POLITECNICO DI TORINO

Master's Degree in Computer Engineering

Master's Degree Thesis

Automation in the Cybersecurity Incident Handling Process

Academic Supervisor:
Prof. Gatteschi Valentina
Prof. Diana Berbecaru
Company Tutor:
Dr. Mennuni Massimo
Dr. Rinaldi Stefano

Candidate:
Davide Ioan Manco

Academic Year 2021-2022

Acknowledgements

There are a few people I must mention now that my academic journey has come to an end.

First of all, those who facilitated my path in my Master's thesis project, Prof. Gatteschi Valentina who gave me the opportunity to carry out my project in Reale Mutua Assicurazioni, and Prof. Berbecaru Diana Gratiela who supported me in my work.

Thanks also to all the professors I met during these years, they were essential in my academic training.

Special thanks to Mennuni Massimo and Rinaldi Stefano, for welcoming me into Reale Mutua, allowing me to work on this project and, above all, making me feel like one of the family. Your advice has been invaluable, I will treasure it.

I thank my family, without them none of this would have happened, thank you for being there for me throughout my journey, I will always be grateful.

Thank you to Arianna, a fundamental person, thank you for being understanding and supportive during these years, thank you for always believing in me.

And thanks to all my friends with whom I have shared so much, for the laughs and for the many special moments, you are and will be an indispensable part of my life.

Summary

Cybersecurity is becoming more and more a key area nowadays, cybercrimes are constantly growing and management and response to a cyber attack is an indispensable component.
Incident Handling is a function performed to effectively manage and respond to incidents and protect organizational resources such as sensitive information and human resources from a variety of cyber-attacks. This includes not only reacting to incidents, but also triggering alerts to prevent potential risks and threats. As corporate security covers vast areas, more and more companies rely on external threat intelligence services with which they collaborate.
Cyber Threat Intelligence is definitely a relevant part of the attack prevention process, which can be used to help identify threat actors by providing details about cyber events, including their tools and procedures, and also information about the general risks associated with cyber threats that can be used to guide a high-level organizational strategy.
Threat intelligence companies report threats and compromise events reported on the network, sending them to partner companies. The information sent may represent malicious IPs, malware hashes or fraudulent domains, this data represents the Compromise Indicators (IoCs).
The thesis presents and deepens the relationship between incident handling and threat intelligence, highlighting how the large amount of IoCs sent by threat intelligence companies to corporations, you need to be filtered by an algorithmic component thus avoiding the analysis process data congestion and become slow.
A detailed analysis of the state of the art is followed by the proposal of the algorithmic model devised which therefore aims to divide the information received among those who can represent the real threats to a company and those that are just false positives, thus improving the time needed to classify this information, reducing the resources employed for this activity, and avoiding that lawful activities are blocked by a misinterpretation of the data affecting the company's reputation.
The model is able to analyze a considerable amount of data, clean them from outliers, classify them according to their descriptive tags, interact with external security providers, collect partial results and produce the final reports.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Dealing with security is not just about keeping away threats that act against business systems, but also about protecting people, business processes and information throughout their life cycle. The two things are often confused, for this reason it should be emphasized that protecting systems (and not information) and therefore "cyber security" means taking care only of the "container", running the risk of forgetting the content.
While a large part of cyber security activities can be outsourced (just as it is possible to entrust the same infrastructures that these protect to third parties), dealing with information security should be a central theme of company management: as such, vision and protection strategy that must be implemented internally.
After a presentation that will illustrate the differences between information security and computer system security, I will present the thesis work that combines these two topics, as it aims to improve corporate security and its perimeter to reduce the risk of sensitive data loss or information leakage due to an IT incident.
The two main areas of research are explained in the following chapters, incident handling and threat intelligence. Two related topics, where the implementation choices of one condition the other. This is followed by an in-depth look at the software developed in Chapter 4 of the thesis work.

1.1 Information Security

For a company therefore it is important to define a strategic plan on the management of the informative patrimony, in order to assure the maintenance of an overall architecture of the systems integrated and sure. The implementation of the strategic plan is usually guaranteed of the responsibilities of the internal organs to the company called SGSI(Sistema di Gestione della Sicurezza delle Informazioni), through processes, resources and procedural and organizational controls necessary to maintain an adequate level of data security.
The Standard ISO/IEC 27001:2013 is the international standard for Information Security Management Systems (ISMS). Closely related to ISO/IEC 27002: 2013, this

standard can help organizations meet all information regulatory compliance objectives as well as enable them to lay the groundwork for emerging new regulations.
It is therefore necessary to define the data properties defined by the ISO/IEC standard:

- ***Confidentiality***: ownership be available to unauthorised individuals, entities and processes.
- ***Integrity***: property of an asset and, therefore, of an information to be protected with regard to accuracy and completeness.
- ***Availability***: property to be accessible and usable at the request of an authorized entity.
- ***Non − Repudiation***: non-repudiation is the assurance that someone cannot deny the validity of something. Non-repudiation is a legal concept that is widely used in information security and refers to a service, which provides proof of the origin of data and the integrity of the data. This means that non-repudiation makes it very difficult to deny the provenance and authenticity of the message.
- ***Authenticity***: property that an entity is what it is claims to be.

A proper level of security in order to guarantee this properties is defined in standard by defining a set of standard methodologies and international guide lines.

- The classification of the information, so as to address security measures consistent with the level of criticality of the information.
- The definition of suitable behaviour of internal or external resources involved in business processes
- The definition of permissions necessary to avoid unauthorised reading, modification or deletion of information
- Establishing rules for the identification and management of information security events/incidents

The General Data Protection Regulation (GDPR) is a regulation in EU law on data protection and privacy in the European Union (EU) and the European Economic Area (EEA). The GDPR's primary aim is to enhance individuals' control and rights over their personal data and to simplify the regulatory environment for international business.
Superseding the Data Protection Directive 95/46/EC, the regulation contains provisions and requirements related to the processing of personal data of individuals (formally called data subjects in the GDPR) who are located in the EEA(european economic area), and applies to any enterprise—regardless of its location and the data subjects' citizenship or residence—that is processing the personal information of individuals inside the EEA.
The GDPR was adopted on 14 April 2016 and became enforceable beginning 25 May

2018. As the GDPR is a regulation, not a directive, it is directly binding and applicable, and provides flexibility for certain aspects of the regulation to be adjusted by individual member states. In a company, in the event of an accident that could result in a personal data breach, it is required to comply with the rules in force concerning the obligations to notify the competent authorities and to communicate to interested parties.
The need for notification to authorities and/or communication to data subjects is determined by the data protection office(DPO) which assesses the level of risk to the rights of data subjects and proceeds accordingly. The DPO also has the tasks of draws up the specific register for events involving personal data,report to the authorities when the event is deemed to be relevant in accordance with the criteria laid down in the legislation and develop awareness programs and training on personal data issues.

1.2 Security of computer systems

The first step to take in addressing a complex issue such as that of Cyber Security in the company is to define the context from which this term is born, among the many definitions that proliferate in the literature, some are cited more frequently and among these is the one reported within the ISO/IEC 27000: 2014 standard which describes it as:

"That practice that allows an entity (an organization, a citizen, a nation, etc.) to protect its physical assets and the confidentiality, integrity and availability of its information, from threats arriving from cyberspace"[11]

Looking closely at the definition, it can be broken down into four parts: In the first part the subjects involved are described, in the second part the elements to be protected are exposed, in the third part it is explained from what they are protected through the practice of Cyber Security. Finally, in the last part, we talk about CyberSpace.
The cyberspace as defined by ISO / IEC 27000: 2014 is a complex ecosystem resulting from the interaction of people, software and services on the internet by means of technologies, devices and networks connected to it. The word "ecosystem" therefore describes that cyberspace includes within it a structured environment, which does not only involve machines and systems, but also human beings.
Bruce Schneier, explains very well the reason for the word ecosystem, defining security as a process, and not a product, since it is a science that involves many components, of which the intrinsic safety of the product cannot be trusted. IT security in the company finds application in different sectors, as it is important to be safe and powerful on multiple levels of security, implementing the concept of Security in Depth.

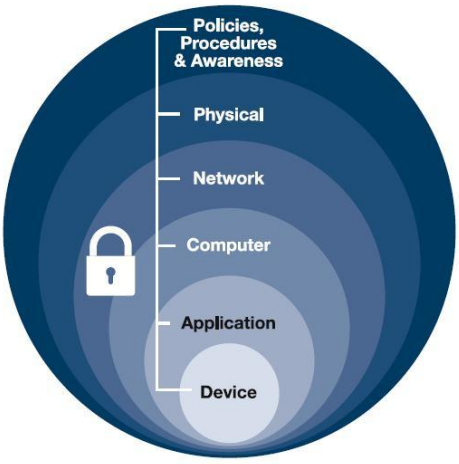

Table 1.1. Defence in Depth [6]

Security in Depth (SID), also known as the Castle Approach, is a computer security concept often presented during a computer scientist's coursework. It is defined as a strategy in which multiple layers of security controls are placed within a technology system. Compared to the castle approach, one layer of security is not enough: one line of defense will not keep all attackers away. The Defense in Depth approach can be executed and structured in multiple different ways; the tradeoff between speed and security can be an important variable in the implementation of this approach. [29]

About CTI topic, threat intelligence and threat hunting are defense proactive activities too. Threat hunting can be explained and illustrated as a technique developed to search for threats within the network architecture before attackers and malicious users can attack. Threat hunting, unlike what is commonly thought, is different from penetration testing and vulnerability assessment.
The aforementioned techniques are activities that simulate an attack from the outside, thus assuming no knowledge of the internal perimeter on the part of the attacker, while those conducting threat hunting assume that a threat is already present within the infrastructure and exploit lateral movements, indicators of compromise, and other information obtained to prove the existence of anomalous behaviour.
We can divide cybersecurity into two macro areas of power, one that is reactive and the other proactive, although it would be more appropriate to talk about services, reactive and proactive. Proactive services useful in a company involve different actors and systems.
It is important train and educate internal personnel about which are the best practise about security in order to achieve some awareness about theme.
It is equally relevant adopting preventive defense systems by design a proper architecture designed to block the upstream attack before it causes damage to internal systems.
Another task that is classified as a proactive security service is the 'Assessment and Vulnerability management', this phase is divided into penetration testing, code review and vulnerability assessment, with the aim of reducing system vulnerabilities and minimizing the risks of potential attacks. Otherwise reactive services usually consist in post incident reports from constituency or other events related to threats or attacks such as compromised hosts, malware, vulnerabilities or other type of similar incidents.
Examples of such reactive services are alerts and warnings, incident handling (detection and response), post incident analysis and forensic analysis.

1.3 Enisa Threat Report

Before deepening the discussion on IoCs, it is necessary to introduce which are the major vulnerabilities exploited as an attack vector. To do this, they will rely on ENISA's annual report.
ENISA is the European cybersecurity agency, the agency works with organizations and businesses to build trust in the digital economy, promote the resilience of EU infrastructures and ensure the digital security of EU citizens. ENISA works primarily for the benefit of public organizations, such as EU institutions and agencies.
Every year Enisa publishes a report describing the main vulnerabilities exploited in an attack, the report refers to the previous year, the year 2022 has not yet been published.
The document, called 'ENISA Threat Landscape' describes as main threats:[7]

- **1# *Malware***: Malware is a common type of computer attack in the form of malicious software. Malware families are diverse and often include cryptominers, viruses, ransomware, worms, and spyware. The main goals of malicious software are certainly information or identity theft, espionage, and disruption of services. ENISA collected a lot of data and showed how there was a relevant 50% increase in malware aimed at stealing personal data or stalkerware and 400,000 detections of spyware and adware preinstalled on mobile devices.

- **2# *Web based attacks***: Web based attacks ranked 2^{nd} in the threat landscape, maintaining the same position from 2018. Some common attack vectors are: [28]

  - DRIVE-BY DOWNLOADS that downloads malicious contents to the victim's device. In such attacks, the end user must consult the legitimate website that has been compromised.
  - FORMJACKING. In this technique, malicious actors inject malicious code into legitimate website's payment forms. This attack mainly captures bank information and other personally identifiable information.
  - MALICIOUS URL. This is defined as a link created with the intention of distributing malware or facilitating a scam.

- **3# *Phishing***: Phishing is a fraudulent attempt to steal user data like login credentials, credit card details or even money using social engineering techniques. This type of attack is usually launched through e-mail messages, appearing to be sent from a trust source, with the intention of convincing the user to open a malicious attachment or click a fraudulent URL.

- **4# *Web application attacks***: Web applications attacks ranked 4^{th} in the threat landscape descending from the 3rd position in 2018. 84% of the vulnerabilities observed in web applications were security errors, and the most common attack takes advantage of SQL injection.

- **5# *Spam***: Spam techniques cover the two phases of the cyber kill chain: Weaponization and Delivery. There are multiple channels used as spam in order to steal information or compromise the victim's machine, some channels described in the report are SMS channels, fake fork or email spam. [23]

Chapter 2

Incident Handling and Response

Before presenting all the stages and actors involved in incident management, it is advisable to define what an IT incident is, and how it should be treated correctly. An IT incident is any disruption in an organization's IT services that affects anything from a single user or the entire company. A security incident breaks the security properties presented by ISO / IEC 27002. Any event or set of these that imply a violation of ICT security policies that is a source of damage to the ICT assets or to the information assets of the organization and for which it is necessary to apply contrast and / or containment measures by of the structures in charge.
In other words, a security incident represents a particular type of alarm whose events imply a clear finding of damage, already suffered at the time of their detection and reporting. An example of a security incident could be unauthorized access or unauthorized disclosure of confidential information.
Incident management is a function performed to effectively manage and respond to incidents and protect organizational assets as sensitive information and human resources from a variety of cyber-attacks.
It is a set of well-defined processes to identify, analyze, prioritize, and resolve security event to recover the system to normal service operations and avoid further recurrence of the incident. This includes not only reacting to incidents, but also triggering alerts to prevent potential risks and threats.
To handle IT incidents correctly, the company must be aware about the risks present in their system and manage them. The security administrator actor must identify the software that is open for an attack before anyone takes advantage of the vulnerabilities. Incident Management includes vulnerability analysis, security awareness training to improve service quality, proactively resolve issues and reduce the impact of incidents. Organizing training sessions to spread awareness among employees is an important part of incident management that aids users in better recognizing suspicious events or incidents and being able to report them to the appropriate authority. Incident management covers all components of information security, such as risk analysis and management; vulnerability identification and mitigation; and threat assessment. Incident response is a component of incident handling that is part of the services provided as part of incident management. It combines the incident handling process

of triage, reporting, detection, analysis, containment, eradication, and forensics investigation. These processes, when performed accurately, can help the organization in fighting against the incidents and preventing or reducing relevant losses.

2.1 Cost of cyber events

The costs generated by cyber events can be differentiated among first and third-party losses. First-party are cost related to direct consequences of IT incident, which are incurred by the company. For instance, in the case of a data breach, this would include the cost of forensic investigation in order to determine the cause or the cost of notifying involved consumers. Losses incurred by third parties are related to costs incurred due to private litigation (e.g. court decisions) or fines or fees imposed by government agencies.[21]
Talking about first party losses, these are split it in tangible and intangible losses, where tangible cost refers to the organization's direct expenditure due to an incident, which could be the cost of replacing the damaged infrastructure or the amount spent for implementing the incident handling process, including the salaries of team members as well as the cost of hardware and software tools.
Intangible Cost refers to the expenditures that the organization cannot calculate directly or value accurately. Intangible cost is difficult to be identified and quantified. It is the loss of not quantifiable assets such as damage to corporate reputation or loss of business credibility and trust.

2.2 Preparing for Incident Response

Preparation is necessary for any well-executed endeavor, and incident response is for sure included. Incident preparation is needed to develop response capabilities, and to facilitate the response process too. As Defense in Depth policy suggests, the prevention adopted in the pre-incident phase involves different areas or layers,such as workstations, network, policies and procedures. In addition, this phase is important to ensure that all aspects of the incident response plan (training, execution, hardware and software resources, etc.) are approved in advance. Response plan have to be well documented, thoroughly explaining everyone's roles and responsibilities. [14]

2.2.1 Preparing Host

About workstations, good choices could be recording cryptographic checksum of critical files, increasing secure audit logging, setting up a proper host's defenses and backing up critical data.

- Checksum of critical files: the integrity of files and data must be verified. The response team checks the integrity of system information and the last time the system information was accessed. To check these attributes, it is needed to match the current system state against a "known-good" system state. Any changes to the system state should generate an alert.

- Increasing audit logging: almost every operating system and many applications provide significant logging capabilities. Exploiting logs after any suspected incident might improve the research about the cause of the incident. Unfortunately, the default logging of the software is not optimized. To achieve a better logging operation, customization is often necessary.

- Set the host defense: if each host is completely secure, many security incidents would be avoided. Pre-incident step has the goal to set up a proper defense system on the workstation. Actions taken to secure host reduces the exposure to security.
  Some guidelines suggested by Kevin Mandia in Incident Response Computer Forensics say to be aware about all operating system and application software is the most recent using the latest release disable unnecessary services. If employees do not use an application or network service, it should not be running. Unnecessary services introduce unnecessary risk.[14]

- Backup of critical data: regular, complete system backups can be a useful reference during incident response. Backups might help to discover what was deleted and what was added, which checksum alone cannot show. In addition, some backups save time/date, which can be useful for checking the times files and directories were last accessed, modified, or created.

2.2.2 Preparing A Network

Literature suggests many network-based security measures that can be taken to improve incident response capability. As a matter of fact, network-based logging is absolutely essential, because there are many cases in which network monitors hope to catch evidence. In fact, network administrators play a critical role during incident response.

- Installing firewalls and Intrusion Detection Systems: when the network's components, intrusion detection systems (IDS), and firewalls are configured optimally, the intrusions are more complex and so less probable. The way in which configure these systems depends on the response posture of the organization.
  Company may decide to deny certain attacks and not log, or permit attacks and log in detail. The configuration of these devices is not simple. The focus is that, rather than configuring network devices to simply protect the network, the company have to configure them to log activities too.

- Configure properly access control lists: the router is typically configured with access control lists (ACLs) that allow certain types of traffic while prohibiting potentially dangerous traffic.

- Use Network Time Protocol: a good rule in order to get better interaction with logs is synchronizing all machines at the same time using the Network Time Protocol (NTP). Block all external access to this port, and have all of the machines on your network synchronized. All logs record the same time for an event this

allows a more valuable event analysis of the time on the router compared to the firewall, compared to the victim machine and the network monitor and other sites.

2.3 Incident Handling and Response (IH&R)

Incident handling and response (IH&R) is a process made of precise steps when reacting to a security incident or cyberattack. It is a set of procedures, actions, and measures taken against an unexpected event occurrence.
Nowadays, organizations need to be constantly vigilant against breaches, and having an incident response plan is key.
A proactive approach that defines strategies before, during and after a potential breach allows the organization to make the most informed decisions to defend it.
For this reason, cyber threat intelligence and incident handling are two activities that are closely related to each other, also thanks to the work carried out by IoCs.
The integration indicators of compromise in defense systems is an additional tool that allows for better implementation in terms of identification and response to the attack.
We will see how in the containment and eradication phase, for example, an accurate analysis of the IoCs, can make the attack under consideration, the actors and the systems involved better-known thanks to the description of the IoC.
A qualitative description of the attack, therefore, can allow a company to understand if it can be a potential victim, which is why the integration of indicators of compromise is an important proactive defense that is disadvantageous to give up. IH&R identifies the incident when it occurred and estimates its impact and its cause.
The rigid steps of preparation, detection, containment, eradication, and recovery allow the company to recover from the impact of an incident quickly and efficiently.
IH&R process involves defining user policies, developing communication protocols among actors involved in processes, building incident response teams, auditing the organizational assets, planning incident response procedures, incident reporting, prioritization, and managing response.
There are several benefits get by the adoption of a well-defined IH&R process, many are strictly related to each of them, identify relevant data and resources that require some level of protection could reduce for sure the impact of potential attacks.
In addition, a proper response increases efficiency and productivity throughout the organization by reducing the time needed for reacting to cyber attacks.

2.4 Detection and Reporting

This step requires observing the environment and analyzing events from various sources such as log files, error messages, and other resources, such as intrusion detection systems and firewalls, that may produce evidence of possible relevant incidents.
The detection phase has to be done through gathering events from various sources and observation of system/application deviation from normal operation. The report of the event can come from the SOC or from the internal system of the company.

2.5 Classification and Prioritization

The activity in this phase must determine the event relevance and classification and severity of the incident in terms of impact and priority. To classify the event is a crucial step, even if you do not have all the information about it, a preassigned is made, in order to activate the components ready to manage that type of severity.
The team has to consult each available source able to say if the event is legitimate or if it is a simple error. In order to verify any data modification, the team audit different sources such as logs because information related to the incident might be available in several places like IDPS, firewall or router logs.
Event correlation is a technique that involves logs that describes a set of event that occurs in a fixed amount of time, because a single log might not prove evidence, but analyzing the together could describe operations performed at a different level by an attacker. Another analysis adopted by the IH&R team is called 'Profiling' which is the process that involves systems and network and consist of the detection of the anomalous behavior of an expected activity.
Monitoring network traffic or checksum of critical files are some examples of profiling.
The classification of an incident depends on the potential targets and the severity of its impact. The purpose of incident classification is to gather all the required information to determine its category and the time required for resolving.
IH&R team evaluates incident details and correlates with indicators and classifies incidents based on their severity, affected resources, and attack methodology.
Once the incident is classified, the prioritization step can begin, it plays a crucial role in the incident handling and response process, because determines the sequential process of responding to a security incident. Organizations adopt a common set of terminology to categorize the incident in order to unambiguously communicate security incidents and events across different departments.
Incident categorization enables the team to prioritize the incident and focus on the incidents that require more effort.
Low-level incidents require usually few time to resolve but they are not be underestimated because can be a starting point for major security incidents. In case of low-level incidents often the business continuity is not compromised. Middle-level incidents require a higher effort and the incident handler must perform a very precise analysis because often this kind of event can represent a false positive. The time to resolve increases too.
High-level incidents are the most risky for a corporation, which compromise the business continuity and cause the highest loss in terms of resources and data. Incidents can affect many services offered by the company, compromising a relevant number of customers. A denial of service attack or the presence of harmful worms which lead to corruption of data are some examples.

2.6 Notification and Dispatching

After the classification and validation phase, it is key to notify about the incident that occurred. The notification could prevent other important assets from becoming victims of the attack. The IH&R team has the task of reporting the event including a descriptive report of the threat, since the report could leak some sensitive information is needed waiting for approval from the internal department.
Once approved by the competent authority, the IH&R team communicates the relevant matters about the incident with necessary stakeholders too. As already explained, this step requires many actors which cooperate with each other, because the response guidelines have to be common in each department and external entity involved. Internal communication is a key requirement not only in the notification phase but also in every phase.
During the response process, different company departments work simultaneously, so secure communication among these is essential. An organization must be secure that will be correct information flow during the response phase, for this reason some company adopts secure communication channel while other prefer communication out of band where internal data are kept separate from regular communication.

2.7 Containment

Containing a cybersecurity incident is all about limiting the damage and stopping the attacker reducing the potential economic losses caused by the event. So the main goal is to avoid the spread of the attack to similar resources across the company.
As the starting point, it is necessary to define a containment strategy that depends on the area involved in the incident and how is complex the attack's pattern.
The strategy might include an external support, if required. Secondly, response team has to take a decision about shutting down the system and isolating the network by disconnecting it.
Shutting down the systems and disconnecting them from the network is the best option if files and critical data are involved, but an objective of the containment phase is also to keep alive the business continuity or at least make sure to restore the function at the earliest. If the containment response fails, the team has to roll back and reviews the containment strategy until the latest is successful. EC-Council suggests some guidelines to follow in the containment phase and presents the most common techniques used by the response team.
To begin, it's a good choice to disable the compromised system services temporarily especially if the event exploited an unknown vulnerability. If an attacker got access by compromising an account, IH&R has to disable accounts and change the password on the affected systems in order to minimize any loss of data. Some complete backups have to be done periodically, they will be fundamental to restoring the services and for performing further investigation about the event.
The guidelines presented by EC-Council suggest also to choosing a safe location for storing data, and to creating forensic backups to proper media.
Regarding the response team and internal personnel, it is important to inform ever

of the latest details about the security event, in addition, each operation during the containment phase have to be reported in a standard way following the standard procedures and policies defined in the containment strategy. About technician hints, EC-Council argues that honey-pots also play a vital role in enhancing security.
Once the containment step is concluded, a forensic analysis might be required. In this step, IH&R team collects evidence about the incident and in the same time create a chain of custody document. The investigators have the task to perform analysis based on information collected defining the perpetrators of the crime and filling the forensic document.

2.8 Eradication

The eradication step is essential because one of the main goals of the incident response teams should be to eliminate the RCA (Root Cause Analysis) exploited by the malicious actors to attack the company. All the actions that unfolded during this phase should be thoroughly documented. As soon as the incident was contained, it is possible to proceed with the eradication phase. In order to eradicate, IH&R has to perform vulnerability analysis to check whether the network is still vulnerable.
If the vulnerability is no more presented then verify if a similar issue exists in similar systems.
IH&R team has also test their environment as final checking before initiating the recovery process. Update the antivirus software with new malware signatures and patterns and install the latest patches on the system are the first operations adopted by the team. Once the cause was eradicated, it is necessary the rebuild of the affected systems, databases and networks. Generally this phase requires a complete reimaging of a system's hard drive(s) to ensure that any malicious content was removed and to prevent reinfection.
This phase is also the point where defenses should be improved after learning what caused the incident and ensuring that the system cannot be compromised again. Tools for detecting missing security patches helps incident handlers to install the latest patches.

2.9 Recovery

The purpose of this phase is to bring affected systems back into the production environment carefully, to ensure that it will not lead to another incident. It is essential to test, monitor, and validate the systems that are being put back into production to verify that they are not being infected again by malware or compromised by some other means. The time to recover a system generally depends on the extent of the security breach.
IH&R team selects an appropriate recovery plan according to the availability of resources and the results of a cost-benefit analysis. It is also defined the duration period and method of monitoring for abnormal behaviors, which ensures that the recovered system does not have any traces of the incident.

The purpose of the recovery plan is to prepare the organization to survive in the event of an incident and continue its normal business operations.

2.10 Post-Incident Activities

After the event is solved, the IH&R team have to perform some activities to improve the response to future attacks, and to be better prepared to handle and respond to future incidents.

2.10.1 Documentation and Reporting

Incident Response team while handling and responding to an incident. The documentation describes the security breach in detail and illustrates the response method applied enriching it with how had handled the event and the reason for the decisions taken. The documentation must be very concise and clear, understandable by everyone. Documents have to be filled in a standard way, enhancing the accuracy. Report writing tools help incident handlers to generate efficient reports on the detected incident during the incident handling and response phase.

2.10.2 Lesson Learned

Reviewing and revising policies is a key phase avoiding future potential incidents. Simply, it is fundamental to update policies based on lesson learned about the latest attack. First of all, the reviewing phase should be done in terms of evaluating the time and cost of the response process leading to improving the impact of the next security incident. Procedures, tasks, settings and configuration should be updated with the feedback of the users, a precise document reports what fails during the response, and which systems need to be updated or replaced because obsolete.

2.10.3 Incident Disclosure

As the last step, the corporation informs stakeholders and users that all is solved by incident disclosure ticket. An incident disclosure ticket describes incident details to various entities, for this reason not all of the incidents can be disclosed. It is important information filtering and keeping safe sensitive information like financial data or user accounts, attaching improper information can represent an intrinsic vulnerability. The IH&R team usually requests approval from management to disclose the incident information to stakeholders, once get approval, the relation team will disclose the detail of the incident to the involved agents.

Chapter 3

Cyber Threat Intelligence

In recent years, Cyber Threat Intelligence (CTI) has become a highly discussed topic in Information Security (IS) but the little literature on this does not clarify well its definition and therefore companies tend to use their own definition to distinguish their product and this can lead to confusion. According to Gartner, Threat Intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard. [16]
A short definition is presented in the paper of Brown, Gommers & Serrano as follows: "Intelligence is about reducing uncertainty in a situation of conflict or of business objectives (also known as "business risk")". While Cloppert offers different definitions of CTI that based on operations, analysis and domain. Hence, he defined the Cyber Threat Intelligence Operations as actions taken in cyberspace to compromise and defend information and capabilities available in that domain.[2]
Based on Cloppert, threat intelligence is not only focused on a nation that is bound by some technique to influence national policy, but It is more on technical aspects such as tools and techniques. [4]
The following table describes some actors involved in cyber threat intelligence:

Role	Description
SOC analyst	This is a person who must have undoubted cybersecurity skills, and must be up-to-date on major SIEM systems and data aggregators, and have data science skills because there is an increasing growth of compromise alerts. It is important to configure data aggregation and analysis tools to interface them with systems this requires relevant programming and scripting skills.
SIEM	The SIEM is a solution in which the powers of the SIM and the tasks of the SEM are combined, the former being a data management system that collects data automatically and orchestrates the logs. While, the SEM is a product that monitors and manages events occurring in the network in real time, often having a console for reporting and responding automatically to compromising events. The SIEM therefore has the task of analysing the collected logs, highlighting abnormal behaviour, producing more detailed reports and enabling optimisation of incident management time.
IPS	The IPS is the network security tool that continuously monitors a network for suspicious activity and is tasked with preventing it by using alerts to system admins or blocking it.
Threat Hunter	A threat hunter deals with analysing the behaviour, targets and methodologies that professional hackers might adopt to cause economic damage to a company. Thus, a threat hunter needs to collect and analyse significant amounts of data, to enable him to estimate forecasts of potential attacks.
CTO	A company's CTO is primarily involved in strategic threat intelligence which focuses on understanding high level trends, and once understanding it can derive some business decision, for instance change the threat intelligence provider.

There are many different definitions to explain this term. There are three overarching, but not categorical - classes of cyber threat intelligence:

- ***Tactical***: technical intelligence (including Indicators of Compromise such as IP addresses, file names, or hashes) which can be used to assist in the identification of threat actors.
  Tactical intelligence focuses on the time window of the immediate future, the main task being to detect IoCs useful for the preventive defence of a system. It is very important that the information produced is reliable, because this could affect the response time to the attack, but especially the detection time.
  These IoCs can be URLs, domains, IP addresses or file hashes and have the important advantage of being machine-readable. Some providers make tactical intelligence their main business, although the services offered can be found from open source sources that also provide real-time feeds, it is estimated that the main difference is in the reliability of the information.
  In fact, one often finds oneself with old or recycled information, i.e. sent back

again at different intervals. Using a tactical threat intelligence service certainly requires a feed management system, because a large amount of data is downloaded, and it is practically impossible to analyse it by a human component.

- ***Operational***: details of the motivation or capabilities of threat actors, including their tools, techniques and procedures.
  Operational threat intelligence has as its first objective to profile malicious actors in order to gain a better understanding, by analysing patterns, of hackers. It's like studying one's adversaries, which is what Operation CTI is all about. The questions it aims to answer are not only 'who' is involved in the attack, but more importantly why they chose to act as they did and 'how' they operated to carry out the attack.
  These responses constitute the CTI, i.e. the tactics, techniques and procedures adopted in the attack. In this category of CTI, the human component is crucial, as computers cannot estimate and calculate an operational report of the attack. Human analysis is always necessary for data conversion.
  Unfortunately, however, despite the significant expenditure of resources in this area, the window of validity of this information is not that long, because cyber criminals tend to change their CTIs very quickly.

- ***Strategic***: intelligence about the overarching risks associated with cyber threats which can be used to drive high-level organizational strategy.
  Strategic threat intelligence has the task of understanding why a certain attacker decided to move in that particular way, why the companies involved were chosen from that particular sector. These answers certainly help a company to then make certain business choices in response to what happened or is estimated to happen.
  Hackers never act alone, numerous events show how external entities are often present to finance the work of cybercriminal groups. For instance, geopolitical ties have always been a relevant factor.
  This type of threat intelligence illustrates how the most relevant cyber events potentially occurred due to friction between states or were desired and financed by international currents.
  Strategic CTI helps company managers make informed decisions and helps them understand the potential risks to which their companies may be exposed. With these suggestions, a decision can be made, for example, to invest in improving the security architecture involved in the protection of certain high-risk data, or to improve certain malware detection tools because there is a high prevalence of ransomware for companies operating in a certain sector at the time.
  Strategic intelligence is the most complex form of threat intelligence because it not only requires accurate data analysis, but also requires constant updates on the global geopolitical situation, and the information generated must be treated with high discretion.

3.1 CTI Maturity Model

Most generally, a maturity model is a tool for assessing an organization's effectiveness at achieving a particular goal. They enable organizations to identify where their practices are weak or not taken seriously and where their practices are truly embedded. In the context of cyber security, maturity models can help to distinguish between organizations in which security is baked in and those in which it is merely bolted on. One of the main reasons that maturity models are used is that organization-wide improvements can take time; in cyber security, a maturity model gives an organization's leadership a way to measure the progress made in embedding security into its day-to-day and strategic operations. The Cyber Threat Intelligence Maturity Model comprises five levels of maturity from Level 1 (Initial) to Level 5 (Optimising). [27]

- ***Level 1 – Initial***
  Level 1 describes an organization that performs little or no Threat Hunting, and instead has a reactive stance, relying on alerts generated by SIEM tools and other defensive infrastructure. Threat Hunting occurs rarely, if at all, and is ad-hoc and basic; it is performed by existing staff e.g. SOC analysts, and on their own initiative.

- ***Level 2 – Managed***
  At Level 2, steps have been taken to start implementing a proactive Threat Hunting capability. Existing staff are occasionally led on hunts by a dedicated and experienced Threat Hunting lead, with the focus on targeting IOCs.

- ***Level 3 – Defined***
  Level 3 is the minimum level required for a company to operate a competent Threat Hunting capability and start realizing benefits. A team of dedicated hunters, led by the Threat Hunting lead, follow a formal Threat Hunting process and hunt on a frequent schedule, with the focus on targeting IOCs, using techniques such as statistical analysis. Normal systems behavior is adequately understood for key systems to allow identification of abnormal activity. Hypothesis and hunt information is recorded in a central knowledge repository, and workflow management tools are used to track workloads and progression. While identified IOCs are provided to the CTI and Protective Monitoring functions for the development of the subsequent SIEM detection rules.

- ***Level 4 – QuantitativelyManaged***
  At Level 4, the Threat Hunting capability is well established, and utilizes quantitative metrics to improve performance and show benefit. The Threat Hunting team is supplemented by SOC analysts. Hunting is very frequent, and targets IOCs at the top of the POP (Piramid of pain) (i.e. adversary TTPs). Mission critical systems are identified, contributing toward the hunters understanding of the organizational context and therefore starting to develop their proper defense.

- ***Level 5 – Optimising***
  At this level, the Threat Hunting team is fully integrated into the wider SOC

with action plans created to mitigate any underperformance. The hunts are conducted continuously, with successful analyses and IOC discoveries shared throughout the community. [20]

3.2 Cyber Kill Chain

Cyber kill chain is a model for incident response teams, digital forensic investigators and malware analysts to work in a chained manner. Inherently understanding, Cyber kill chain is modeling and analyzing offensive actions of a cyberattacker.[30]
To analyze complex and structured attacks, the cyber kill chain provides a framework to break down a complicated attack into mutually nonexclusive stages. The layered approach permits the analysts to split into smaller and easier problems and at the same time it also helps the defenders to examine each phase by developing defenses and mitigation for each of the phases. Cyber kill chain defines the flow of a cyber attack in this 6-layer model:

1. ***Reconnaissance***
   Reconnaissance consists in gathering information about the potential victim which can be a person or a corporation entity. Reconnaissance can be broken down into target identification, selection and profiling. All available sources are involved in reconnaissance step such as personal websites, blogs or mailing lists. It is possible to distinguish two kind of reconnaissance: Passive and Active Reconnaissance.
   In passive reconnaissance the information gathering about target is made without letting him/her know about it, while active reconnaissance requires much deeper profiling of the target which might trigger an alert to the target.

2. ***Weaponization***
   It is the one phase that the victim doesn't see happen, but can detect. Weaponize phase of the cyber kill chain aims with designing a penetration plan based on the information gathered from the reconnaissance step. It is the technique used to obfuscate shellcode, the way an executable is packed into a trojan document, etc. Technically it is binding software/application exploits with a remote access tool (RAT). Remote Access Tool is a piece of software which executes on the target's system and gives remote, hidden and undetected access to the attacker. Detection of this is not always possible, nor is it always predictable, but when it can be done it is a highly effective technique.

3. ***Delivery***
   This is the critical phase where the payload is delivered to its target. In most the cyber-attacks it is common to have some kind of user interaction like downloading and executing malicious files differently from some attacks which are performed without user interaction by exploiting network devices or services.
   This phase is very sensitive for an attacker, because the delivery leaves tracks on the target system in fact the attacks are usully done exploiting anonymous services or compromised email accounts. Attacker use often more one channel

to deliver a malicious payload, to achieve the guarantee that at least one is successful, also a failed delivery can represent an information useful about the target.

Delivery Mechanism	Explanation
Email Attachments	The email content is composed to entice the user to download the attachment.
Phishing Attacks	Sensitive information like usernames, passwords, credit card details etc. are stolen by an attacker that acts as a trust fake entity.
Drive by Download	Target download malicious content from the internet.
USB/Removal Media	Infected files are kept in removable media which silently infects other systems.
DNS Cache Poisoning	Vulnerabilities in DNS are exploited, DNS traffic follows a well-crafted path toward a fake server controlled by an attacker.

. MITRE ATT&CK - Delivery mechanisms [18].

4. ***Compromise / Exploitation***
   Once all the previous steps have been performed, the exploit can be triggered, then it will install / execute the payload on the victim's computer. It is clear that exploit is the most critical part of the chain technically. The payload will connect to its Command and Control counterpart to inform about successful execution and wait for further commands to execute.

5. ***Command and Control***
   An important part of remote cyber attacks is the command and control (CC) system. The CC system is used to provide remote instructions to compromised machines. It also acts as a place where all data can be exfiltrated, although this stage is not always present.
   Over the years CC has evolved into many strategies, in fact there are many attack patterns. There are mainly three types of CC communication structures, namely the traditional centralized structure, the new decentralized peer-to-peer architecture.

   - Centralized Structure: in this model there is usually the use of a single malicious server. It is the simplest structure, the number of bots instantiated in the CC attack depends on the hardware capabilities of the server. If the server is knocked down, the whole architecture fails.

- Decentralized Structure: malware authors started using peer-to-peer P2P architecture for command and control. The benefits of using this architecture are scalability, each node is responsible only for a subset of the total botnet and fault tolerance.

6. ***Exfiltration***
   The exfiltration phase is conceptually very simple: this is when the data is taken. Critical data stolen from the victim's computer is typically packed and file-encrypted before being sent to the attacker's collection point such as the server at the top of the botnet.

3.3 Indicators of Compromise (IoCs)

These indicators are often analyzed and used at different levels, because they often encode useful information not only at the network level. With varying degrees of confidence depending on the source they will help identify a network compromise or simply report a compromising event related to the same business as the company that is using the IoC. These IoCs are primarily used by network security deputies to protect their network infrastructure. Examples of IoCs can include:

- Unusual DNS lookups.
- Suspicious files, applications, and processes.
- IP addresses and domains belonging to botnets or malware CC servers.
- A significant number of accesses to one file,
- Suspicious activity on administrator or privileged user accounts.
- An unexpected software update.
- Data transfer over rarely used ports.
- Behavior on a website that is atypical for a human being.
- An attack signature or a file hash of a known piece of malware.

Indicators of compromise (IoC) are fundamental to cyber threat intelligence (CTI), since they improve and speed the detection of malicious activities in technological infrastructures. They show and describe the use of technological capabilities involved in an attack, such as the tools adopted, both tactics, techniques, and procedures (TTPs) developed by hackers. Information related to TTP is not easy to find among the various providers deputed to infosharing this type of information.
These providers mainly focus on sharing basic indicators (hash, ip and domain), which are useful since they provide immediate results once they are uploaded to network security components, such as EDR or firewalls, but they present a non-negligible problem: their lifecycle, because they mostly describe dynamic information with a short time validity. [25] In his research, Cloppert classifies three classification

categories for IoCs, based on their levels of structure and complexity and also in relation to the granularity of the data they represent: [5]

- ***Atomic***: Atomic indicators describe that indivisible information, which cannot be broken down into smaller parts and retain their meanings in the context of an intrusion. Examples of atomic indicators include IP addresses and domain names.

- ***Computed***: Computed indicators are those which are derived from data involved in an incident. Examples of computed indicators include hash values and regular expressions.

- ***Behavioral***: Behavioral indicators are collections of computed and atomic indicators, often subject to qualification by quantity and possibly combination logic. An example of a complex behavioral indicator could be repeated social engineering.

Behavioral IoCs are produced by operational threat intelligence, while atomic and computed ones are associated with tactical threat intelligence.
Tactical intelligence has a shorter life cycle than operational intelligence, and can also be bypassed more easily. Although they represent a less useful information than the behavioral ones, the atomic and computed ones are considered by many organizations to be the data with the most added value of a threat intelligence company.
The main reason for this evaluation is that the indicators representing tactical intelligence are usually expressed in machine-readable formats, which is why they can be easily loaded into the security components, providing immediate reports and results once an infringement has occurred.
IoCs are mainly produced by manual investigations of companies or derived from company reports and then once analyzed they are then shared on a large scale, so that a range of individuals and organizations can adopt them in their defense systems.
Some advantages related to the use of IoCs: [13]

- IoCs don't need large resources to use: IoCs are highly scalable and easy to deploy, making them a truly valuable resource for smaller entities. IoCs are also cheap to use.

- IoCs reduce the effort needed to defend against an attack: the match is one-to-many, as it is enough to block a data described by an ioc, an IP for example, to protect several users within a company.
  Moreover the shareability and reproducibility of iocs is another important advantage, because it allows a defender to use in a preventive way that information to organize in the best way its process of response to the attack.

- IoCs can be attributed to a specific threat actor: this means an organization can prioritize or accept trade-offs against a subset of malicious actors, binding IoCs to threat actors allows the company to focus its defenses against particular risks.

- IoCs permits for discovery of historic attacks: a network defender can use recently acquired IoCs with historic data, like DNS queries or email attachment hashes, to hunt for signs of past compromise.

SANS Institute illustrates how IoCs play a relevant role in the Incident Handling Process:

Simplified Objective	SANS Incident Handling Step	Explanation
-	Step 1: Preparation	This step takes place prior to an incident and does not take IOC into account.
Detecting possible infection	Step 2: Identification	IOC is used to describe the malware in object.
Preventing further infection	Step 3: Containment Step 4: Eradication	IOC is used to document the changes made to the infected host's file system and registry configurations; network and the host IPS could be configured for the purpose of containment and eradication adding information described by IOC.
-	Step 5: Correction	In this phase IOCs are not involved.
Profiling infection	Step 6: Lesson learned	IOCs that could describe the profile of the attacks in order to determine if it is a targeted attack.

. Using IOC in IH&R - SANS Institute [10].

3.4 MITRE ATT&CK and TTP

MITRE ATT&CK is a popular globally searchable repository of adversary tactics and techniques. It can be used as a powerful tool to illustrate and categorize adversaries' moves based on their attack dynamics.
ATT&CK is a well-constructed list of clear behaviors by hackers, which have been categorized by different techniques and tactics and put into matrices that summarize them. This matrix is a comprehensive representation of the behaviors that cybercriminals adopt during an attack in order to compromise networks and systems.
MITRE has divided ATT&CK into three different matrices covering different categories and systems: Enterprise, Mobile, and PRE-ATT&CK. Each matrix describes techniques and tactics related to the domain of that matrix.
The Enterprise matrix lists tactics and techniques that can be used on common operating systems such as macOS, Linux, and Windows. While Mobile describes tactics and techniques applicable to mobile devices.
Finally, PRE-ATT&CK contains tactics and techniques that illustrate activities that hackers implement before the actual attack begins; these are often information gathering activities and preliminary study of the victim.
The taxonomy used by ATT&CK is very useful for companies to adopt: [18]

- ***Threat Research***: ATT&CK illustrates the main defensive systems and informs about the main vulnerabilities, this is an important advantage for Threat Hunters who can use this information to detect yet unknown cyber attacks.
- ***Integration of tools***: Different systems and technologies can coexist by adopting and standardizing their operation on the ATT&CK taxonomy.
- ***Detection and Investigation***: The SOC(Security Operations Center) along with the incident response team can consult techniques and tactics listed in ATTCK that have been recently discovered. This information will allow them to diagnose the internal network by looking for which systems may be involved in these vulnerabilities, and by improving the systems one can use this information for accurate preventive defense.

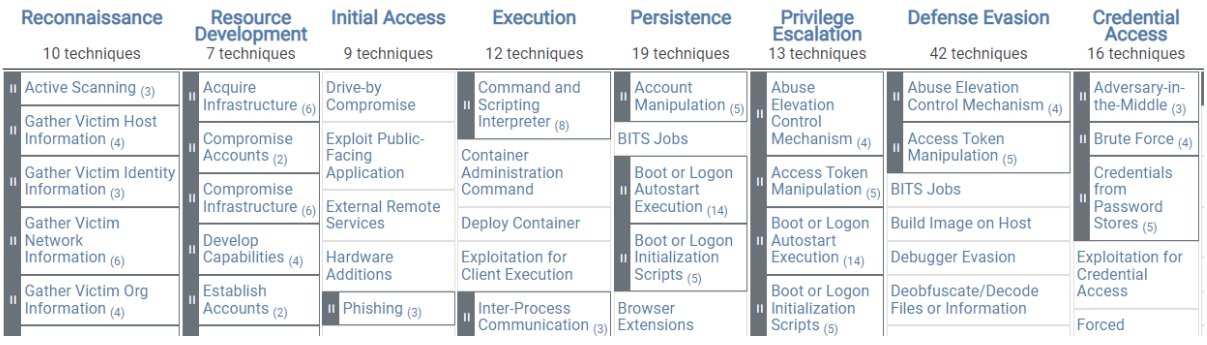

Figure 3.1. Section of MITRE's Enterprise matrix [18]

Before moving on to the next section, it is appropriate to present what is meant by the use of the words technical tactics and procedures.
This information is very useful for a very accurate attack analysis, because it allows one to study the opponent's actions and remedy them before another attacker can replicate the attack again.

- ***Tactics***: Tactics describe why our attacker performs certain operations; each of these tactics is described by MITRE and each is identified by an ID. Some of these may be for example Privilege Escalation or Credential Access.
  So the tactics present why a cybercriminal acts in a certain way, and the general strategy adopted to accomplish his goal.
  The description of this presents the tactic at a high level without specifying methods and tools.

- ***Techniques***: Techniques describe how an opponent operates in order to carry out his attack. MITRE also adopts identifiers for the techniques to distinguish them, and subcategories have been defined for each category, e.g., for the account discovery category, subcategories have been defined for each type of account, whether local, domain, or email.
  So the techniques describe in more detail the behavior of the attacker, they are not as high-level as the tactics.

- ***Procedures***: The procedure describes step by step how the attack is carried out, often referring to a specific technique and tactic.
  The procedures then codify the actor of the attack with which technique he has chosen to adopt, how he intends to apply it and what aids he has chosen to use to carry out the attack.
  All this information can be valuable for studying a potential future incident from the example already recorded and thus being able to predict the flow followed by the attacker.
  This type of information is the most difficult to retrieve and, in fact, it is often commercially disclosed and not distributed on open source.

3.5 Pyramid of Pain

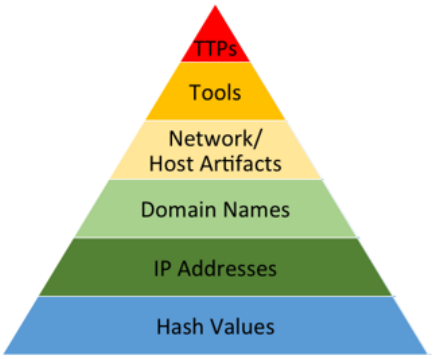

Table 3.1. Pyramid of Pain IOC types [3]

The categories of IoCs can be summarized in a "Pyramid of Pain" that can be used for attack prevention, detection, and mitigation. The levels of the Pyramid of Pain start from hashes to TTPs, and the pain ranges from code recompilation to the creation of a new attack strategy. At the lowest level are the hashes of malicious files. These are the easiest for a defender to collect and can be distributed to firewalls, for example, to block malicious downloads.

IPs and domains, on the other hand, rank at the next level. The latter are referred to as blockable, with a good percentage of false positives often causing malfunctions and blocking legitimate traffic; attackers can bypass blocking by changing IP ranges, assigning it from a new provider, and changing their code if the IP address is hard-coded. Domain names are more granular than IP addresses and are more difficult for an adversary to change.
Network and host artifacts, such as modified timestamps of files left on the endpoint or a pattern that becomes on the network, are even more difficult to modify because they relate specifically to the attack in progress and may not be under the attacker's direct control. Tools and TTPs are located in the top two levels of the pyramid; these levels describe the attack methods adopted by a malicious actor, i.e., how it executes the attack.
An example would be the implementation of a malicious code to perform a reconnaissance of the victim's network,for the purpose of lateral movement between workstations then finding a valuable endpoint, and then downloading a ransomware payload. Tools, on the other hand, describe the software pattern used to conduct the attack, while TTPs address the broader attack strategy used. The information on TTPs and tools is the most difficult to find and requires significant diagnostic effort on the part of the defender, but it is critical to the attacker.

3.6 Sources of IoCs

The indicators of Compromise can be received from different sources, the latter being classified into two categories: external agencies or internal sources.

- External Agencies: The external agencies may be commercial or industry sources or free IoC sources such as the MISP. Examples of commercial IoC sources include antivirus or antimalware vendors, and all of these have a huge repository or collection of IoCs which are used and constantly updated. Some free IoC sources available include the Malware Information Sharing Platform or MISP, or the AlienVault OTX, which is a nice resource across many different areas. Another possible alternative is IBM X-Force Echange that is a cloud-based, collaborative threat intelligence platform that helps security analysts focus on the most important threats and help speed up time to action. TIP combines human intelligence with a global security feed, providing a unique view into potential actors and threats.

- Internal Sources: There are several methods that you can use to collect logs and events that are analyzed to detect IoCs. And these can come from commercially available systems, some free systems. Such as internal logs and event viewers. Some of the main ones include unusual outgoing network traffic and geographic anomalies connection or for instance could be an account user who is logging from a foreign location.

A further classification is made regarding external sources that are divided into TIP and MISP.

- MISP: It is a platform used primarily for the exchange of information, enrichment and correlation of external data. This platform uses STIX/TAXII only for the exchange of information stored in proprietary JSON format and is normally used as data storage and IoCs correlation. Currently there is a strong use by public administrations for the exchange of IoCs and links with other European agencies.

- TIP: It is a platform mainly used to share information, enrichment, correlation and analysis /investigation of an organization's internal and external data. This platform can be native to STIX/TAXII standard and all internal elements and its components use the standard of Threat Intelligence; it is used as data storage and as an analysis device /survey. Currently there is a strong use by private companies, for the conversion and exchange of IoCs, which in turn are used to make cyber information "actionable", exploiting the link to SIEM and SOC.

Moreover, regarding the technical functionalities, Ing. Mattia Siciliano reports the following differences. [22]

Feeds Ingestion	Threat Intelligence Platform	MISP Threat Information Sharing Platform
Ingestion of Public Feed	✓	✓
Ingestion of Commercial/Private Feed	✓	✗
Ingestion of Feed in structured formats	✓	✓
Ingestion of Feed in unstructured formats (e.g. pdf)	✓	✗
Users	**Threat Intelligence Platform**	**MISP Threat Information Sharing Platform**
Threat Intelligence Teams	✓	✓
SOC / CERT Teams	✓	✓
Fraud & Risk analyst	✓	✓
Management and Executive Teams	✓	✗
Scalability	**Threat Intelligence Platform**	**MISP Threat Information Sharing Platform**
Production/Consuming scalability	✓	✗
Architecture scalability	✓	✓

. TIP & MISP differences [22].

Other advantages of TIP over MISP:

- Lifetime for information sharing: the time-to-live information of the IOCs threat indicators depends on the intelligence process you want to adopt. Nevertheless, intelligence needs to be analyzed by consumers in order to establish priorities for action. Currently, time-to-live information that is not provided by most feeds, can instead be set to the TIP that has the function of directing the intervention to the SOC or CERT.

- Internationalization: TIP is a technology created for the management of cyber threats and information sharing of Cyber Threat in a standard STIX format, whereas MISP technology is a European technology only and is mainly used only in Europe by government agencies and for Information Sharing purposes.

3.7 InfoSharing & Standards

A fundamental theme in cyber threat intelligence is surely the theme related to info sharing. This is the cornerstone of every threat intelligence company. Info sharing is basically an ecosystem where there is a real-time sharing of actionable information (that is, information analyzed, contextualized, timely, accurate, relevant and predictive) of Cyber Threat, able to increase the defenses of an organization in order to prevent, identify and mitigate the Cyber Threat before there can be a real impact on the organization itself.
It is therefore important to understand what to share, with whom, how and why to do it. About what to share depends a lot on the company that takes care of it, surely a public part will be interested in sharing a potential threat to a critical entity, while in a private context, info sharing is contextualized in the merits of the company's own or similar field of interest. UK Government argues that collective defense is the main reason for sharing information.
Regular cyber threat information sharing significantly assists organizations mutually to pre-empt, prevent, detect, and respond to serious cyber incidents and threats, while improving the preparedness and resilience of the wider ecosystem. Awareness of the various threats that may affect other organizations allows better use of internal resources and capabilities. Info sharing aims to improve threat awareness by learning from the perspectives of other similar organisations and aids building contacts in case of a cybersecurity event requiring collaboration with other organizations.
National Institute of Standards and Technology (NIST), the European Union Agency for Network and Information Security according to UK Governement suggests some principles about how shared information, based on experiences of security professionals and practitioners engaged in the routine sharing of cyber threat information.
First of all in order to information to be useful, companies need to have good inventory management and documentation. The cybersecurity team needs to establish best practices for how to communicate information regularly on threats and vulnerabilities to management/business. These best practices must also clarify which channel is most effective (e.g., email, briefing, report). Senior security team plays a crucial role in setting the tone and championing information sharing. [9]

3.7.1 STIX/TAXII Standard

Structured Threat Information Expression (STIXTM) is a structured language used to describe cyber threat information developed by OASIS (Organization for the Advancement of Structured Information Standards). STIX gives a structured, common framework for disclosing cyber threat intelligence, improving intelligence accuracy, interoperability and automated processing efficiency.
TAXII is a protocol that operates at the application layer to deliver cyber threat information in a simple and scalable way. TAXII leverages HTTPS as a protocol for exchanging network threat intelligence and it is auditable by a set of APIs. Once the information is normalized in TAXII format it is delivered to TAXII server using the TAXII transport mechanism.
All clients subscribed to the TAXII server can get the latest threat information from the TAXII server. TAXII can also transmit threat information in other formats, greatly increasing the flexibility of threat intelligence sharing, differently from STIX. TAXII permits multiple companies to securely share threat information in faster way. Specifically, the full realization of TAXII means: [12]

- Security and Privacy: since TAXII defines standard mechanisms to protect the integrity, confidentiality and attribution of information on cyber threats, these features can be included in tools that automatically ensure the correct level of security and privacy protection.
- Speed: cyber threat information sharing is faster. Defined services and message exchanges enable automation for what is now a largely manual undertaking. Defenders can receive data in real time.
- Improved and advanced analysis: STIX/TAXII standardization and automation allow companies to better organize analyst time, the effort previously used to manually produce indicators of compromise, such as cutting and pasting IP addresses from a PDF file can instead focus on analyzing threat data.

Chapter 4

Feeds Classification Based On Scoring System

4.1 Introduction

The project aims to integrate the information received from different Threat Intelligence sources in order to classify them, thus acquiring a greater awareness of the potential usefulness of the feed received.
As described in the previous paragraphs, IoCs represent useful information for the preventive defense of our systems, and also play an important role in incident handling during the various phases. IoCs are often shared using JSON files, which follow different standards starting from the STIX/TAXII presented in Chapter 3.5 to other less common ones, sometimes the shared information is often textual and networked using PDFs.
The amount of information shared between the various threat intelligence companies represents an important amount, which would be difficult to manage by internal personnel. Furthermore, the filter work relating to this information would require a continuous flow. Carrying out my study, I was able to observe how the research on this topic was few in number, but many shared the idea of a model that could quantify and estimate the quality of this information.

4.2 MISP Source

MISP Threat Sharing project consists of multiple initiatives, from software to facilitate threat analysis and sharing to freely usable structured Cyber Threat Information and Taxonomies. Some features of MISP:

- An efficient IoC and indicators database allows to store information about malware samples, incidents, attackers and intelligence.
- Automatic correlation finding relationships between attributes and indicators from malware, attacks campaigns or analysis. The correlation engine includes correlation between attributes and more advanced correlations.

- A flexible data model where complex objects can be expressed and linked together to express threat intelligence, incidents or connected elements.
- Built-in sharing functionality to ease data sharing using a different model of distributions.
- Storing data in a structured format (allowing automated use of the database for various purposes) with an extensive support of cyber security indicators along fraud indicators as in the financial sector.
- Export: generating IDS (Suricata, Snort and Bro are supported by default), OpenIOC,plain text, CSV, MISP XML or JSON output to integrate with other systems (network IDS, host IDS).
- Customizable taxonomy to classify and tag events following your own classification schemes or existing taxonomies. [17]

4.2.1 MISP Taxonomy

Taxonomies that can be used in MISP (2.4) and other information sharing tools and expressed in Machine Tags (Triple Tags). A machine tag is composed of a namespace (MUST), a predicate (MUST) and an (OPTIONAL) value. Machine tags are often called triple tags due to their format. [17]

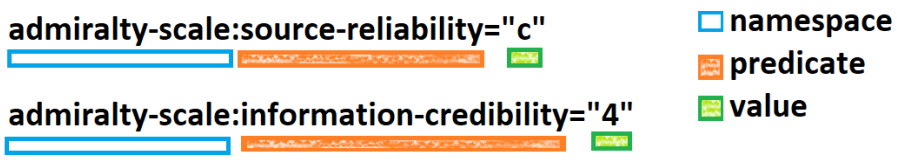

Figure 4.1. Taxonomy Example

4.2.2 Example of IoC

The following is an example of an event received by the MISP provider, we see how each event contains a list of descriptive tags which is the taxonomy, while the list of IoCs is contained in the Attributes section of the .json file.

```
1 "Event": {
2    "publish\_timestamp": "1607324084",
3    "info": "OSINT - Egregor: The New Ransomware Variant
        To Watch",
4    "published": true,
5    "date": "2022-07-8",
6    "analysis": "2",
```

```
    "timestamp": "1657307916",
    "uuid": "0b988513-9535-42f0-9ebc-5d6aec2e1c79",
    "threat\_level\_id": "1",
    "Tag": [ {
        "name": "type:OSINT",
        "colour": "#004646" },
    {
        "name": "osint:lifetime=\"perpetual\"",
        "colour": "#0071c3" },
    {
        "name": "osint:certainty=\"50\"",
        "colour": "#0087e8" },
    {
        "name": "tlp:white",
        "colour": "#ffffff" },
    {
        "name": "misp-galaxy:ransomware=\"Egregor\"",
        "colour": "#0088cc"
    }],
    "Attribute": [{
        "deleted": false,
        "value": "http://49.12.104.241:81/78.bin",
        "disable\_correlation": false,
        "type": "url",
        "comment": "",
        "category": "Network activity",
        "to_ids": true,
        "timestamp": "1606485600",
        "uuid": "7df62701-db13-41e4-987c-dcd58b98b7c5"
    },{
        "deleted": false,
        "value": "http://49.12.104.241/sm.dll",
        "disable_correlation": false,
        "type": "url",
        "comment": "",
        "category": "Network activity",
        "to_ids": true,
        "timestamp": "1606485600",
        "uuid": "6b2c6a04-37bd-4796-a56a-29489fd91efc"
    }
}
```

4.3 Related Work

4.3.1 Sources

In my study we worked with the three main sources of feeds that can be consulted for free as they are open source sources.

- The Computer Incident Response Center Luxembourg (CIRCL) is a government-driven initiative designed to gather, review, report and respond to computer security threats and incidents.
  url = 'https://www.circl.lu/doc/misp/feed-osint/'
- Botvrij.eu provides different sets of open source IOCs that you can use in your security devices to detect possible malicious activity.
  url = 'http://www.botvrij.eu/data/feed-osint/'
- DigitalSide Threat-Intel Repository This repository contains a set of Open Source Cyber Threat Intellegence information, monstly based on malware analysis and compromised URLs, IPs and domains.
  url = 'https://osint.digitalside.it/Threat-Intel/digitalside-misp-feed/'
- Shadow Server - This server provides a lookup mechanism to test an executable file against a list of known software applications. Unlike the previous sources, this is used to build a whitelist useful for scoring.
  url = 'https://bin-test.shadowserver.org/api'

4.3.2 Scoring Model

The system provides for the scoring of each IoC, this mathematical model will then be used to classify the information as a truly valid indicator of compromise (TP), or it will be discarded as false information, which does not represent any type of real threat, and therefore classified as false positive (FP).
There are many reasons why shared information can be of little use. One of the main reasons is that the information can be recycled, which is taken from an old data set and for some reason re-shared, therefore non-original information.
Furthermore, the information could be discarded for another reason. If the company receives an IP address as an IoC, but it is processed by the systems in the week following receipt, it can certainly be said that the data does not have the same relevance, as that same IP address may have been reassigned.
The model takes into account all these variables and tries to summarize them by calculating an integrative score that is conditioned by all these possibilities.
It is absolutely certain that with the same IoC involving an IP address, the algorithm must propose a higher scoring to the most recent IoC, unless this IP address is contained in a WhiteList, but we will discuss this other possibility.
The base_score is canceled if the IOC received is present in our trusted whitelist.
The score is calculated as follows:

$$base\_score = \frac{W_{tag} \times score_{tag} + W_{event} \times score_{event} + W_{vt} \times score_{vt}}{W_{vt} + W_{event} + W_{tag}} \quad (1)$$

- $W_{vt} \in [0,1]$,$W_{event} \in [0,1]$,$W_{tag} \in [0,1]$ represents weights used in base score elaboration, in case IoC refers to recent date W_{vt} is 0. W_{ioc} is used for $score_{ioc}$ which represents the overall reliability data of the event received to which multiple IoCs are linked.

- $score_{tag}$ is a metric based on taxonomy which described the event, MISP uses a rich taxonomy that is implemented in a simple JSON format. Anyone can create their own taxonomy or reuse an existing one. Taxonomy can be freely reused and integrated into other threat intelligence tools.
  Taxonomies are licensed under Creative Commons (public domain) except if the taxonomy author decided to use another license), despite this only a few parameters are useful for scoring the quality of IoC.
  The following table 4.3.2 shows the taxonomy subjected in this study:

namespace and predicate	value	Description
admiralty-scale:source-reliability	[a, b, c, d, e, f]	The Admiralty Scale or Ranking (also called the NATO System) is used to rank the reliability of a source and the credibility of information. Value "a" means information completely reliable while "f" means info unreliable.
admiralty-scale:information-credibility	[1, 2, 3, 4, 5, 6]	Information-credibility describes the number of confirm received for the truthfulness of the IOC. "1" confirmed by other independent sources, while "6" there is no basis exists for evaluating the validity of the information
estimative-language:likelihood-probability	[almost-certain, very-likely, likely, roughly-even-chance, unlikely]	estimative-language should indicate and explain the basis for the uncertainties associated with major analytic judgments, specifically the likelihood of occurrence of an event or development, and the analyst's confidence in the basis for this judgment.
estimative-language:confidence-in-analytic-judgment	[high, moderate, low]	Confidence in a judgment is based on three factors: number of key assumptions required, the credibility and diversity of sourcing in the knowledge base, and the strength of argumentation. Each factor should be assessed independently and then in concert with the other factors to determine the confidence level.
osint:certainty	[30, 50 ,75, 93]	it describes how much confidence there is on the veracity of the data.

According to taxonomy value, in the scoring system is associated a numerical value $\in$ [0,100]

$$score_{tag} = \sum_{i=0}^{n}(tag_i \times weight_i) \times \frac{1}{\sum_{i=0}^{n} weight_i}$$

4.3.3 VirusTotalScore

VirusTotal is a website that allows free analysis of files and/or URLs to find virus or malware inside. It uses more than 90 antivirus software including Kaspersky, Avira, BitDefender, AVG, Malwarebytes, Microsoft and McAfee. VirusTotal allows you to send files with a maximum size of 650 MB. On September 7, 2012, Google announced the purchase of VirusTotal.
If an IOC has recently been shared but is authentic to a given past, virus total can be integrated into our scoring system. VirusTotal produces an output for each antivirus of which it is a partner, classifying the threat as: undetected, malicious or harmless, or timeout if the antivirus is late in finding a response.
This information is also important in evaluating the entire event in its completeness, thus contributing to the scoring of all the attributes present in the received json.

$$score\_vt = \frac{\#antivirus\ detecting\ a\ threat}{\#total\ antivirus} + reputation\_rate$$

VirusTotal has developed its own file reputation system. Whenever you submit a file or URL, you'll see a chart that shows the reputation of the file or URL and ranges from -100 (fully malicious reputation) to 100 (fully harmless reputation). The reputation of each file or URL is built by (among other factors) Virus Total Community user votes, which are recorded by clicking either the malicious or harmless icon below the reputation chart.[8]
The community score is important because unlike the security vendor rate, it provides feedback from a human person who is supposed to have voted from past experience. It can be used in the case of recent IoCs, as it is not certain that security vendors have already updated their blacklists.

4.3.4 Decay Time

As previously described, it is appropriate to give the right weight for each type of IoC, and also it is of fundamental importance to take into account the date on which our IoC was generated and received. The decay time is a variable introduced precisely to try to introduce the time factor in the mathematical model.
The decay time linked to a hash of a file will certainly be different from that of a domain address, as it will also be different from an IP address.
As a starting point, I tried to plot the scoring curve by adopting a linear model as Eq.4:

$$\text{final\_score} = \text{base\_score} - \delta(T_t - T_{t-1}) \tag{4}$$

Symbol	Explanation
T_t	Describes the time our compromise indicator was received.
T_{t-1}	Describes the date our indicator of compromise refers to. $T_{t-1} <= T_t$
δ	Decay Time

Table 4.1. Symbol Table

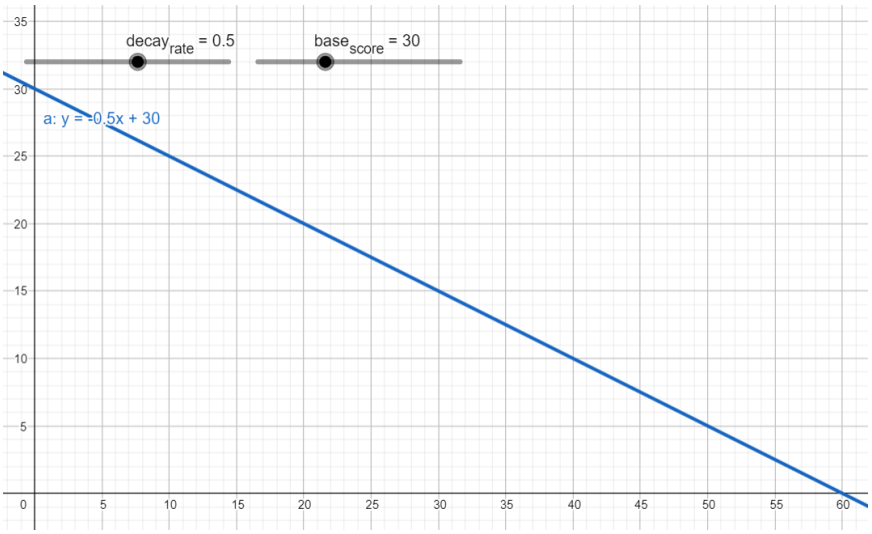

Figure 4.2. Scoring Trend Adopting Linear Interpolation

As shown in 4.2 the decay time does not respect the real trend of the systems. Associating a linear decay time to an IoC related to a hash of a file is not correct, as it is assumed that the integrity of a malicious file remains intact for a long period , and therefore its detected hash remains the same. Decreasing its scoring in linear time would not respect the real dynamics of the systems.
therefore considering what has been said before, I looked for an interpolation that could be faithful to the real characteristics of our systems, adapting the decay rate

to the type of IoC under consideration (hash, IP, domain ..).
The final equation is described by equation 5 [24]:

$$\text{final\_score} = \text{base\_score} \times (1 - (\frac{t_{current} - t_{ioc}}{\tau})^{\frac{1}{\delta}}) \tag{5}$$

Symbol	Explanation
τ	τ represents the expiration time. A company can set it according to the time window it deems best. In the example, a decay time of 5 days has been reported, therefore 120 hours. This information is useful for clearing the scoring of those IoCs that refer to too old date, for which it would be useless to process the information as it is post-dated.
t_{ioc}	The date to which the indicator refers.
$t_{current}$	The date on which the feed is received

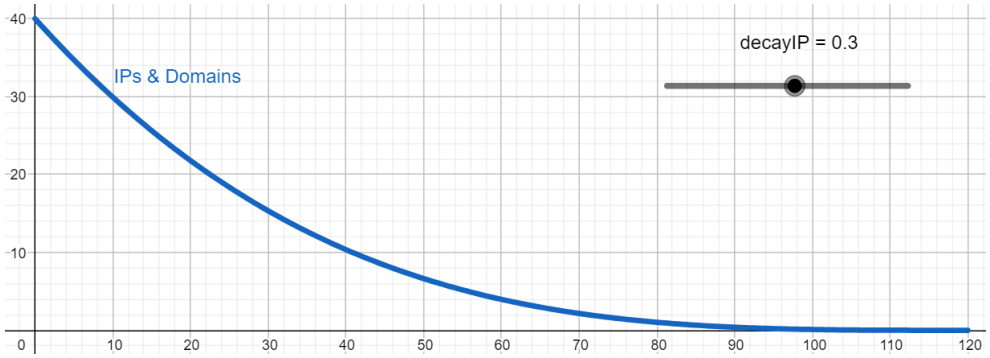

Figure 4.3. Example of Decay Time

As shown in 4.3 is important the usage of the most appropriate δ(decay time) according to the kind of data elaborated. IP and Domain are information similar to each other and we can assume their validity follows the almost same trend. To achieve these properties it is important define:

Decay Time	Application
$\delta \in (0,1)$	Domains, Links and IPs.

4.4 Code

This paragraph describes the steps that follow one another for the processing of the compromise indicators, starting from the input json file, data normalization processes and ranking algorithms for the classification of information will follow. As described in the previous section, there are three types of scoring.
The code is written entirely in Python, which is a "high-level" object-oriented programming language suitable for developing distributed applications, scripting and numerical computing.
The following flowchart describes step by step the operations computed by scoring software.

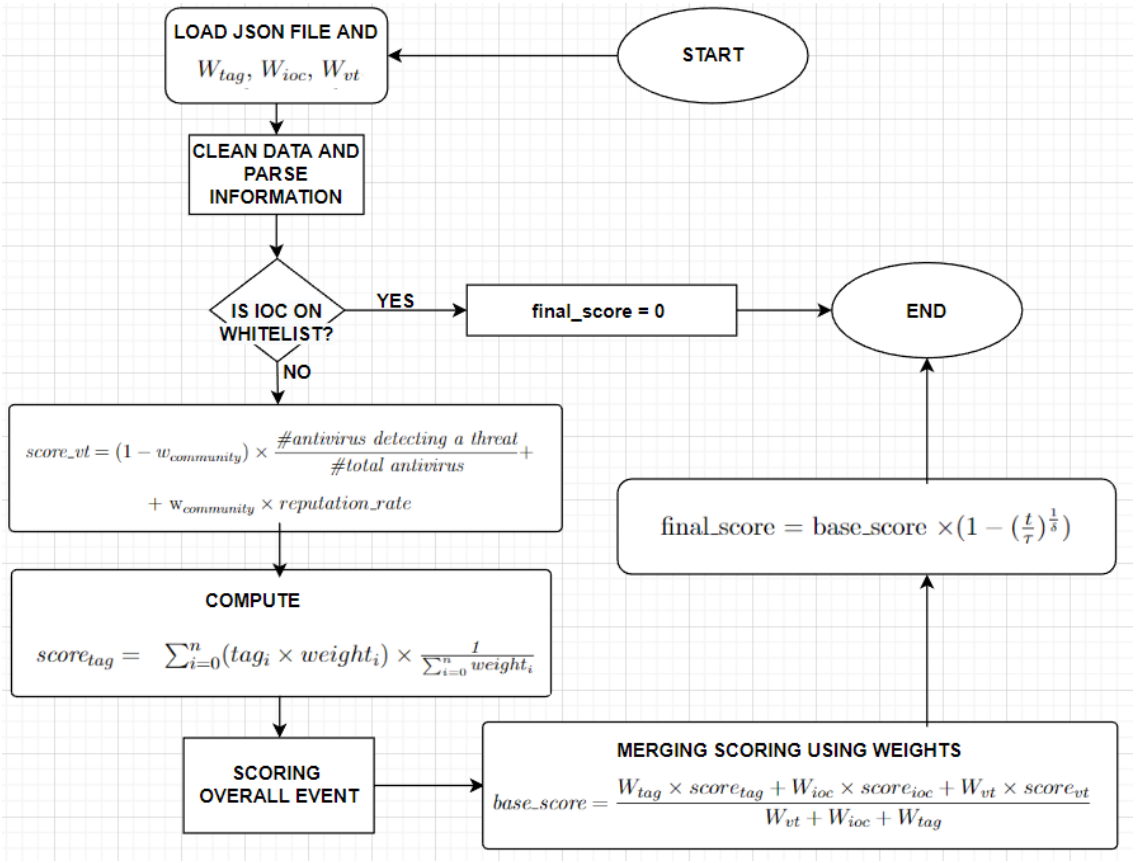

Figure 4.4. FlowChart - Summary and description of execution flow

Here is a comment on the previous flowchart:

- As a starter point, the code asks four input parameters, the json file to be analyzed and the three weights provided by Eq.1 (4.3.2), W_{tag}, W_{event}, W_{vt} useful for the final scoring.

1. The first operation consists of analysing the file to check that there are no errors, or missing data, all ouliers present in the fields useful for producing the

final report are eliminated.

1.1 Then the IoCs are enumerated, each IoC representing a piece of information, and it is checked whether this data is present in the whitelist sources described in sec 4.3.1, if so the analysis on that IoC ends with a final score equal to 0.

2. Now software involves VirusTotal API, from which data is then taken regarding the report produced by the security vendors, the whois call that is attached in the response, and the community score regarding the data provided.
   The community is an important part of the scoring, because if the score provided to the platform is recent with respect to the date of whois creation or last modification, then it is an important vote to measure the reliability of the data.

3. In the third phase an analysis of the entire file is performed, so no longer related to the individual IoC, the score of the tags used to describe the event is calculated, where each tag has a weight and a value, the tags used are described in the previous table.
   Then the number of indicators that have the virus total score, below a certain threshold, are counted, this measure is useful in maintaining a history of the reliability of the source from which we receive this information, this measure could prove useful in future product developments.

4. The final base score of each IoC is then calculated by merging all the previous results. These are the scores that do not, however, take into account the temporal validity of the information. 4.3.2

5. The last stage extends this concept, using the whois call previously made to VirusTotal, calculates the temporal scoring according to equation 5 4.3.4. As a final step, the data is then reported in a final report encoded in a .csv file. This file can eventually be used for uploading into corporate defence systems.

Once the script has been launched, the program will behave as follows in fig 4.5. It will wait in input for the json file to be analyzed and the weights to be used in calculating the base_score.

```
PS C:\Users\david\Desktop\TESI\TesiMaterial\Codice\daTerminale> python mainCmd.py iocList.json 1 1 1
50 IoCs found:
VirusTotal Scoring...
Taxonomy and Overall Event Scoring...
Merging Previous Results...
Results were stored in output.csv in current directory.
```

Figure 4.5. Terminal Snapshot

$score_{tag}$ is calculated by parsing the tags describing the IOC using an additional struct which is shown in fig. 4.4.

```
SCORING_STRUCT = {
    admiralty-scale:source-reliability="a": [0.98, 0.7],
    admiralty-scale:source-reliability="b": [0.84, 0.7],
    admiralty-scale:source-reliability="c": [0.70, 0.7],
    admiralty-scale:source-reliability="d": [0.56, 0.7],
    admiralty-scale:source-reliability="e": [0.42, 0.7],
    admiralty-scale:source-reliability="f": [0.28, 0.7],
    admiralty-scale:source-reliability="g": [0.14, 0.7],
    admiralty-scale:information-credibility="1":[0.96,0.7]
    admiralty-scale:information-credibility="2":[0.80,0.7]
    admiralty-scale:information-credibility="3":[0.64,0.7]
    admiralty-scale:information-credibility="4":[0.48,0.7]
    admiralty-scale:information-credibility="5":[0.32,0.7]
    admiralty-scale:information-credibility="6":[0.10,0.7]
    estimative-language:likelihood-probability=
    "almost-certain": [0.90, 0.5],
    estimative-language:likelihood-probability=
    "very-likely": [0.60, 0.5],
    estimative-language:likelihood-probability=
    "likely": [0.40, 0.5],
    estimative-language:likelihood-probability=
    "roughly-even-chance": [0.30, 0.5],
    estimative-language:likelihood-probability=
    "unlikely": [0.15, 0.5],
    estimative-language:confidence-in-analytic-judgment=
    "high": [0.70, 0.5],
    estimative-language:confidence-in-analytic-judgment=
    "moderate": [0.40, 0.5],
    estimative-language:confidence-in-analytic-judgment=
    "low": [0.10, 0.5],
    osint:certainty:[0.5, 0.5]
}
```

Listing 4.1. Scoring Struct describes which tags are taken in account according to table 4.3.2.

Subsequently, fig. 4.4 and fig. 4.4 show how the $score_{vt}$ is calculated and processed, the snapshot shows the case of an IP address score, in fact the code interfaces with VirusTotal using the ***VirusTotalAPIIPAddresses*** module, the VirusTotal-V3 library provides a different module depending on the type of IoC. Figure 4.4 then shows how the returned data are processed and parsing in json so that the subsequent steps are more convenient in terms of data processing.

```
def call_ip(lista_ip ,x):
    resultjson=""

    vt_api_ip_addresses = VirusTotalAPIIPAddresses(virus_total_api_key)

    try:
        result = vt_api_ip_addresses.get_report(lista_ip)
    except VirusTotalAPIError as err:
        print(err, err.err_code)
    else:
        if vt_api_ip_addresses.get_last_http_error()==vt_api_ip_addresses.HTTP_OK:
            result = json.loads(result)
            resultjson= codifica(result,x)
        else:
            print('HTTP_Error['+ str(vt_api_ip_addresses.get_last_http_error())+']')

        return resultjson
```

Listing 4.2. Interface to VirusTotalAPI.

```
def codifica(result, x):
    malicious=result['data']['attributes']['last_analysis_stats']['malicious']

    sosp=result['data']['attributes']['last_analysis_stats']['suspicious']

    harmless=result['data']['attributes']['last_analysis_stats']['harmless']

    undected=result['data']['attributes']['last_analysis_stats']["undetected"]

    repudiation=result['data']['attributes']['reputation']
    tot= int(mal)+int(undect)+int(sosp)+int(harmless)

    try:
            rate= int(malicious)/tot
    except ZeroDivisionError:
        rate=0
        print("ZeroDivisionError")

    domain_js={'rate':(rate), 'malicious': int(malicious), 'undetected':int(undected),
               'sospicious':int(sosp), 'harmless':int(harmless), 'tot': int(tot),
               'repudiation':int(repudiation)}

    return domain_js
```

Listing 4.3. Parsing result from VirusTotal.

At the end of the analysis, the results are saved in a csv file in the script directory, at the end of this a summary graphical representation of the results produced is

shown.

The figure 4.6 shows on the X axis the window of validity of the IOC, while on the Y axis the relative score represents everything on its curve with the relative decay time of that type of IoC.

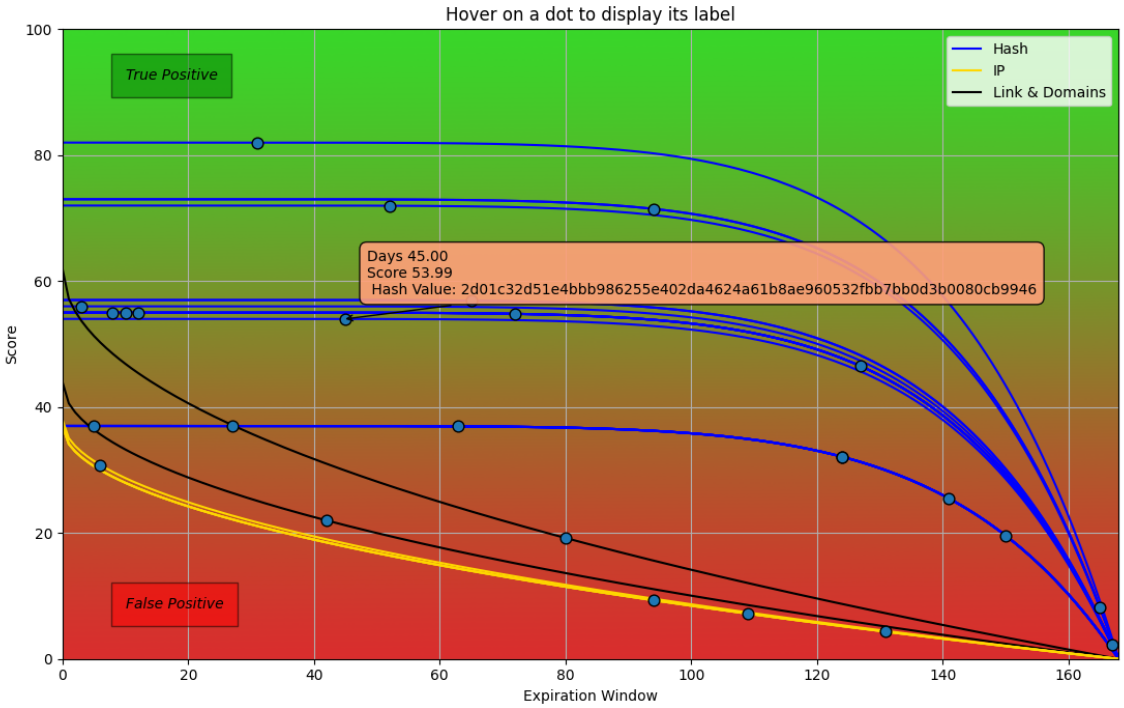

Figure 4.6. IOC scoring graph

4.4.1 Tool

To develop this project some libraries were used that helped me in the development of the code by reducing errors and development time.

- ***VirusTotal API 3*** is the default way to interact with VirusTotal. This new API was designed with ease of use and uniformity in mind and it is inspired in the http:/s/jsonapi.org/ specification. [26]

- ***Matplotlib*** is a comprehensive library for creating static, animated, and interactive visualizations in Python. Matplotlib makes easy things easy and hard things possible. Using this library we can create publication quality plots, make interactive figures that can zoom, and update and customize visual style and layout. [15]

- ***mplcursors*** provides interactive data selection cursors for Matplotlib. It is inspired from mpldatacursor, with a much simplified API.

- ***NumPy*** is the most commonly adopted package for scientific computing in Python. It is a Python library that provides a variety of different objects, a multidimensional array object and various objects derived from it (such as masked arrays and matrices), and a collection of routines for quick operations on object vectors, including operations such as mathematical and logical operations. Used in the project to interpolate data used to draw the final chart [19]

- ***request***: the requests module allows to send HTTP requests using Python. The HTTP request returns a Response Object with all the response data (content, encoding, status, etc). Used in the project to query our resources of whitelist.

4.5 Results

In this section we will analyze the results of the work carried out, also presenting the limits of the model produced and the potential improvements to be adopted.

4.5.1 I Step

In the first phase, a lot of data related to previous events and IoCs were collected, these were used to test the scoring model by analyzing its behaviors and results according to the various descriptive models of IoCs provided. As described in 4.3.1, the data was collected from multiple open source sources, each of these adopted a descriptive model for compiling the IoC report, albeit following the TAXII standard. The project analyzed about 200 thousand indicators and 3 thousand compromise events, most of this dataset was provided by the Computer Incident Response Center Luxembourg (CIRCL), the composition of the dataset is represented as follows 4.2.

IoCs analyzed	CIRCL	Botvrij.eu	DigitalSide	Other Sources
206.170	184.618 (89.5%)	13.259 (6.5%)	7.437 (3.6%)	856 (0.4%)

Table 4.2. DataSet

The choice to use CIRCL as a major data provider is simply linked to the fact that in addition to being the most popular among open sources, It is the one that provides more data in real-time than the others. This first dataset was used to select the best taxonomy to use for the evaluation of the IOC, and also to find the right weights to assign to the selected taxonomy, based on the number of occurrences consistent with the listed IoCs. Furthermore, the most used taxonomy groups for the description of an event were collected, so as to best represent a real model, the taxonomy groups used are described in the next step in tables 4.4, 4.5.2, 4.6.

4.5.2 II Step

In this second phase, the results are evaluated. To do this, a dataset was built containing different types of indicators including hash, ip and domains. The dataset has been merged into a .JSON file, which in addition to listing the different IOCs also contains the taxonomy.
The evaluation was performed by iterating a different taxonomy set for each simulation with different weights described in equation 1 (4.3.2).

Category	TP	FP	Total
IP	30	20	50
Domain	15	15	30
Hash	30	20	50

Table 4.3. Dataset used for test

The choice of taxonomy set is taken according to the average case in IoCs analyzed in the first phase, and it is described by taxonomy set 1(4.4), while taxonomy set 2(4.5.2) and 3(4.6) are respectively the best case in which we can describe an IoC using standard TAXII, while taxonomy set 3 is the worst case.

namespace and predicate	value
osint:certainty	50

Table 4.4. Taxonomy Set - 1

namespace and predicate	value
admiralty-scale:source-reliability	a
admiralty-scale:information-credibility	1
estimative-language:likelihood-probability	very-likely
estimative-language:confidence-in-analytic-judgment	high
osint:certainty	50

Table 4.5. Taxonomy Set - 2

namespace and predicate	value
admiralty-scale:source-reliability	g
admiralty-scale:information-credibility	6
estimative-language:likelihood-probability	unlikely
estimative-language:confidence-in-analytic-judgment	low

Table 4.6. Taxonomy Set - 3

Next the results of the tests performed are reported, each simulation is numbered and represents a different configuration, all possible combinations of the weights were tested, and for each configuration involving an active w_{tag} a different Taxonomy Set was simulated.
The accuracy expressed in the table 4.7 were expressed by data category, IP, Domain and Hash. Accuracy was analysed for each indicator type because we wanted to observe the behaviour of the system by analysing its classification by data type.
While the accuracy of the overall system is reported in the confusion matrix 4.8.
We also wanted to report the confusion matrix because it is an excellent tool for assessing the quality of the classification model's predictions. In particular, the matrix highlights where the model goes wrong, in which instances it responds worse and which ones better. By looking at it, one can analyse how many objects in the dataset are classified as expected.

Case	w_{tag}	w_{event}	w_{vt}	Tax. Set	IP	Domain	Hash
1	0	0	1	-	38/50(76%)	29/30 (96%)	49/50(98%)
2	0	1	0	-	28/50(56%)	15/30(50%)	43/50 (86%)
3	0	1	1	-	42/50(84%)	29/30(96%)	49/50(98%)
4	1	0	0	1	28/50(56%)	15/30(50%)	43/50(86%)
5	1	0	0	2	28/50(56%)	15/30(50%)	43/50(86%)
6	1	0	0	3	20/50(40%)	15/30(50%)	22/50(44%)
7	1	0	1	1	42/50(84%)	29/30(96%)	49/50(98%)
8	1	0	1	2	50/50(100%)	29/30(96%)	49/50(99%)
9	1	0	1	3	20/50(40%)	23/30(76%)	46/50(92%)
10	1	1	0	1	28/50(56%)	15/30(50%)	43/50(86%)
11	1	1	0	2	28/50(56%)	15/30(50%)	43/50(86%)
12	1	1	0	3	20/50(40%)	15/30(50%)	22/50(44%)
13	1	1	1	1	42/50(84%)	29/30(96%)	49/50(98%)
14	1	1	1	2	50/50(100%)	29/30(96%)	50/50(100%)
15	1	1	1	3	20/50(40%)	15/30(50%)	36/50(72%)

Table 4.7. Final Results

Case	Overall Confusion Matrix			Precision	Recall	Accuracy
1		PREDICTED: P	PREDICTED: N	81%	100%	89%
	ACTUAL: P	60	14			
	ACTUAL: N	0	56			
2		PREDICTED: P	PREDICTED: N	100%	62%	66%
	ACTUAL: P	74	0			
	ACTUAL: N	44	12			
3		PREDICTED: P	PREDICTED: N	86%	100%	92%
	ACTUAL: P	64	10			
	ACTUAL: N	0	56			
4		PREDICTED: P	PREDICTED: N	100%	62%	66%
	ACTUAL: P	74	0			
	ACTUAL: N	44	12			
5		PREDICTED: P	PREDICTED: N	100%	62%	66%
	ACTUAL: P	74	0			
	ACTUAL: N	44	12			
6		PREDICTED: P	PREDICTED: N	8%	100%	43%
	ACTUAL: P	6	68			
	ACTUAL: N	0	56			
7		PREDICTED: P	PREDICTED: N	86%	100%	92%
	ACTUAL: P	64	10			
	ACTUAL: N	0	56			
8		PREDICTED: P	PREDICTED: N	98%	100%	99%
	ACTUAL: P	73	1			
	ACTUAL: N	0	56			
9		PREDICTED: P	PREDICTED: N	44%	100%	68%
	ACTUAL: P	33	41			
	ACTUAL: N	0	56			
10		PREDICTED: P	PREDICTED: N	100%	62%	66%
	ACTUAL: P	74	0			
	ACTUAL: N	44	12			
11		PREDICTED: P	PREDICTED: N	100%	62%	66%
	ACTUAL: P	74	0			
	ACTUAL: N	44	12			
12		PREDICTED: P	PREDICTED: N	8%	100%	43%
	ACTUAL: P	6	68			
	ACTUAL: N	0	56			
13		PREDICTED: P	PREDICTED: N	84%	100%	92%
	ACTUAL: P	64	10			
	ACTUAL: N	0	56			
14		PREDICTED: P	PREDICTED: N	98%	100%	99%
	ACTUAL: P	73	1			
	ACTUAL: N	0	56			
15		PREDICTED: P	PREDICTED: N	20%	100%	54%
	ACTUAL: P	15	59			
	ACTUAL: N	0	56			

Table 4.8. Confusion Matrix

4.5.3 Conclusions

Conclusions are made by analyzing the different accuracy for each type of data 4.7 and the subsequent matrix of confusion 4.8.
We can observe as the first thing that Virus Total alone is not enough despite having a remarkable accuracy, it is shown that alongside the scoring system, they use the taxonomy set 1, the overall accuracy increases by three percentage points.
The best result was achieved in case study 8, where it is shown that using appropriate tags, the classification of the IoC is excellent, reaching an accuracy of 99%, unfortunately, often a description like that of the taxonomy set 2 is rarely present in the IoC received by the companies.
It is shown that the model also works with the taxonomy set 3 which represents a poor description of the event, although the information is classified as unreliable, the system reaches an accuracy of 68%.
About the w_{event}, the system has not proved elastic to this measure, the most appropriate use of which would be to measure the reliability of the source over time, because including it in equation 1 (4.3.2) has proved irrelevant.
In all the case studies there has been an excellent measure of recall, so it has happened little often that a False Positive has been classified as a True Positive, while some TP has been classified as FP has been shown in case studies 2, 4, 5 and 11, which demonstrate that the scoring system based only on taxonomy is not enough, but must be combined with VirusTotal. In any case, neither of the two systems can work alone, they have proved complementary.
The system is believed to be reliable in classifying hashes, this is an expected result considering the large number of malware blacklists recorded by signature.
It also proves to be good with the classification of domains, while the classification of IPs is more difficult, partly because if there is little taxonomy the scoring is mainly deputed to the VirusTotal component, which in the case of a recent IP has limitations.

4.5.4 Limits and Future Works

The limitations found in the design of the model were mainly found in the availability of IoCs that could represent recent information, of which the dataset constructed for the testing phases of the product is evidence.
Another strong limitation found during the research is that threat intelligence companies tend to attach to compromise events sparse documentation/taxonomy related to the event itself, and this does not simplify the scoring model, very often the information is deprived of tags or there are only a few tags that are useful for the purposes of the final score.
There are multiple extensions and future developments that can be implemented to the scoring model. As a first suggestion, it is possible to extend the compatibility of the product to more supported standards in addition to STIX/TAXII.
Another use is to extend the scoring product to multiple threat intelligence sources so that the final reports produced by the software can also estimate the reliability of the source from which the feeds are received, thereby computing a final weighted average of the same IoC compared to multiple sources.

This will allow the company to make assessments regarding the threat intelligence companies it collaborates with. Since these choices influence the safety management of a company, taking them in accordance with periodic reliability measurements is certainly an important turning point in the company business model.
Another possible implementation to extend my model is the addition of an additional step at the end of the processing of the results. In fact, it is possible to upload the final report directly into the company's defence systems like EDR, thus enabling better time management in the detection and prevention phase. The security analyst will thus no longer have to manually search for compromises, but can invest the saved time in other useful actions and phases of incident handling. In addition, with this approach it is possible to make the logic implemented in the EDR work for an automatic response to the attack.

Chapter 5

Case study

This section will analyze, how the feed scoring software presented in the previous chapter could have improved some phases of Incident Handling.
As documented earlier, the steps of Incident Handling are rigid and potentially time-consuming, so it is important, especially in an enterprise context, to reduce risks and improve defense systems regarding those deputed to attack prevention and detection.
The case studies analyzed will relate to a phishing campaign that aims to exfiltrate sensitive data, and ransomware that aims to create a command and control botnet to encrypt company data and demand a ransom.
Before proceeding to the presentation of incident management steps, it is appropriate to present a possible network and defense architecture of the enterprise network, what actors are involved, and where the feed classification software would be placed.

5.1 Logical Scheme

Figure 5.1. Brief presentation of the architecture

A brief presentation of the components involved:

- EDR technology is used for real-time threat monitoring; its main tasks are to analyze data about network traffic that may pose a threat and about incident

response affecting corporate endpoints.
It provides end-to-end visibility into the activity of every endpoint in the enterprise infrastructure, all of which can be done from a single console and with security intelligence tools that can be used for further investigation.

- The IPS is the network security tool that continuously monitors a network for suspicious activity and is tasked with preventing it by using alerts to system admins or blocking it.
  It is more complex than an intrusion detection system (IDS), which is limited only to the detection of malicious activity without it being able to be stopped in any way.

- DNS security tools are mainly used in those specific cases where the protocol involved is exploited by hackers to bypass traditional protection systems that analyze normal internet traffic. This approach is common in communications between trojans and command and control servers useful for sending commands to compromised systems and for the data exfiltration process.
  Well-structured malware often use the DNS protocol to "hide" commands sent to infected computers, as the basic configuration of traditional protection systems tend to overlook the content of DNS queries sent.

5.2 Case 1 - Phishing Campaign

In this section, how a phishing campaign can be stopped or its spread contained by software on IoCs classification, Incident Handling steps will be analyzed.

5.2.1 Prevention

There are numerous precautions to take to avoid running into a phishing campaign. First, it is important for staff to be prepared to follow small precautions such as verifying a site's security and keeping the browser up to date.
Regarding the involvement of realized software, if the company has been informed of a phishing campaign targeting companies in the same business, filtering of IoCs using the appropriate taxonomy is appropriate.
There are many tags used by cyber threat intelligence companies to describe a phishing campaign: [17]

- circl:incident-classification : It is a tag used by CIRCL to classify the incident. The value to set in this case is "phishing".

- circl:methods-tactics : It is often useful to describe, techniques, tactics and procedures (TTP) used by cyber criminals, setting the predicate with "data-exfiltration" is a reasonable setting considering it is the first purpose of a phishing campaign.

- information-security-indicators : The information security indicators are a comprehensive set of operational indicators that organizations can use to assess their

security posture. Setting the predicate to "PHI.1" looks for IoCs related to phishing targeted at company workstations that harms the company's image or business.

- common-taxonomy:information-gathering : Describes an active and passive gathering of information on systems, unauthorized monitoring and reading of network traffic or attempt to gather information on a user or a system through phishing methods. In this case it is appropriate to set to "phishing".

It may be appropriate to lower the score threshold for which the IoC will be recognized as valid information to an appropriate value.
With this approach, more IPs and domains potentially involved in the phishing attack will be loaded into the EDR component than using a standard threshold.
Thus, there will be the advantage of a larger number of indicators that can generate an alert, but there will be a higher risk of blocking legitimate traffic to harmless hosts and domains.

5.2.2 Detection and Containment

In the detection phase, the scoring software is not directly involved, but nevertheless it could certainly help considering the considerations adopted in the previous section (5.2.1). When it comes to detection of a phishing campaign, it is advisable to monitor detection channels, both automatic and manual, customer and staff channels for clues of a data breach or compromise.
At this stage, detection is mainly from a report from internal staff, or it could come from the EDR as a report of unauthorized access from a foreign location with which a data exfiltration event and thus phishing event can definitely be associated.
One possible choice is to temporarily block incoming traffic from states with which you have no business agreement and no business, because you don't know where the phishing traffic is coming from, this is to prevent another workstation from being involved in the attack.
Second, identify the systems impacted or at risk of impact and if the dynamics are not known isolate the affected systems and apply restrictive access control rules to prevent any kind of access to the production network. You can apply whitelisting/blacklisting techniques where you specify for the compromised user the list of authorized and blocked applications, in addition, content filtering is also useful, thus restricting access to documentation that could compromise other resources or the company's reputation.

5.2.3 Eradication, Recovery and Post-Incident

Once the incident has been detected and contained it is necessary to proceed in the eradication of the cause of the incident and prevent it from replicating.
So the main thing is to reinforce the authentication methods of internal personnel, so that a credential theft is not compromising to the user, a multifactor authentication use is the most appropriate solution in such cases. It is also important to reduce the exposure of sensitive data, and make it accessible to a narrow user base.

It is important to verify that credentials are saved securely, to prevent a compromise to the server from putting many users at risk. You also need to make sure that no session data is exchanged in an insecure manner. Regarding security on data, it is important to classify the data and adopt protection methodologies according to the classification. Important assure that sensitive data are encrypted with recent and strong algorithms so that an adversary cannot exploit any implementation weakness. Once this is done, a secure key management system should be adopted, and keys should be stored in controlled environments or on dedicated hardware.
In the recovery phase, once the systems are restored, it is necessary to search for all services accessible from the profiles whose credentials were stolen and proceed to change passwords as soon as possible, and then reintegrate the compromised systems.
In the post-incident phase it is needed to report details of the incident and remediated on the network, including timing, actions taken, as well as the effect on users.
Draw up guidelines on possible aspects of new training campaigns for internal staff to help prevent a recurrence of a similar incident.

5.3 Case 2 - Ransomware Attack

According to an analysis conducted by the SonicWall Institute, there were 304.7 million ransomware attacks in the first half of 2021, a 151% increase since 2020. [1] Protecting a company from a ranson attack is critical; the data and economic losses would be significant.

5.3.1 Prevention

The techniques adopted as prevention activities for a ransomware attack are similar to all those related to malware prevention. As a main precaution, it is advisable to backup systems regularly and keep copies in a different environment to the internal network, also it is good to use different credentials for backups so that if the network is compromised, storage space remains secure.
The main attack vector of a ransom attack is the mail channel, so it is advisable to train internal staff to a secure management of email attachments.
Regarding the scoring system, as in case 1 (5.2), it is advisable to filter the feeds received from CTI companies, configuring the best taxonomy to filter the data.

- cccs:malware-category : It identifies the malware category, simply setting it to "ransomware".
- circl:incident-classification : It is a tag used by CIRCL to classify the incident. The value to look for is "ransomware".
- ecsirt:malicious-code : malicious-code describes that the hash of shared file, is intentionally created for a harmful purpose and in order to activate the code it is needed a user interaction.
- In addition, the MISP uses an entire taxonomy to describe the state of malware spread, targets, communication methods and its complexity.

In this case, once the scoring software automatically loads the IoCs in defense systems, such as the EDR it is advisable to set the block response action, as an alert could be ignored by a careless user.

5.3.2 Detection and Containment

Detection of ransomware is a very important step because it is important to do so before the cyber kill chain of the attack reaches workstations with a higher level of administration and it is also crucial to stop malware before it encrypts data.
At best detection can be implemented by signature. Each malware has its own signature, characterized by domains, IPs and other information that identify it.
Defense systems can then recognize them by their signature and block them, but malware change over time and also change their signature and can not block what you do not know. Ransomware can also be detected by abnormal traffic.
More complex ransomware attacks often have a double purpose, encrypting data for

ransom, but they also steal data before encrypting it to sell to third parties. This implies large data transfers to systems outside the network, a definitely abnormal behavior that can be detected.
In addition to relying on EDR's automatic detection systems, other complementary detection techniques can be followed:

- Dynamic analysis: this type of approach is also called behavioral analysis. Port monitoring, process and registry monitoring methods are the most commonly used techniques. In the case of ramsomware, there is always a tendency to open ports to allow connection to the remote control system, or they disguise their behavior by using seemingly harmless processes for the purpose of spreading the malware to other workstations.

- Static analysis: using this method we look for malicious files in the system, the analysis is also called dump analysis. The analysis is performed without running the malicious code, and some common detection techniques are using fingerprinting files in order to detect differences, or using malware scanning tools. On Windows systems, analysis of PEI (Portable Executables Information) is possible, which are the executable files that contain metadata about a file, such as imported and exported functions and linked libraries.

The containment phase begins once an infection has been detected on a host that needs to be isolated from the rest of the network to prevent further spread of the malware . It is important that once the system is isolated, the status of the information is maintained so that if necessary forensic analysis can be performed.
One approach used in the containment phase is malware analysis, after isolating the systems involved, the malware can be analyzed either by reverse engineering techniques to identify data erasure functions, or in sandboxing mode, in this mode the malware is executed in a controlled environment, the security team then analyzes its behavior and studies data obfuscation and extraction techniques in order to find an implementation flaw to exploit for data recovery without paying a ransom.

5.3.3 Eradication, Recovery and Post-Incident

In case of ransomware, rebuilding a host is typically more resource-intensive than other eradication solutions, it should be performed only when no other eradication method is sufficient. In general, rebuilding should be taken in account by company for any system that has any of the following incident characteristics, such as the attacker get administrator-level access to the system or unauthorized administrator-level access to the system was available to anybody through a backdoor. Whether the system does not function properly after the malware has been eradicated by antivirus software, it is an indicator about the malware has not been eradicated completely, so even in this case the rebuilding solution should be evaluated by security team.
The primary objectives in the recovery phase are definitely to restore damaged systems and restore data, then resume business continuity as soon as possible.
This step will be faster the more well defined the data recovery plan is. If an inventory

of data has been provided that illustrates how it has been categorized and where it is stored, the recovery process will be faster.It is also important to define all endpoints in the recovery plan, to establish recovery priorities and level of security.
Restoring data from uninfected secure backup copies is the main task of this phase, the restore can be done with command line work or more easily, using recovery tools offered by our security vendors that process this task automatically.
Because severe malware incidents can be extremely expensive to manage, it is important for organizations to conduct lessons learned for severe malware incidents.
The process of lesson learned for malware is no different than that of any other type of incident. Some activities to be performed in this can be, changes to security policies. Security policies could be changed to prevent similar incidents.
For example, if the ransomware has been spread with a particular extension by attaching it to an email, it is reasonable to change the policy to block the sending of emails with a document attached to that extension, user training on security is a common denominator in the lesson learned phase, regardless of the attack, the internal staff is an asset for the company and as such must be protected therefore investing in training and periodic courses is useful to reduce the number of infections or improve the user actions in reporting incidents. Depending on the dynamics of the incident, it may be appropriate to modify the configurations of the operating system or detection tools to allow detection and a prompt response for a similar future attack.

Chapter 6

Conclusions

6.1 General considerations

The thesis in the company allowed me to grow both personally and professionally. Experiencing a corporate environment has shaped me greatly, observing internal dynamics and taking advice from colleagues will surely prove invaluable throughout my career.
Working in a team was definitely what motivated me the most, as was interfacing with a real business process. The thesis in the company allowed me to have direct contact with managers of threat intelligence companies, to confront myself with them in order to actualize as concrete a model as possible and to understand how a company of that level works with data and how that data is analyzed.

6.2 Technical considerations

This work allowed me a broad exploration of the topic of cyber threat intelligence with a focus on indicators of compromise between companies in the same industry in order to prevent attacks and malicious actors.
A threat analysis was done through the corporate EDR console in order to gather elements for proper prioritization and severity classification. At the end of the work, a good ability to evaluate the sources of information sharing services related to the publication of cyber threats and indicators of compromise was acquired, understanding what information represents concrete and useful data for the final project.
The development of source code related to the final project presented me with new libraries useful for implementation purposes for communicating with external services for receiving cyber event feeds.
Commenting on the results, it can be seen that the measurement that includes only VirusTotal despite having a remarkable accuracy remains inferior to the combination of the measurement with the taxonomy, in fact it is shown that by combining the scoring system with the use of the taxonomy set 1, the overall accuracy increases by three percentage points.

The best result was achieved in case study 8, where it is shown that by using appropriate tags, the classification of IoCs is excellent, reaching an accuracy of 99%; unfortunately, a description such as that of taxonomy set 2 is rarely present in the IoCs received from companies.
It is shown that the model works even with a sparse taxonomy (taxonomy set 3), the system achieving 68% accuracy despite the indicators not being rich in tags.
As for the w_{event}, the system did not prove to be elastic towards this measure, whose more appropriate use would be to measure the reliability of the source over time, because including it in equation 1 (4.3.2) proved to be irrelevant in conditioning the results.
In all case studies there was a very good measure of recall, so that it was not often that a False Positive was classified as a real threat.
The final project is able to classify indicators of compromise according to descriptive tags by interacting with external security vendors, collect partial results, and produce the final reports.
The model is shown to achieve relevant accuracy in classifying data.
The final results demonstrate how the integration of a data scoring model can effectively improve the classification of feeds received from threat intelligence services. Improving the prevention and detection phases of an Incident Handling process, in terms of detection time and reducing the risks to which a company may be exposed.
The limitations encountered in the design of the model were mainly found in the availability of IoCs that could represent recent information, of which the dataset built for the testing phases of the product is evidence.
Another limitation found in the course of the research is that threat intelligence companies tend to attach little documentation/taxonomy to compromise events related to the event itself, and this does not simplify the scoring model, very often the information is untagged or there are only a few tags useful for the final score.
Future developments of the model include extending the compatibility of the product to more supported standards besides STIX/TAXII. Improving the measurement of w_{event} and extending the reception of feeds to more threat intelligence sources, so that the final reports produced by the software can also estimate the reliability of the source from which the feeds are received.
An additional implementation to extend the model is to fully integrate it into an enterprise environment; it is possible to upload the final report directly into an organisation's defence systems such as EDR, thus enabling better time management in the detection and prevention phase. The security analyst no longer has to search manually for compromises, but can invest the saved time in other useful actions and steps of incident management. Furthermore, with this approach, it is possible to make the logic implemented in the EDR work for an automatic response to the attack.

Bibliography

[1] *2022 SonicWall Cyber Threat Report*. URL: https://www.sonicwall.com/2022-cyber-threat-report/.

[2] Sarah Brown, Joep Gommers, and Oscar Serrano. "From Cyber Security Information Sharing to Threat Management". In: Association for Computing Machinery. DOI: 10.1145/2808128.2808133. URL: https://doi.org/10.1145/2808128.2808133.

[3] Stephan Chenette. *Emulating Attacker Activities and The Pyramid of Pain*. URL: https://www.attackiq.com/2019/06/26/emulating-attacker-activities-and-the-pyramid-of-pain/.

[4] Michael Cloppert. *Defining Cyber Threat Intelligence*.

[5] Mike Cloppert. "Security intelligence: Attacking the cyber kill chain". In: *SANS Computer Forensics* 26 (2009).

[6] *Defense in Depth*. URL: https://www.networkaccess.com/defense-in-depth/.

[7] *ENISA Threat Landscape - List of top 15 threats*. URL: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-list-of-top-15-threats.

[8] *Flag files and URLs as malicious or harmless*. URL: https://support.virustotal.com/hc/en-us/articles/115002146769-Vote-comment.

[9] UK Government. "Cyber-threat intelligence information sharing guide". In: (Mar. 2021).

[10] Adam Kliarsky Hun-Ya Lock. "Using IOC in Malware Forensics". In: URL: https://www.giac.org/paper/grem/2593/ioc-indicators-compromise-malware-forensics/125039.

[11] *ISO/IEC 27000:2014*. URL: https://www.iso.org/standard/63411.html.

[12] Charles Schmidt Julie Connolly Mark Davidson. "The Trusted Automated eXchange of Indicator Information (TAXII™)". In: *THE MITRE CORPORATION* ().

[13] Ollie Whitehouse Kirsty Paine. "Indicators of Compromise (IoCs) and Their Role in Attack Defence". In: *UK National Cyber Security Centre* (July 2020). URL: https://www.ietf.org/archive/id/draft-paine-smart-indicators-of-compromise-01.xml.

[14] Kevin Mandia. *Incident Response & Computer Forensics, 2nd Ed.*

[15] *Matplotlib — Visualization with Python*. URL: https://matplotlib.org/stable/users/release_notes.

[16] Rob McMillan. *Gartner Research: Threat Intelligence*. URL: https://www.gartner.com/en/documents/2487216#:~:text=Threat%20intelligence%20is%20evidence%2Dbased%22.

[17] *MISP taxonomies and classification*. URL: https://www.misp-project.org/.

[18] *MITRE ATT&CK - Deliver Techniques*. URL: https://attack.mitre.org/.

[19] *Numpy Documentation*. URL: https://numpy.org/doc/stable/.

[20] Cyber Security Programme. "Capability Maturity Model". In: *Digital, Data Technology Gov UK* (). URL: https://hodigital.blog.gov.uk/wp-content/uploads/sites/161/2020/03/Detecting-the-Unknown-A-Guide-to-Threat-Hunting-v2.0.pdf.

[21] Sasha Romanosky. "Examining the costs and causes of cyber incidents". In: *Journal of Cybersecurity* 2.2 (Aug. 2016), pp. 121–135. DOI: 10.1093/cybsec/tyw001. URL: https://doi.org/10.1093/cybsec/tyw001.

[22] Mattia Siciliano. *La Cyber Threat Information Sharing: differenze di approccio tra Malware Information Sharing Platform (MISP) e Threat Intelligence Platform (TIP)*. URL: https://www.ictsecuritymagazine.com/articoli/la-cyber-threat-information-sharing-differenze-di-approccio-tra-misp-e-tip/.

[23] *Spam - ENISA Threat Landscape*. URL: https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/etl-review-folder/etl-2020-spam.

[24] "Taxonomy driven indicator scoring in MISP threat intelligence platforms". In: (). URL: https://doi.org/10.48550/arXiv.1902.03914.

[25] Antonio Villalón-Huerta, Ismael Ripoll-Ripoll, and Hector Marco-Gisbert. "Key Requirements for the Detection and Sharing of Behavioral Indicators of Compromise". In: *Electronics* 11.3 (2022). ISSN: 2079-9292. DOI: 10.3390/electronics11030416. URL: https://www.mdpi.com/2079-9292/11/3/416.

[26] *VirusTotal's API*. URL: https://developers.virustotal.com/reference/overview.

[27] Anne W. "Maturity models in cyber security". In: *National Cyber Security Centre* (). URL: https://www.ncsc.gov.uk/blog-post/maturity-models-cyber-security-whats-happening-iamm.

[28] *Web-based attacks - ENISA Threat Landscape*. URL: https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/etl-review-folder/etl-2020-web-based-attacks.

[29] *What is Security in Depth?* URL: https://www.networkaccess.com/defense-in-depth/.

[30] Tarun Yadav and Arvind Rao. "Technical Aspects of Cyber Kill Chain". In: DOI: 10.1007/978-3-319-22915-7_40.