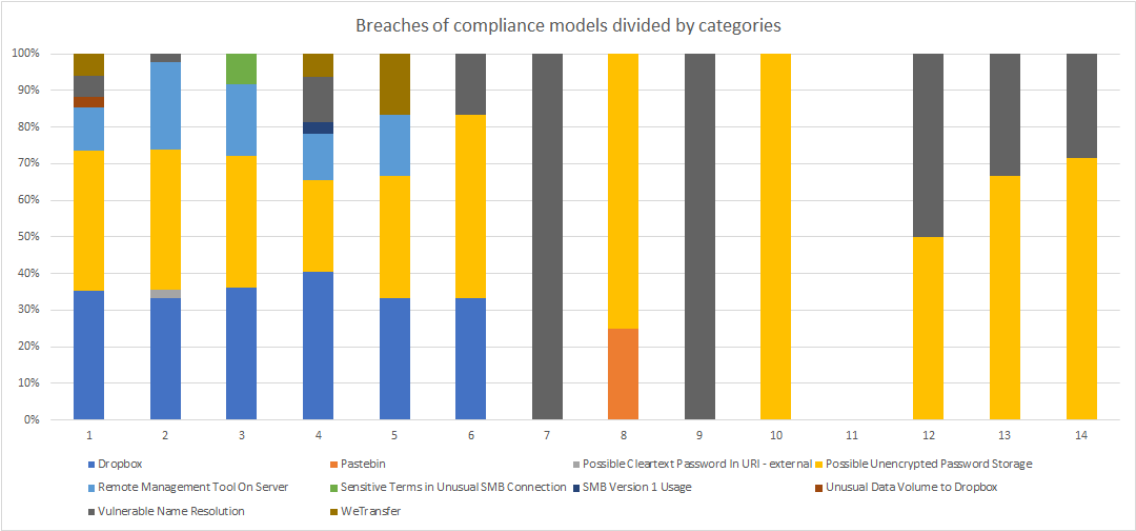

Figure 6.4: Breaches of compliance models divided by categories

breached models, the actual security threats which were escalated and resolved, the time spent for the analysis phase and the total effort made to use and customize the platform.

6.2 Positive findings

Among the positive points identified during the monitored period, there are visibility, ease of use, customization, manageability, the notification system and the threat investigation process. All these characteristics are briefly explained in the following subsections, along with a little explanatory example.

6.2.1 Visibility

As regards visibility, the solution allowed the system administrator to have a detailed view of the events that occurred within the network. All the devices that were making connections were identified and, in some cases, the system administrator was surprised that he was not aware that specific detected devices were connected to the network. Furthermore, it was possible to identify incorrect network-level configurations, which were mitigated and resolved by adjusting the firewall settings, better segregating the subnets and devices. Besides, a SaaS connector was configured to allow Darktrace to monitor the company's Office 365 platform, collecting the events and logs generated by O365 and reporting any incorrect or unusual behaviour.

6.2.2 Ease of use

Darktrace turned out to be very easy to use thanks to the user-friendly interface. For instance, the example reported in Figure 6.5 graphically represents the connections made from the central device to other network devices, like servers and clients, and external locations geographically located in the globe. Then, other two fundamental characteristics are the informative alerts created once an event breaches a model and the correlation made

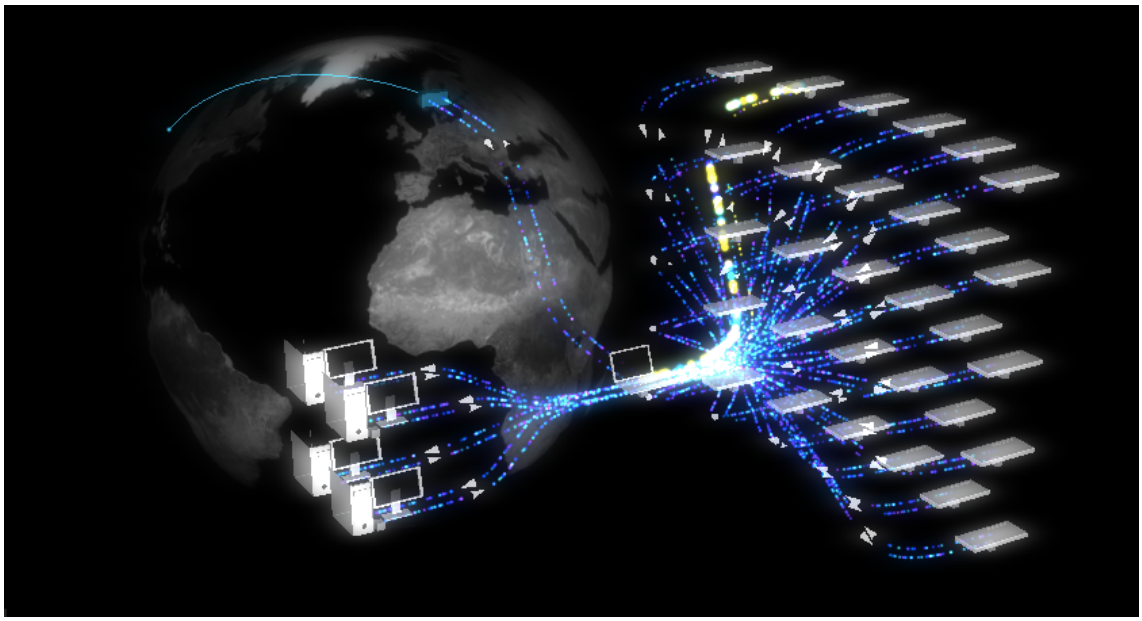

Figure 6.5: Darktrace graphical representation of device connections

by the Cyber AI Analyst able to autonomously put together different events and create storytelling about the identified attack or compliance infringement.

6.2.3 Customization

As regards customization, it was possible to specify an informative name for each subnet. This simple action led to a better contextualization of security events since, in addition to the IP and subnet, a descriptive name such as "Sales Department" or "IP Phone Subnet" was available. Besides, the devices used by the network administrators have been appropriately tagged. In this way, Darktrace stopped reporting actions made by those devices, which could be interpreted as risky. Two examples related to this topic are:

- A device that performs administration operations such as network scan or multiple SSH accesses.

- a server that connects OT devices but also makes connections to the Internet. If these two types of connections are made from the same machine, it can be an

anomaly. The Enterprise Immune System will detect it since a server should make only one of these two connections for security reasons.

Another example is that some subnets have been further divided to facilitate device tracking. The network mask has been changed from /24 to /30 to separate devices under DHCP from those with static IP addresses. Thanks to this differentiation, the corporate immune system was able to track the devices better knowing that some of them were changing IP due to a DHCP server, and therefore were identified by using more of the hostname rather than just the IP address. Finally, some devices were accessed by multiple users. Therefore the system was set up to track them according to the different access credentials used, associating the actions with the active credentials instead of just the device.

6.2.4 Manageability

Usually, manage an entire network could be problematic for a single person or a small group of people. There are too many events to investigate to gather enough information and decide if a breach occurred. Darktrace solved this problem with the Cyber AI Analyst, which employs the Artificial Intelligence engine to correlate security breaches to give an extended and context-rich explanation of what is happening or happened inside the network. Moreover, the Cyber AI Analyst suggests some remediation procedures to solve the issue and, along with Antigena Network, which is the IPS part of the solution, helps the system administrator to manage the network and keep security intact.

6.2.5 Notifications

There are two primary notification sources used to inform the platform administrator about a potential breach. The first one is through e-mail, which is a standard method also used by other security solutions, while the second one is unique and consists of smartphone push notifications. The Darktrace's notification application works for both iOS and Android operative systems. It receives the most relevant alerts created by the

Enterprise Immune System, enriching the message with all the needed information and allowing the receiver to activate the autonomous response whether the latter is in "human confirmation mode". There is a third, less used, way to warn the administrator, which is through a phone call. This method is only used in critical cases, very much related to an ongoing attack.

6.2.6 Threat investigation

Once an alert shows up in the platform environment, it is mandatory to examine it and decide whether it is a true positive or not. Based on this statement, with Darktrace technology, it is possible to dig into the events analyzing packet per packet what was transferred between the source and the destination to define when a breach started. Furthermore, the events are rich in context, and there is the possibility of using the "Time Machine" to go back to the point where the violation began and verify second by second the actions taken by the threat source. This feature is shown in the example reported in Figure 6.6, which represents the moment when Antigena went into action and blocked the malicious actions of the device, characterized by a luminescent halo surrounding it. Note that the image has been taken from a demo environment and is not related to the company used as a real use-case for privacy reasons.

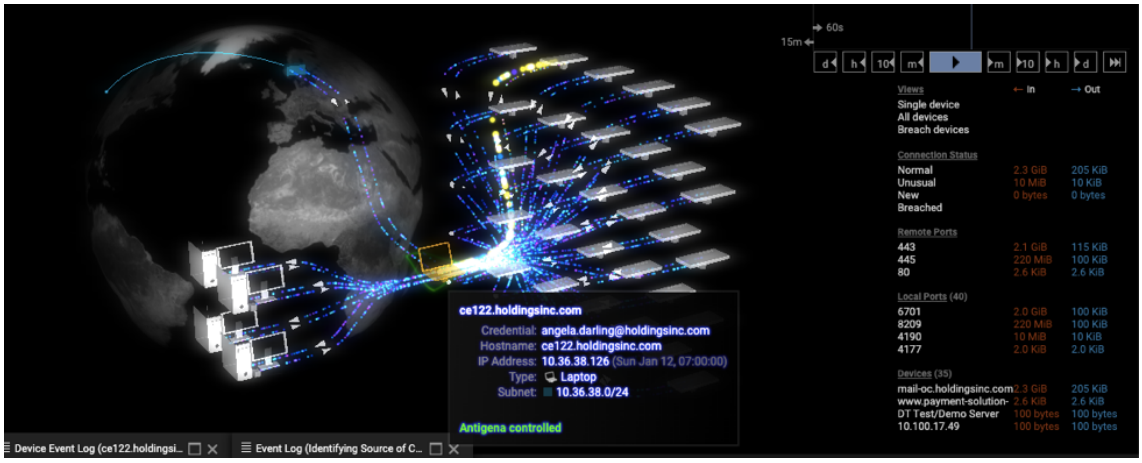

Figure 6.6: Darktrace time machine view

6.3 Negative findings

Some negative aspects were identified during the testing period. These were mainly related to false-positive alarms caused by differences between what the technology modelled and the actual behaviour of the entity in real life. Furthermore, initially, the daily effort required to analyze the reported events is not negligible. Finally, the presence of devices used by multiple users leads the appliance to generate some false alarms, as in the case of multipurpose servers.

6.3.1 False positives

Not mentioning the first two weeks of deployment, which were full of false positives since the learning process started from zero previous knowledge, some of the daily alerts reported by the Enterprise Immune System were false positives. It is not a severe drawback since the alerts fired every day are usually less than ten. Therefore, they can be quickly investigated by the network admin in a few minutes. Nonetheless, this is still a negative aspect, and it is worth to mention.

6.3.2 Daily commitment

Following the just mentioned topic, the platform administrator should make a little effort every day to keep the breach section clean, investigate the alerts and decide whether they are false positives, real attack sources or compliance infringements. This process takes less and less time proportionally to the platform usage. Therefore it can be considered a drawback only for the first two or three months. Then the administrator will need only five to ten minutes per day to evaluate the events.

6.3.3 Shared device

When multiple users share a device, it is needed a human action to instruct the appliance about this particular behaviour. Therefore, the administrator has to divide the subnet further and set the proper tracking method so that the appliance can correct the model which characterizes the device. The system will correlate the device actions with the active credentials and creates a pattern of life which comprehends these two elements. Giving this fine-grained contextual information to technology will increase the precision and effectiveness of the alerts.

6.3.4 Multi-purpose server

It happened that a device was used to perform multiple types of connection identified as anomalous. Although it was reported in the positive findings that the platform could be configured to allow this behaviour, this process needs human intervention, and it is a drawback. The platform applies some tags automatically, based on the analyzed entity behaviour, and the administrator can remove and add more tags following his knowledge about the network and the device itself. The problem here was that certain manually applied tags were overwritten from time to time by the platform Machine Learning engine. Therefore, to circumvent this tag replacement process, a higher priority tag was applied to the server so that the Enterprise Immune System was no more able to remove it, guaranteeing the correct model to be applied for that device, avoiding false-positive alerts.

Chapter 7

Analysis of results

In this chapter are reported the facts that happened during the weeks when the Darktrace technology activities were monitored. There were some relevant facts identified and investigated to understand the threat source. Threats and compliance issues were solved to strengthener the company security posture. Some of them are reported in the following sections.

7.1 TeamViewer automatic login

One of the most dangerous threat happened during a Sunday of April. At about two a.m., Darktrace identified anomalous connections starting from a server inside the network and going to many external and rare sites. These connections were a clear symptom of Command and Control strategy, accompanied by crypto-mining activities. Antigena was not active during that period, and the platform administrator noticed the alerts only on Sunday morning. The breach was then investigated thoroughly to identify the weak point and understand how to fix the problem. It was discovered, using the Time Machine functionality of Darktrace along with the rich of context alerts, that the incriminated server, managed by a third party software house, was making TeamViewer connections from two weekends during night hours. Thanks to this information, the software house was interrogated, and

the outcome was that TeamViewer was used during working hours to accomplish routine business tasks, but the "automatic access" option was checked to facilitate the operations. Therefore, after malware was able to land inside the remote endpoint, it connected to the server using a TeamViewer session. From that point, the malware downloaded a copy of himself from inside the managed server and started the rare connections and the Monero mining activities. Fortunately, the server was isolated from the rest of the network. Therefore, it was enough for the use-case company to close the TeamViewer session, created a firewall policy to block any connection starting from that machine during night hours and restore an old backup. Darktrace was able to correlate the events and understand that the TeamViewer program was used correctly during the week, but the weekend usage was anomalous. Moreover, it was able to indicate the Monero mining actions and the Command and Control connections, helping during the investigation phase. The incident was discovered by Darktrace 10 hours before the software house noticed that was breached.

7.2 Rogue router

Another major event occurred a few weeks after Darktrace was implemented. A router unknown to the network administrator has been identified within the corporate perimeter. By studying the connections, the device communicating with that router was recognised. It was discovered that the unauthorised router belonged to an employee and was being used to bypass security controls that prevented the use of cloud storage, which was not compliant with company policies. After clarifying the situation, it turned out that the employee had no intention of stealing corporate data, but wanted to be able to work more easily using that cloud storage.

7.3 Deleted Office 365 user

Before analysing this event, it is good to know that when an Office 365 user is deleted, it will be sent to the trash for 31 days. Besides, the user's O365 license will not be revoked before this month has passed. Darktrace identified a user who was performing large uploads in his OneDrive space. That action was judged odd compared to his "pattern of life". After a quick investigation, the user's credentials matched those of an employee laid off a week earlier. Immediately, the events identified by Darktrace were analysed and compared with the OneDrive logs. After this analysis process, all files uploaded to the user space in the offending period were discovered. At that point, all uploaded data was discarded, and the user's Office 365 license was revoked.

7.4 Torrent downloads

This case was not as dangerous as those explained above, but it was still a compliance problem and could have led the company to receive a considerable fine. During the first few days of Darktrace's distribution, the platform identified extended torrent connections from midnight to four in the morning almost every day. By examining this violation, the offending device was identified, along with the authenticated user credentials. It was traced back to a desktop computer assigned to a trainee to carry out his work activities. This person downloaded a torrent client and scheduled to download numerous movies and TV shows every night. Thanks to Darktrace, this behaviour was immediately identified, and the intern was advised not to do it again. The torrent client was uninstalled and never used again.

7.5 Extensive cloud storage usage

During the first few weeks, it was discovered that many users were using Dropbox, Google Drive and Box to save corporate files and sensitive customer data, in contrast to corporate policies that only use OneDrive as cloud storage. Thanks to these reports made by Darktrace, it was possible to identify the users included in these violations and the data that were uploaded. After careful analysis, the affected employees were informed, and the compliance violation never occurred from week seven onwards.

7.6 Possible WannaCry ransomware

During the fourth monitored week, a Windows XP machine began communicating using the SMBv1 protocol and contacting external sites deemed rare. Although all these actions were failing, the behaviour was associated with the ransomware called WannaCry, which was probably using the EternalBlue exploit to contaminate other devices within the network. Fortunately, this legacy machine was kept online only to maintain a rarely used service and, therefore, there was a firewall policy which allowed specific communications from and to that machine, reducing the exploitable surface. Since Darktrace identified the device and the possible cause of these strange connections and actions, it was decided to restore the machine and update the operating system, installing the compatible version of the service and, consequently, it was possible to mitigate the ransomware risk. To be sure, a thorough threat analysis was conducted using the Enterprise Immune System, but no compromise indicators were identified. Therefore the corporate network was rated as safe.

Chapter 8

Conclusions

In this thesis, the internal threats that a medium-sized company must take into consideration to secure its network have been analysed. It was seen how to correctly position a next-generation IDS, which is capable of detecting attacks never seen before. During the testing period, some relevant ongoing attacks were discovered, and the pattern of these threats was not identifiable by static rules or signatures-matching systems. Instead, the applied technology was able to detect tricky tactics and procedures, thus helping to maintain a secure business environment. Furthermore, both during the monitored period and after, the UEBA system helped to improve the company safety posture, modifying its network and adapting it to the business needs. Moreover, several weaknesses identified before the technology deployment have been solved. In particular, it has been stated the solution manageability and the deep network visibility, all from a single platform.

Given an IDS that adapts to the network peculiarities and any medium-sized company of any commercial sector, the use case shown in this thesis can be generalised for the following reasons:

- It is not necessary to have previous knowledge of the network, devices or users.
- The system evolves as the company changes, initially reporting warnings of suspicious behaviour but learning new habits after a few days.

- If new elements are added to the network, technology will observe them and unify them with their fellow men, considerably speeding up the learning phase.
- It is possible to tune the IDS to adapt it more effectively to business needs, and the optimisation process is most likely easy to perform. Besides, the solution requires near-zero maintenance once in production.
- A small number of people is needed to monitor the alerts produced by this technology, depending on their aptitude for IT security and their technical level.
- During the thesis development, the same technology was implemented in three other companies of different sizes, from four hundred users up to about eight thousand users, monitoring several IP addresses that varied from one thousand to fifteen thousand for the largest reality. The trend and results obtained are extremely similar to those reported in this document.

There are only a few shortcomings that can jeopardise the good results of behaviour-based IDSs. One is a lack of data, given by incorrect network configurations or bad policies. If employees are barely controlled, their traffic to the company's network resources or intellectual property will be lost. Besides, another rare but concrete example is a business that changes continuously, not allowing the system to determine the "daily activities" of company entities.

Future developments of next-generation IDS could include deep integration with OT systems, granularly controlling the connections and commands sent to and from them and understanding the meaning of those parameters. The goal should be to identify any off-scale values, thus ensuring that OT devices are not tampered with or compromised by apparently legitimate but dangerous commands. Besides, when designing these systems, it should be kept in mind that is the technology which should adapt to the business requirements, and not the other way round.

References

[1] Jeffrey Hunker and Christian W. Probst. Insiders and Insider Threats - An Overview of Definitions and Mitigation Techniques. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, 2(1):4–27, 2011.

[2] Dawn M. Cappelli, Andrew P. Moore, and Randall F. Trzeciak. The CERT Guide to Insider Threats: How to Prevent, Detect, and Respond to Information Technology Crimes. In *The CERT Guide to Insider Threats*, 2012.

[3] Guerrino Mazzarolo and Anca D. Jurcut. Insider threats in Cyber Security: The enemy within the gates. *arXiv preprint arXiv:1911.09575*, 2019.

[4] Symantec Corporation. Implementing an Effective Insider Threat Program. Technical report, Symantec Corporation, May 2017.

[5] Jon Fielding. The people problem: how cyber security's weakest link can become a formidable asset. *Computer Fraud & Security*, 2020(1):6–9, 2020.

[6] Verizon. 2019 Data Breach Investigations Report. Technical report, Verizon, 2019.

[7] Cybersecurity Insiders. Insider threat - 2018 report. Technical report, CA Technologies, 2018.

[8] Accenture security and Ponemon Institute. Ninth Annual Cost of Cybercrime Study. Technical report, Accenture, 2019.

[9] Przemyslaw Kazienko and Piotr Dorosz. Intrusion Detection Systems (IDS) Part I-(network intrusions; attack symptoms; IDS tasks; and IDS architecture). *Retrieved April*, 20(2009), 2003.

[10] Ajay Yadav. Network Design: Firewall, IDS/IPS. *Infosec*, Apr 2018.

[11] Mofti Rafie Abdel-Ghani Ahmed et al. *Enhancing Hybrid Intrusion Detection and Prevention System for Flooding Attacks Using Decision Tree*. PhD thesis, Sudan University of Science and Technology, 2019.

[12] Chani Jindal, Mukti Chowkwale, Rohan Shethia, and Sohail Ahmed Shaikh. A survey on intrusion detection systems for android smartphones. *International Journal of Computer Science and Network*, 3(6):12–17, 2014.

[13] R Vinayakumar, Mamoun Alazab, KP Soman, Prabaharan Poornachandran, Ameer Al-Nemrat, and Sitalakshmi Venkatraman. Deep learning approach for intelligent intrusion detection system. *IEEE Access*, 7:41525–41550, 2019.

[14] G.V. Nadiammai and M. Hemalatha. Effective approach toward Intrusion Detection System using data mining techniques. *Egyptian Informatics Journal*, Dec 2013.

[15] Darktrace. Machine Learning in the Age of Cyber AI - A Review of Machine Learning Approaches for Cyber Security and Darktrace's Underlying Technology. Technical report, Darktrace, 2019.

[16] Darktrace. Darktrace Cyber AI - An Immune System for Email. Technical report, Darktrace, 2019.

[17] Eurostat. Your key to European statistics. *Small and medium-sized enterprises (SMEs) - Eurostat*, 2004.