Semester:
ICTE4

Title:
MISP Engine to Assess and Evaluate Threat Events Based on Data Quality

Project Period:
February - June 2021

Aalborg University
Copenhagen
A.C. Meyers Vænge 15
2450 København SV

Semester Coordinator:
Henning Olesen

Secretary:
Charlotte Høeg

Semester Theme:
Master Thesis

Supervisor(s):
Henning Olesen
Peter Anglov

Project group no.:
4.2

Members
(do not write CPR.nr.):
Niklas Christensen

Pages:
101

Finished:
03/06-2021

Abstract:

This written report will investigate the concepts of threat intelligence sharing and the improvement of such. Low data quality in threat intelligence sharing is having negative consequences on the trust in the threat communities. This report will explore the concepts of data quality, reputation system and trust. It will explain how an algorithm programmed in Python, can help establish more trust in the MISP community. MISP is a threat intelligence platform, which is open source and free. Therefore, large amount of data is published every day. This data is evaluated manually by network forensics, the algorithm programmed will be tool for the forensics, to the better and faster make an actionable decision on cyber threats.

MISP Engine to Assess and Evaluate Threat Events Based on Data Quality

-

Data Quality is important in threat intelligence sharing to maintain a high level of trust. This report will investigate the data quality in the MISP events and come up with a suggestion to enhance the trust between members in the community.

Author:
Niklas Christensen

Supervisors:
Henning Olesen
Peter Anglov

Master of Science Thesis
Aalborg University
Innovative Communication Technologies and Entrepreneurship

Acknowledgement

The creation of this project has only been possible, because of the generous help from Dennis Rand. Dennis Rand has been extremely supportive in aiding with personal experience and very valuable insight knowledge about threat events and the MISP platform.

The supervisors of the project, Henning Olesen and Peter Anglov, must also be credited for helping to shape the project, and steer it in the correct way. A special thanks to Peter Anglov must be emphasised for creating connections to relevant experts in the field and being very supportive throughout the entire project.

A special thanks must also be given to Nicolaj, Alexandre Dulaunoy (CIRCL), Ming Shen, Emmanouil Vasilomanolakis and Cedomir Stefanovic, for assisting with knowledge in their respective field of study.

List of Figures

Definitions

- TIP – “*Threat Intelligence Platform” -* solution to facilitate the management of cyber threat intelligence and associated entities such as actors, campaigns, incidents, signatures, bulletins, and TTPs [53].
- CTI – “Cyber Threat Intelligence” - Threat intelligence is knowledge that allows you to prevent or mitigate cyberattacks [52].
- MISP – *“Malware Information Sharing Platform” [7].*
- MISP Event – *“MISP events are encapsulations for contextually related information represented as attribute and object”* [14].
- MISP Galaxy – *“MISP galaxy is a simple method to express a large object called cluster that can be attached to MISP events or attributes”* [14].
- Objects – *“MISP objects are in addition to MISP attributes to allow advanced combinations of attributes.”* [14.]
- Attribute – *“Attributes in MISP can be network indicators (e.g., IP address), system indicators (e.g., a string in memory) or even bank account details.”* [14].
- Event Features – *“Event features are the pieces of information that the data quality algorithm will measure upon” [14].*

Abbreviations

- DQ – “*Data Quality”*
- ZMQ – *“Zero MQ”*
- GUI – “*Graphical User Interface”*
- DQD – *“Data Quality Dimensions”*
- KPI – *“Key Performance Indicators”*
- FE – *“Forsvarets Efterretningstjeneste”*

1 Introduction

The digital era is upon us [1]; the age of digitalisation is characterised by shifting from industrial processes to a more information-based approach with help from computers, machines, smartphones, and more. These are often used as a medium for communication, which essentially means that data is being sent back and forth, processed, analysed, and stored. Data in today's society is very valuable, data is embedded in many systems, and many systems is becoming digitalised.

Digital data is highly essential in today's economy; all over the globe and across borders, the data is being sold, gathered, and in some cases stolen; some even argue that data is a new currency [2]. According to the Oxford Learners Dictionaries [3], the definition of *currency* is:" *the fact that something is used or accepted by a lot of people*", which data is. 'We' as the users of different online trading and communication platforms like Facebook, Instagram or eBay, accept that 'we' are allowed to consume 'their' services, but as a trade-off, they are allowed to collect, store and sell information about 'us' [4]. The collected information is often highly sensitive and can be used to identify individuals; therefore, this data must be kept secure and protected. One way of being proactive in protecting sensitive resources and infrastructure is to apply cybersecurity software and frameworks capable of analysing and evaluating network traffic and seeking out harmful and actual cyber threats.

1.1 Background

In today's network landscape and with the possibilities of modern computing, data takes many different forms and is used for many objectives. Threat Intelligence data will be the primary focal point of this thesis.

'Threat intelligence' and the sharing of such plays an essential role in protecting against malicious network activity, and this term is central to this report. The definition of Cyber Threat Intelligence (CTI) is:" evidence-*based knowledge about adversaries – their motives, intents, capabilities, enabling environments and operations – focused on an event, series of events or trends, and providing a decision advantage to the defender.*" [5]. It can be learned from this definition that CTI is the act of collecting, organising, and sharing data about potentially harmful activity, and as a result, trying to shed light on malicious traffic and giving the defender an advantage when making essential security decisions.

As mentioned earlier, data is precious in modern societies, and criminals will often try to conduct attacks on organisations to steal this data or control systems to gain a benefit. Cybercriminals are always trying to think of new ways to attack infrastructures, and it is the job of network forensics and cybersecurity analysts to develop and better information systems to withstand such attacks.

The cost of developing and improving computer systems to better withstand cyber-attacks and lower the risk of being successfully attacked is high. In 2020 the cybersecurity market was set to grow to around 170 billion dollars worldwide [6]. However, not preparing for cyber-attacks can be even more costly; it can be hard to estimate how much a cyber-attack will cost to recover, but online resources [60][61] can agree on one thing, the numbers are often high. The high numbers can be crushing for companies who do not possess the capital to recover and prepare for cyber-attacks.

Many Threat Intelligence Platforms (TIP) which specialise in collecting and organising data about threats, are already on the market. This project will dive deep into the platform of MISP (Malware Information Sharing Platform), which is a widely known platform supported by NATO and the EU [7]. The MISP is gathering threat intelligence information from all parts of the world, and a negative result of that is an overload of data [69]. This data overload is highly problematic because the network forensics must evaluate each incoming threat and decide upon an action. That is a problem on its own; however, the forensic teams also must deal with the fact that anyone can upload threats reports to the MISP platform (feed community). This essentially means the data quality is only as good as the person uploading it; this has consequences on the overall data quality of the events, meaning that some events are not helpful at all; this can lead to lowered trust in the community.

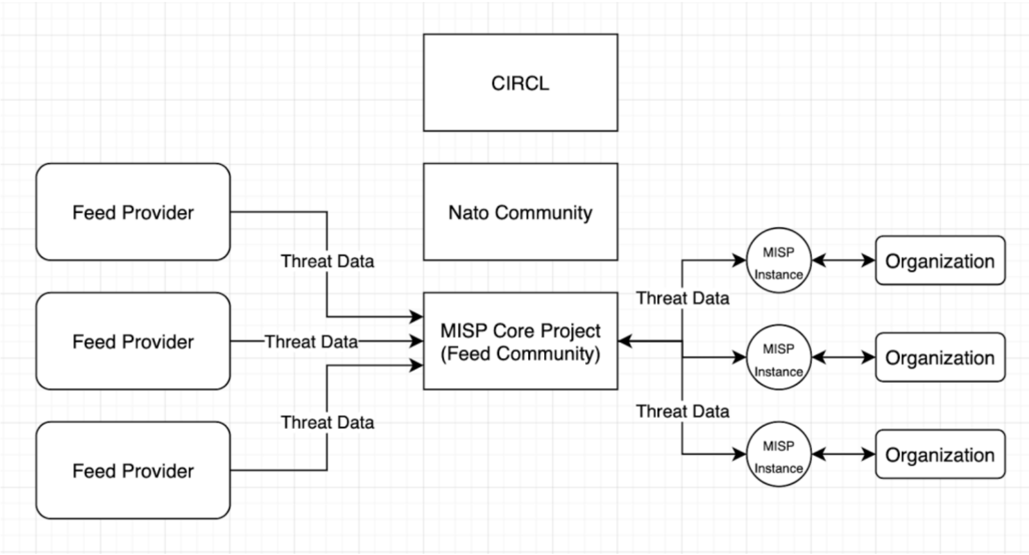


Figure 1 - The MISP landscape, with a few known communities.

The MISP platform will be the TIP of interest in this thesis. Figure 1 describes the organisation structure of the MISP project [8]. In the MISP landscape, different communities exist; the most ’typical’ one is the ‘feed community’, where anyone is allowed access, and is given the ability to publish threats events and become so-called feed providers. Feed providers [48] are specialised in analysing and sharing an immense amount of threat intelligence data. Besides the feed community, other private communities also exist [8], for example, the CIRCL community and the NATO community [8]. These communities require that users get approved before fetching being able to fetch threat information. All users of the MISP have their private MISP instance, where threat data is stored and sent to.

The main community of interest in this project will be the feed community, since it is freely available, however, the entire concept of MISP communities will be explained thoroughly later in the report.

The MISP framework already has many features, like input validation, false positive warning lists, etc., however, a sorting algorithm, going through each incoming threat event, calculating how good the data quality is, scoring, and tagging it, has not yet been developed. This thesis will focus on creating one. This written report will research the market of TIP’s, find what it already made, and dive deep into the MISP community. By doing that, the needs of the MISP platform can be analysed and lead to a possible solution as a proof of concept (PoC) including necessary architecture plans, diagrams, and explanations. Furthermore, it will explain the thoughts done while creating the PoC, using the concepts and inspiration from the analysis. Lastly, this report will be handed over to the founder of eCrimeLabs [6], Dennis Rand, and The Danish Defence Intelligence Unit (Forsvarets Efterretningstjeneste) [7]. This is done to further develop the concept and its functionality, which may later be used in production.

1.2 Motivation

The foremost motivation was the help to battle malicious activity on the internet. MISP is a good tool for creating protective measures against malicious activity, and this project's main contribution is a system that can enhance the tool of MISP. It is the assumption that doing this project, the outcome will be more trust in the MISP communities, because of enhanced data quality, and eventually better protective measures, making the internet a bit safer to navigate in.

Another reason for this project is the personal interests in it. Another project [66] was created where the main contribution was to rate the feed providers based on the information they provided, but the focus point was validation of the data, such as domain names, IP addresses and more. This project takes another angle and will look at the data quality of the events, the first piece of information that comes into the MISP instance.

1.3 Challenges

The first task in the project will be to investigate and research relevant topics, which include: reputation systems and trust, how trust can be created and maintained in an environment or community, where data quality is not always prioritised, and speed combined with response time is essential. Therefore, the topic of data quality must also be investigated to its fullest. Data quality is a broad term, so definitions of the concept must be established, how others have used it must be researched, including necessary metrics and dimensions. The project will include a PoC, therefore similar algorithms must be researched, so inspiration can be gathered. The algorithm is important to the project and is therefore considered to be the main contribution of the project. Other's work is essential; therefore, similar solutions must be studied; an excellent approach to investigate other services could be to contact other companies or experts to hear about different solutions. Threat intelligence is a relatively unstudied field; therefore, contacting experts could provide an insight into this specific and technical topic.

1.4 Problem Formulation

Based on the initial research and background investigation about threat events, data quality, and other relevant frameworks; a problem formulation could be structured:

How can an engine for the MISP platform be designed and developed to assist forensics in evaluating incoming events based on the data quality, to respond more precise to threats?

The thesis will investigate how a TIP such as the MISP platform can be improved by structuring a system, which purpose is to evaluate the threat events based on its data quality. The intention is help network forensics in their job of acting on the threats. The system should also enhance the trust in the community. To practically demonstrate the solution to the problem formulation a PoC will be created, that has the basic functionalities implemented.

1.5 Delimitations

The most limiting factor is time; if more time were provided, a more profound investigation could be made, and a better technical solution could be provided, however, due to the fact that it is an academic thesis, deadlines must be kept.

The report will not propose a fully functioning solution ready to be put into production; it will propose a proof of concept (PoC), where the most basic functionalities will be presented to demonstrate that this problem can, in fact, be solved to some extent. The PoC will not include a designed GUI (Graphical User Interface) since this is time-consuming to design such a solution and is not the focus of this report; however, a low-fi prototype will be provided to illustrate how it potentially could look like in the future.

The field of threat intelligence data and measurement of data quality is relatively new, and therefore few studies and algorithms have been created; this will also be limiting, in fact, that only a few solutions can be used to build upon.

1.6 Initial idea

This brief subchapter will give a short explanation of the concept and provide an early figure of the idea. The MISP instance owner will subscribe to different feeds and synchronise events from communities and trusted partners, which will provide the subscriber with a number of events, depending on the activity of the feed provider. These events will present information about threats from malicious activities on the internet, this information can and should be utilised into the organisation’s infrastructure to minimise the risk of successful attacks.

The events being published to the MISP instance are often ‘user’ made, consequently, the maturity of the way the events are created and populated with data fluctuates in terms of data quality. For example, using the MISP instance, some events are sometimes seen only containing a single attribute (IoC), without any context such as checksums, domains and IPs; making it difficult for the consuming organisation to know if the data can be trusted.

The main contribution of the MISP platform, is to provide an incentive to add more context to the information. Without context making, it is hard for the receiving part to figure out how to utilise the data correctly.

One crucial term mentioned in the report is the IoC [9], Indicators of Compromise; the IoCs are small bits and pieces of information a malicious person will leave after conducting an attack. IoCs can consist of many types of information such as IP, screenshots, email address, checksums, and many more. Discussing the term of IoCs, one often referend diagram is the pyramid of IoC (section 10.1). The pyramid [10] explains how easy or difficult it is for an attacker to alter the IoCs after an attack. The lowest part of the pyramid is the hash values and IPs; these are very easy for the attackers to cancel or change. However, the top-level TTP (Tactics, Techniques and Procedures), which is the attack pattern in the attack, is very hard for the attacker to change. Network forensics should always try to chase the higher levels of the pyramid of an attacker and share this information with others if possible.

The goal of the solution is to create a data scoring model for each threat event providing value to the threat handler, since with the large amount of data coming into the system these days, it is neither an optimal nor a viable path to evaluate all this data manually. The proposed solution will provide valuable knowledge about each event, making it easier to decide how to handle the threat, The data collected will then be used to create a reputation system, that will rate the feed providers, it is the assumption, the solution will create more trust in the community.

This project will provide an open-source solution to evaluate each incoming event in real-time. The solution will be built around a data model, which will give each event a ‘confidence score’, e.g., from 1-100 based on the maturity of the data; this will be stored as a tag to the specific event.

Based on the score, the receiving party will do a risk assessment and decide on how high the confidence scoring has to be to automatically ingest the data into either passive security components such as SIEM systems or active security components such as firewalls and proxies.

Previous scoring systems have been primarily focusing on the attributes (IoCs) themselves; however, this project will be based on how the sending party of an event is describing the event on elements like tags, objects, galaxies, relations, comments, and similar. Nevertheless, the model will not look at or validate the individual IoCs directly.

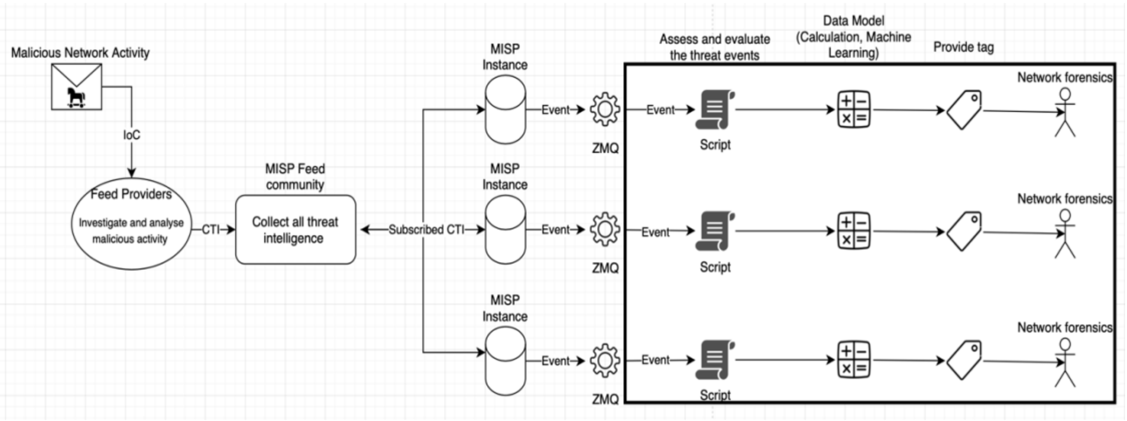

Malicious Network Activity
IoC
Feed Providers
Investigate and analyse malicious activity
CTI
MISP Feed community
Collect all threat intelligence
Subscribed CTI
MISP Instance
MISP Instance
MISP Instance
Event
ZMQ
Assess and evaluate the threat events
Data Model (Calculation, Machine Learning)
Provide tag
Script
Network forensics

Figure 2 - An initial idea to potentially solve the problem of sorting off data of low quality.

Figure 2 depicts the components and the flow of the system. The black box (right part of the figure) in the figure is showing where the contribution to the MISP Platform will be. The focus of the project will be on the incoming threat events and the analysis of those. The first step in the project will be to install the MISP Core instance on a cloud service, this MISP instance will be the central point of the system. The incoming threats will go through the 'ZeroMQ', a queue system (explained more thoroughly later in the report), and it is here where this proposed solution will intercept them and provide a 'confidence score' along with a tag. A data model and algorithm must be developed to provide the confidence score. The initial hypothesis is that, by using an algorithm that will score the events on different measures, including relevant mathematical formulas, it is possible to provide a precise and viable result. Including machine learning models is also heavily considered, and can be applied, if it is found to fit the requirements.

The data model must be dynamic and personal to the individual usage, meaning that the user must be able set the weight of each measure and take out measures that are not needed for the specific case.

To sum it up, the solution will be an ETL [67] (Extract, Transform and Load) solution, fetching data in terms of events from the MISP ZeroMQ, gathering the different features, and analysing them; this will produce a confidence score, which will generate a tag. The event will then be equipped with that tag, which will then be sent to clients SIEM systems or other implemented security handling system, where they can specify the risk, the organisations are willing to take, e.g. the data needs to have a confidence score of 90%, meaning that the data quality of the event must be relatively high, before taking any active action. Furthermore, a simple reputation system will be created to rate feed providers, this will create trust, because users of the MISP can then follow the feed providers in terms of the data quality they deliver.

1.7 Methodology

This subchapter will explain and describe the tools, techniques, and procedures that will be applied throughout the project. This subchapter will explain why they were chosen and how they helped provide valuable knowledge to the project used to answer the problem formulation adequately.

The report will be a scientific report, where development is done and technical specifications will be outlined, leading to the final outcome of the project which will be a fully functioning PoC presenting the basic ideas and functionalities explained throughout the report.

1.7.1 Data Collection

The first step in the creation of the project was to find a sufficient amount of information to gain the correct knowledge about all of the aspects of the report. The information found and used throughout the report, was found using different techniques. The largest part of the data and information found was using desktop research. The desktop research approach will give a good insight into the topic, and what other researchers have already created, this will take the project a long way, and a lot of general knowledge can be found. However, the project is specialised and very technical, working in an area of TI, that is relatively unexplored, and that is why desktop research will not provide all of the answers needed to answer the problem formulation.

To get more specialised knowledge it was decided to reach out to experts in the field. These experts will have first-hand experience with the topic and this experience will be much needed to understand the small niches of the topic, which can be difficult to find during the desktop research. The experts were both contacted via email, and online interviews. The interviews gave a very qualitative view of topic, that included highly specific information about the MISP instance.

It was decided not to purpose the quantitative approach because, as mentioned earlier the project is highly specific, and therefore going out asking random people or creating questionnaires, will not produce the desired outcome.

1.7.2 Prototype

To demonstrate the thoughts of a design, that is considered, a low-fy prototype must be designed, it will be designed a simple sketch, including the most basic functionalities. The algorithm and system that will be create is very abstract and is only background processes of a script. The prototype drawings will therefore help the reader visualise how the system will create value for the user, and how it could look if it one day was put into production. An actual UI of the system is not the focus of the project, and therefore is not implemented, and therefore making drawings of a potential system was found valuable.

1.7.3 Testing

Testing will also be a part of the project; different kinds of testing will be conducted to ensure a high level of performance for the proof of concept. The different methods for testing the system will be some unit testing, stress testing and a test will control data sets, obtained from the experts in the field, this will give a good indication if the system is tuned right, and the correct scores are given to right threat events. The system must run must of the time assessing and evaluating each incoming event; therefore, a stress testing session would be beneficial to see how the system behaves when it is bombarded with threat events.

The unit testing will be done during the implementation of the system, each component and feature will be tested, when it is implemented to make sure that the system is still running, and fully integrated.

1.8 Expected outcome

The expected outcome will be a conceptually designed system, that suggests of the problem formulation can be used, with the most state of the art technologies. The project will include features, that is not expected to be implemented, but it is considered for the future. Besides the proposed system it will furthermore include as mentioned a PoC, that is based on the proposed solution. The PoC will not have all the features implemented, however, the most basic one and the main concepts will be implemented.

2 State of the Art

The State of the Art chapter is essential, because it will provide the reader with vital knowledge to understand the central concepts. The MISP framework, will be the first thing to be explored. Here the central functionalities, terms and techniques will be discussed, these are important to have knowledge about since they will be a great part of the implementation. Academic work will then be presented, this will include important theories, that is very central knowledge for the analysis. Similar solutions to the MISP framework will also be investigated, and inspiration will be gathered that can later be used in the PoC.

2.1 The MISP Framework

The first technology to be investigated is the platform of MISP. The objects to be explored is the concepts, including, concepts like threats events, ZeroMQ, attributes, PyMISP, and more.

The MISP framework [7] was established in 2012, it was started as a single-man project because there were no other options to share threat intelligence besides manually written files. The big potential of a threat sharing platform was quickly picked up by bigger organisations, and in 2016 it was adopted by NATO, and shortly after the EU.

The main purpose of the MISP framework is to share, store, and structure threat intelligence data collectively, with the goal to enhance knowledge sharing, helping one another protect better against cybersecurity attacks.

The MISP framework is open source [7], this allows all organisations who have an interest in cyber threat intelligence, either for protection, sharing, or just curiosity, to join the MISP community, and start using it free of charge.

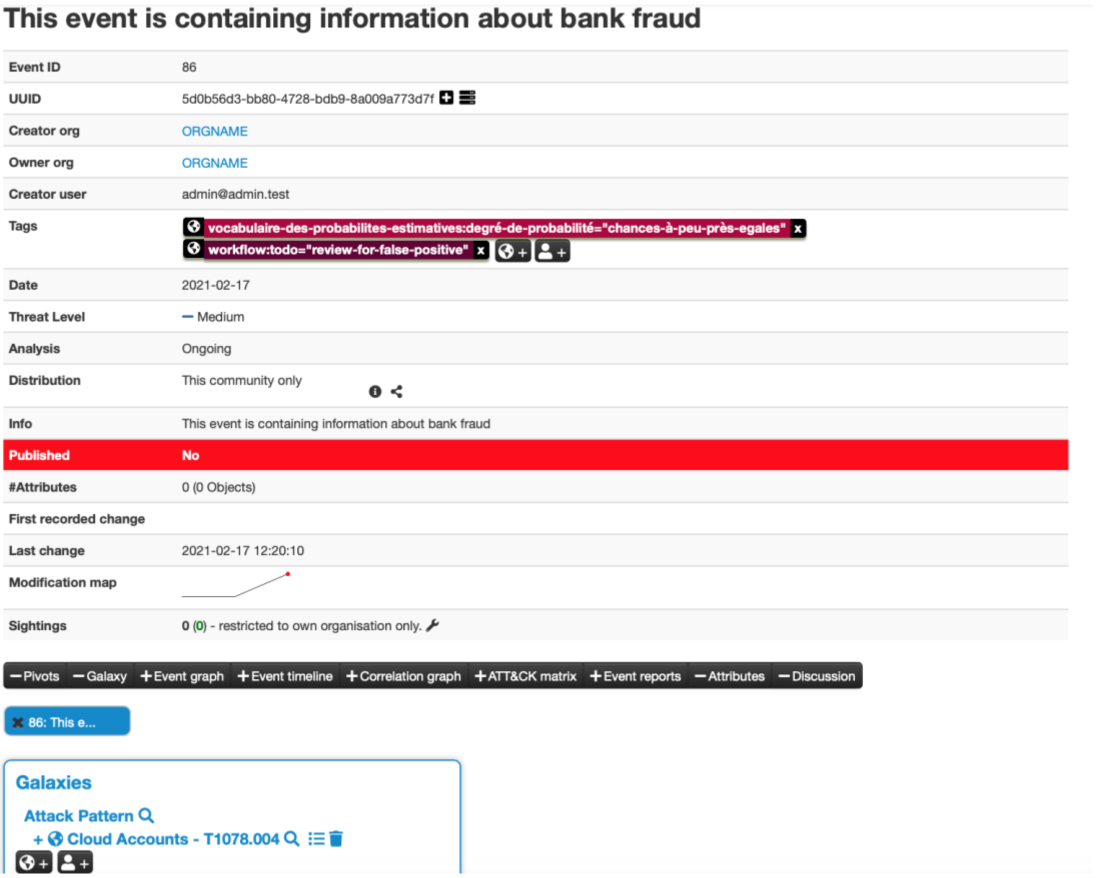
If an organisation wants to start using the MISP platform, the first step is to freely download it, each organisation has its own instance of the platform, but is a part of a larger community with a main MISP (the MISP core). The free community with no access requirements is called the feed community, where all members are allowed to share threat information. The framework is mainly built on a give and take structure, meaning that all information is free, but it is good practice to also share personally obtained information and experience with other members of the community, however, there is no requirement to do so. In other communities like NATO, ‘normal’ members are not allowed to share threat intelligence, only to receive it.

Some organisations have specialised in finding threats and sharing them with the MISP core project, these are in the MISP universe, called ‘feed providers’. They are providing feeds that the average user can subscribe to and get updates on cybersecurity attacks. A large number of feed providers exist, and this combined with the fact that everyone can upload information has a negative consequence [13].

The result is that a large amount of information is shared every day, and the network forensics who manually have to evaluate each incoming threat, deciding whether to block the network traffic or allow it, have a close to impossible task, since there is no mechanism to detect false information or low-quality data, which is very time consuming to evaluate [69].

2.1.1 MISP Events

The MISP platform is heavily based on threat events. Events are encapsulations of contextually related threats such as objects and attributes [14]. Events, as mentioned earlier, are populated with 'attributes', which are the actual attacks that have occurred (IoC), been investigated and lastly published by the free providers. The attributes in the event include the IoC, which is often an IP, Email, domain name, etc. Attributes can be flagged as 'IDS', essentially meaning that the data is so clean it can be processed by machines [14].



This event is containing information about bank fraud

Event ID	86
UUID	5d0b56d3-bb80-4728-bdb9-8a009a773d7f
Creator org	ORGNAME
Owner org	ORGNAME
Creator user	admin@admin.test
Tags	vocabulaire-des-probabilites-estimatives:degré-de-probabilité="chances-à-peu-près-egales" x workflow:todo="review-for-false-positive" x
Date	2021-02-17
Threat Level	— Medium
Analysis	Ongoing
Distribution	This community only
Info	This event is containing information about bank fraud
Published	No
#Attributes	0 (0 Objects)
First recorded change	
Last change	2021-02-17 12:20:10
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Galaxies

Attack Pattern

+ Cloud Accounts - T1078.004

Figure 3 - A threat event in the MISP platform, including several information about the threat [68].

Figure 3 illustrates how a threat event in the MISP platform is structured. The event contains many pieces of information [13], that will benefit the user, such as tags, galaxies, threat level, a report function (user written text), timeline, and much more. When an event is published by a feed provider, the event information is the first data the network forensic will be presented for. It is therefore important that the information creates context and is of high quality.

2.1.2 Essential Event Features

The following paragraph will be based on conversations with an expert in the field, Dennis Rand (part of interviews, presented later in the report); it explains the event features which are the most important and will create the most context to the threat event [13].

Tags

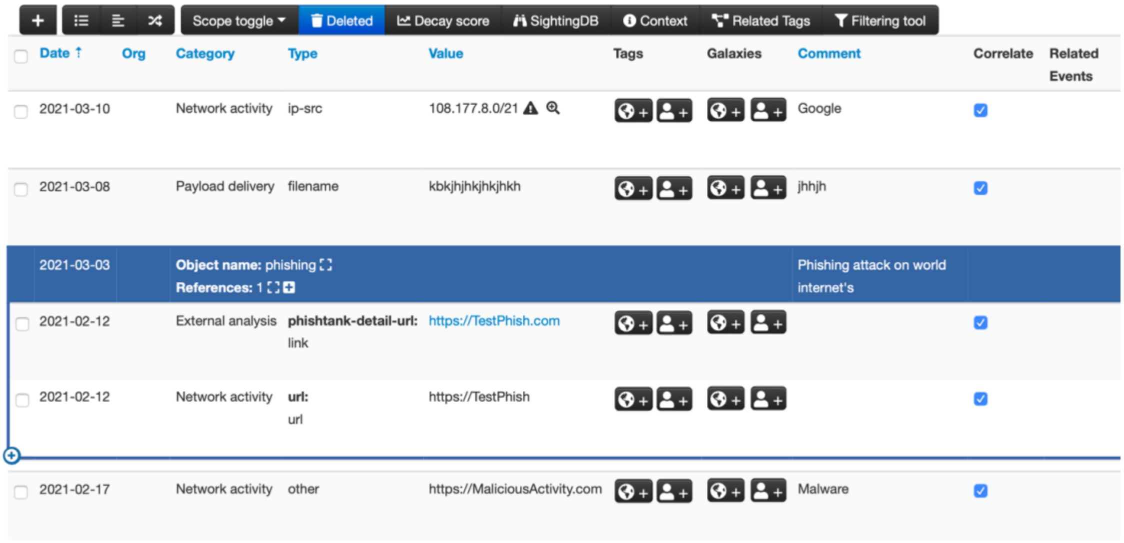
One widely used mechanism in the MISP platform is the utilisation of tags [14]. Tags are considered excellent practice, because it provides more context and information to the event, e.g., a TLP [62] (Traffic Light Protocol) tag explain how the event must be distributed. Because it provides such good context it should always be used and be as precise as possible. The MISP platform contains two different kinds of tags: ‘custom tags’ and ‘taxonomies’. Custom tags are custom-made tags of the private MISP instance; these tags should be used in local environments and should not be used in the global environment. Instead, taxonomies should be. Taxonomy tags are ‘premade’ tags, which describes different behaviours and categories. If possible, taxonomies should always be used and are considered perfect practice, because they describe a very specialised group of events.

Taxonomies in the MISP use a concept called ‘triple tags’, which means a taxonomy must include a namespace, a predicate and a value; the value is optional, separated by a colon. CERT-XLM:abusive-content="spam" is an example of a triple tag structure [14].

Attributes and Objects

Attributes [14], as explained earlier, are the actual threats collected and analysed by security analysts, it is the IoC of the attacks and it highly relevant. They will provide important information about the attacks, such as time, category, tags and explanations. When attributes are provided individually, they do not necessarily provide much logic; for example, an attribute can just be a checksum; this will not provide the network forensic valuable information because no context is provided. To deal with that problem, feed providers should always group the attributes into ‘objects’. Objects allow for advanced combinations of attributes, grouping actual attacks to provide more logic to a threat. For example, grouping a checksum with the IP, MD5, or URL will give more valuable information and context than an individual checksum [14].

Figure 4 illustrates how attributes are listed in an event. It is clearly shown what information is included, such as category, tags and more. An object is also shown; the object consists of the attributes within the blue box. The object also contains further information, such as name and references to other objects.



	Date ↑	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events
□	2021-03-10		Network activity	ip-src	108.177.8.0/21			Google	☑	
□	2021-03-08		Payload delivery	filename	kbkjhjhkjhkjhkh			jhhjh	☑	
	2021-03-03		Object name: phishing References: 1					Phishing attack on world internet's		
□	2021-02-12		External analysis	phishtank-detail-url: link	https://TestPhish.com				☑	
□	2021-02-12		Network activity	url: url	https://TestPhish				☑	
□	2021-02-17		Network activity	other	https://MaliciousActivity.com			Malware	☑	

Figure 4 - The related attributes in the event and object in the blue box

Galaxies

Galaxy Clusters [14] are unique in the way that they are above events in the hierarchy. A galaxy cluster contains many events and is a way of grouping similar events together to create context. Galaxy clusters work very similarly to objects explained earlier in the way of grouping related information together. Using galaxy requires that the feed provider has a solid knowledge about related events to group them. Using galaxies is considered to be very good practice and should always be used if possible [69].

Other Event features

The previously mentioned features are important to provide as much context as possible for the network forensic team. However, the event is also populated with other more general information, which should be filled out before publishing an event. For example, threat level, which is an indicator of how dangerous and urgent the threat is, rated on a scale from 1 to 3. Sighting reports will provide information on when members of the community last spotted the threat. The first/last seen report tells the members when the threat was first and last seen, which can be an indicator of the threat is over or ongoing. Event report, which is a newly added feature which allows for the creator to enter freely written text to an event and can be beneficial when explaining complicated threats. Event extension allows the creator to reference other events if the information is overlapping [14].

These features should always be filled out to an extent where context is so sufficient that a network forensic can make a qualified decision about the threat.

Filtering algorithm

The feed community allows every member to publish data at any given point in time; however, this creates an overload of events and data, and in some cases, events are populated with insufficient data that is not actionable for the network forensics. Some input validation is implemented when creating an event; however, this can easily be worked around. After some research, no evidence was found that MISP has any algorithms implemented to detect and filter out insufficient threat events of low quality.

2.1.3 Message Queue System

Upon installation of the MISP platform, a 'ZeroMQ' (ZMQ) is included [15]. The ZMQ is not a tool developed by the MISP project. The ZMQ provides a message queue for the MISP instance, but it can run without a broker, unlike other message queues. A broker is a software that translates messages between two software instances to 'understand' each other using formal structures.

The ZMQ supports many different communication approaches, such as publish/subscribe, push/pull, client/server, and more. This makes the ZMQ solution very versatile and easy scalable [16].

The ZMQ utilises the concept of sockets. Socket communication is widely used and is often used in server/client communication. When a client wants to communicate with a host, a request is sent; when the request is accepted, a dedicated communication line is established. When the intended communication is finished, the line is shut down, and a new one must be created to further communicate [17]. The ZMQ uses the same concept; however, the communication is asynchronous, meaning that the ZMQ is not bound to timestamps and dedicated connections as in traditional socket communication. The entire process of connection setup, teardown, and reconnect is in ZMQ transparent to the user; it may happen that messages will have to queue because the receiver cannot receive them at the time. In traditional socket communication, the protocol can only form one-to-one, one-to-many or many-to-one; however, ZMQ allows to be connected to many endpoints and still send to many endpoints, which is called many-to-many communication [18].

As mentioned, the MISP project utilises this service, and the ZMQ is located on each MISP instance installed. The ZMQ captures everything that happens in the MISP instance, including incoming events, which are sent as a JSON structure. A benefit of using ZMQ is implementing real-time features to the MISP instance because the ZMQ can intercept everything that happens in the MISP instance [19].

The JSON structure is essential in the MISP instance; everything in the MISP ZMQ is described and formatted in the JSON structure. The JSON format is ideal for data interchange and is widely used because it is scalable and human-readable [20].

Figure 5 shows where the ZMQ is placed in the MISP landscape. It is placed before the private MISP instance but after the MISP core project (Feed community).

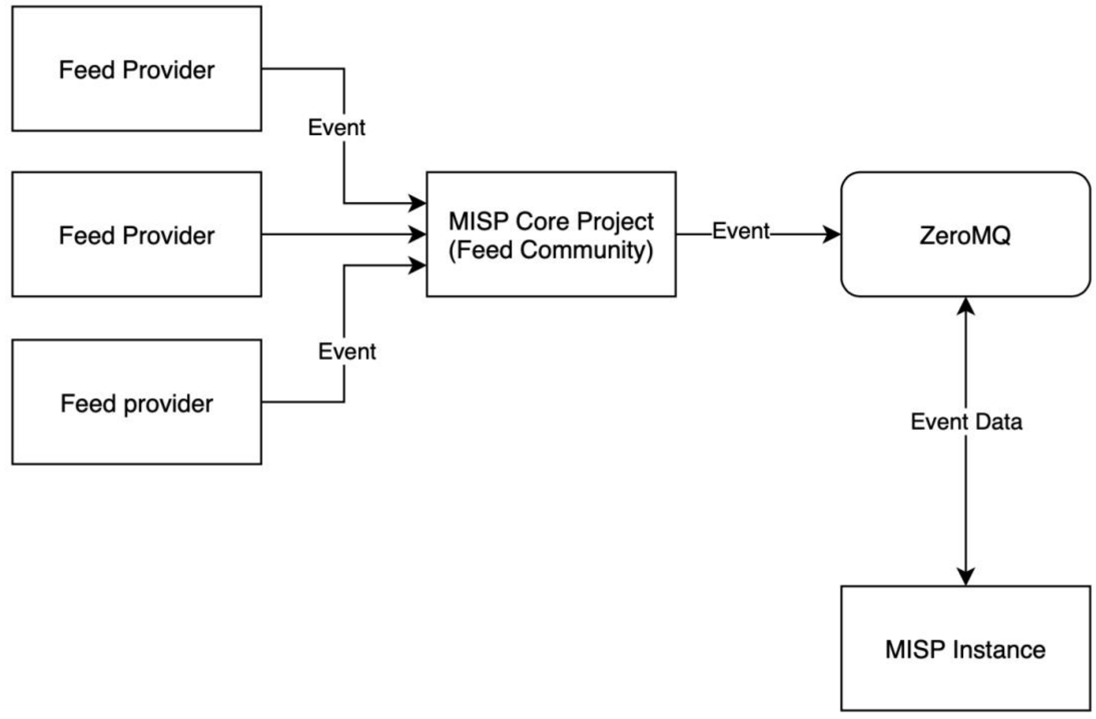


Figure 5 - The private MISP instance setup including the ZMQ.

The feed provider will provide feeds to the core project, which will delegate out the events to the MISP users who are subscribing to those feed providers. When the event is sent to the private MISP instance, it has to go through the ZMQ. The ZMQ will create a queue for all information going to the MISP instance and can be intercepted.

2.1.4 MISP Communities

As mentioned in section 1.1, one significant part of the MISP universe is the MISP communities; the communities are where threat intelligence is shared. The largest community is the feed community. The feed community are based on the feeds provided by, e.g., OSINT and CIRCL and other threat organisations, which is larger organisations that analyses and shares threat intelligence to their members, however, as mentioned ‘regular’ users can also share events.

Many different communities already exist, and some of them are listed on MISP’s webpage; besides those listed, many others exist which are private or hidden. A large number of big actors are using the MISP platform and have created communities where they share detailed threat information. To name a few of those organisations, NATO, CIRCL, CSSA, FIRST, etc. [8].

It is up to the individual community to decide on the rules that must be fulfilled before an organisation can join and get the threat intelligence information. The CIRCLE community consists of over 800 organisations and is free of charge to join. Some of the other communities require an application, however trying to access the application page to investigate what requirements must be fulfilled, they were all unavailable and could not be reached [8].

To request access, a key that utilises PGP (Pretty Good Privacy) must be used; the key is distributed to the user at the point of installation of the MISP instance. The PGP is an encryption scheme and is based on public and private keys; it is often used in email services and is relatively simple. The explanation lies without the scope of this thesis; an article is attached where more information about PGP keys can be read [21].

After getting access to the community using the aforementioned approach, the community data can be fetched. This is done by using the internal API. The API is called PyMISP [70] (further explained in 2.1.5). After getting access to the community, no MISP is actually needed because they provide the user with a web interface to see threat intelligence. The data can then be exported and used in other systems as SIEMS, IDS systems and more [13].

Documentation on how to become a feed provider in the CIRCL community is minimal; CIRCL refers to a page where incidents can be reported. Reporting incidents can be done using email or phone, and because of that, it is assumed that ‘regular’ users cannot create threats events in those specialised communities [22].

2.1.5 PyMISP – API Client

The MISP is a very open and transparent platform and allows its members to see new features and comment on them before releases. One approach that the MISP platform is achieving this openness through is integrating solutions that allow members to quickly access data. MISP even has a library implemented, called PyMISP [70], and the library is coded in Python.

The library, which is a REST API, includes many different essential features, which help users better integrate with other systems. The library is very intuitive to use because it allows includes a guide and tutorials on how to use the API.

The library includes many different examples of code, and to mention a few; it can update events, add tags, remove tags, search in texts, and much more [23]. To utilise the API, an API key must be implemented; the API key is given to each MISP admin in the platform.

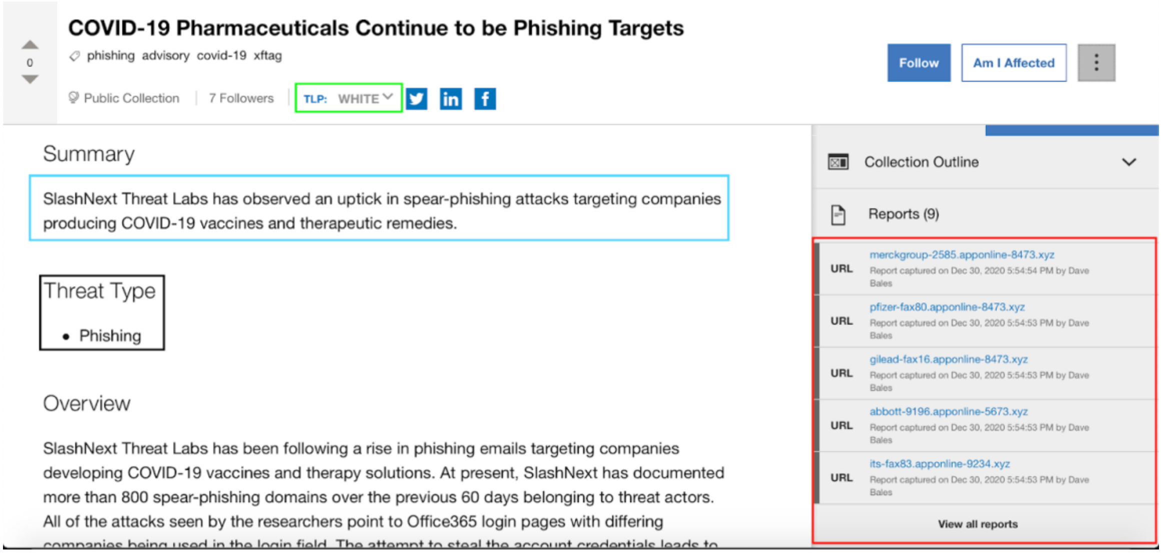
2.2 Similar services

This subsection will investigate the market for similar platforms; this is done to gather inspiration from other sources providing the same (or similar) services. This will enhance the understanding of the concept and determine what has already been tried in the market and where the innovation could be made.

This section aims to find the three most used similar services and investigate their functionality, setup, and algorithms. Depending on the source, a lot of different cyber threat sharing platforms are recommended, and it was not possible to find any academic work on the matter, so therefore the recommendation of [24] was followed. The article suggests ten different platforms, and the two first selected were IBM X-Force and Anomali. These were selected because they were the two topmost recommended platforms. The last selected candidate for analysis is a platform called AlienVault, because, in many ways, it resembles the MISP platform.

2.2.1 IBM X-Force Exchange

IBM X-Force [25] was established in 1996 as a part of an Internet Security System, an independent company, but IBM acquired it in 2006; X-Force is today a part of the IBM enterprise. IBM X-Force is a collaborative threat intelligence platform, whose primary purpose is to help security analysts find threat indicators (IoCs). IBM X-Force is a widely used TIP and their website states: *"X-Force Exchange hosts six of the world's top 10 retailers, and five of the world's top 10 banks." [25]*. This is considered a good indicator that this is a product worth investigating.



COVID-19 Pharmaceuticals Continue to be Phishing Targets

phishing advisory covid-19 xftag

Public Collection | 7 Followers | TLP: WHITE

Summary

SlashNext Threat Labs has observed an uptick in spear-phishing attacks targeting companies producing COVID-19 vaccines and therapeutic remedies.

Threat Type

- Phishing

Overview

SlashNext Threat Labs has been following a rise in phishing emails targeting companies developing COVID-19 vaccines and therapy solutions. At present, SlashNext has documented more than 800 spear-phishing domains over the previous 60 days belonging to threat actors. All of the attacks seen by the researchers point to Office365 login pages with differing companies being used in the login field. The attempt to steal the account credentials leads to

Collection Outline

Reports (9)

URL merckgroup-2585.apponline-8473.xyz
Report captured on Dec 30, 2020 5:54:54 PM by Dave Bales

URL pfizer-fax80.apponline-8473.xyz
Report captured on Dec 30, 2020 5:54:53 PM by Dave Bales

URL gilead-fax16.apponline-8473.xyz
Report captured on Dec 30, 2020 5:54:53 PM by Dave Bales

URL abbott-9196.apponline-5673.xyz
Report captured on Dec 30, 2020 5:54:53 PM by Dave Bales

URL its-fax83.apponline-9234.xyz
Report captured on Dec 30, 2020 5:54:53 PM by Dave Bales

Figure 6 - Setup of a threat event in IBM X-Force [26].

To further investigate the software, it was decided to sign up and get a hands-on experience with the product. The product is free and browser-based, which makes it accessible and easy to start using, however, the acquired version is only a demo, so it is very limited.

Figure 6 shows what is equivalent to a threat event in the MISP platform, a threat report. The report includes some of the same data as a MISP event. The green box in the figure above shows the tagging system, which is very simplified, and the user cannot set local tags. The blue box highlights a summary of the incident. The black box shows the type of threat, and in correlation to the type, lastly all of the IoCs are shown in the red box. In this example, the IoCs are just URLs and include the exact time they were captured. This does not provide much valuable information to the person evaluating the threats [26].

The system is not open-source, and some of the features require payment; one of them is called ‘early warning’ [71], which is a feature where the consumer can get an early warning about ongoing attacks, which is a valuable feature so consumers can quickly update firewalls, SIEMs, or similar systems.

IBM X-Force also allows integration into private systems in terms of API protocols; this enables the consumers to fetch data from the framework to use in their private systems. However, there are bounds to the API; a consumer can only fetch so much information from X-Force, the exact amount is not specified on their web page. Exporting threats is also a possibility; however, the only available format is STIX and STIX 2, which can be a limiting factor when making private integrations [27].

X-Force is also offering a service called “IBM Advanced Threat Protection Feed”, which is a billed service; the service can transform the reports of the threats into machine-readable reports. This makes it possible for companies to pay for the service and then automatically update firewall rules, SIEM systems, and intrusive prevention systems. This led to the same questions as in the MISP platform, how good is the data quality of the incoming threat reports, and can it be trusted?

After a thorough investigation, it was not possible to find if the incoming threat reports are being analysed in terms of data quality. IBM only allows non-users to see threats, and therefore the publishing of events was not tried. An assumption is that regular users of the service are not allowed to publish events, meaning only IBMs own analysts can do so, and in that way, make sure that the data is of high quality [26].

2.2.2 Anomali ThreatStream

Anomali is a threat intelligence-driven company, it was established in 2013 [72], and it has since risen to be a major actor on the market of threat intelligence. Anomali was found worthy of investigating because they state that they provide their service to more than 350 global organisations, and it was also recommended by the previously mentioned article [24]. Anomali claims that amongst the 350 organisations, some of them are in the global fortune 200 and 500, and they also provide threat information to governmental defence organisations [28].

Anomali offers different services, but what is most relevant is their product called ‘Threatstream’. Threatstream is a TIP, where companies can buy access to the threat reports and threat information.

Trying out their product is not a possibility since it is a paid service; however, it was decided to reach out to them by email requesting a demo version of their product. The answer, which also can be seen in full in appendix 10.2.1, was a denial of the request because of high pressure, and the fact that representing the demo is too costly for them, because they have a special team doing the presentation. Therefore, the only documentation available is what has been published on their webpage. To get more explicit information about the service, a follow up email was sent. The email explained the initial solution (section 1.6) and asked Anomali whether they use any algorithms to measure the data quality. A team leader answered: ”*Anomali uses a tool called Macula to perform this task, The results enable users to set parameters around which score variable they would like to focus on, for example everything above a 95% score on severity and confidence”* [Appendix 10.2]. After a further investigation of Macula, it was decided to stop the research of that tool because it has no documentation, so it must be assumed it is an in-house development, however, it is assumed that Macula is an AI tool [29].

2.2.3 AlienVault OTX

AlienVault USM is a part of AT&T cybersecurity, which is a company specialised in helping companies and organisations protect digital assets. They are offering an extended range of services spanning from security compliance to penetration testing [30]. AlienVault is AT&T’s take on a TIP, and they are offering a service where visitors of their web page can try out their service in a controlled environment with demo information. As mentioned, AlienVault is the TIP of the three mentioned in this work that resembles the MISP platform the most, all users are allowed to publish, and the users can subscribe to specified threats of interest.

After getting access to the demo version, the user is presented with a dashboard, where different visuals are presented [31]. In figure 7, various pieces of information are located on the dashboard. The user can see alarms the current day and the current week, what methods of malicious activity are used and what kinds of attacks occurred on a given day.

Besides the dashboard, the user can also navigate to the event page. The event page is where all threat reports are gathered, and users can see the incoming threat along with the specified IoC; this is illustrated in Figure 8.

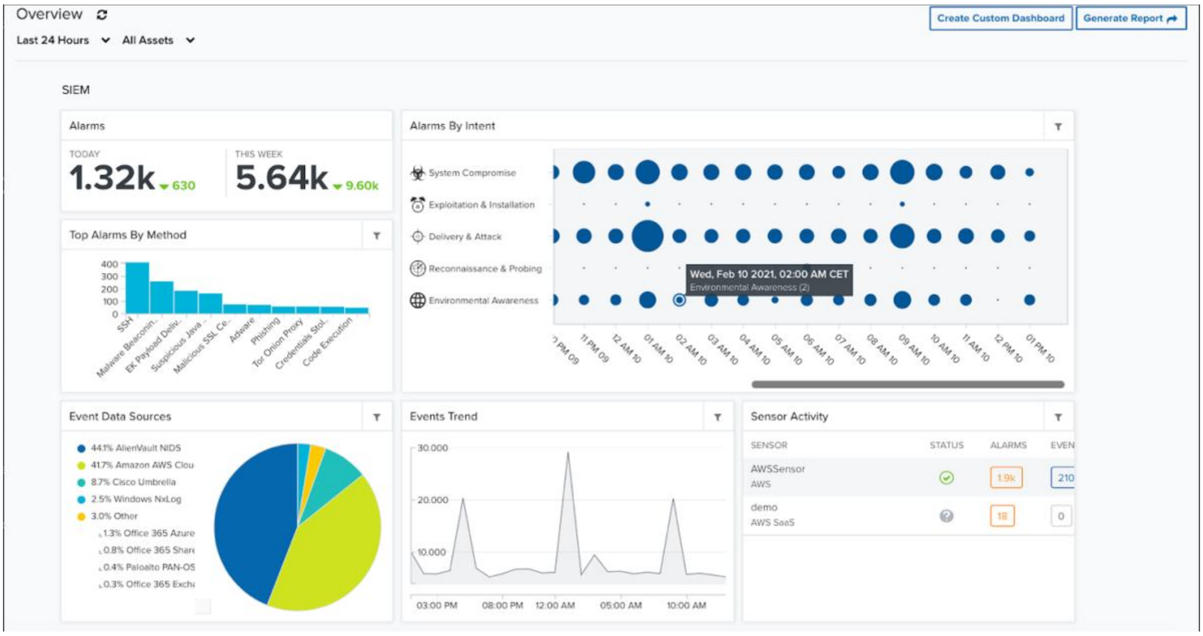
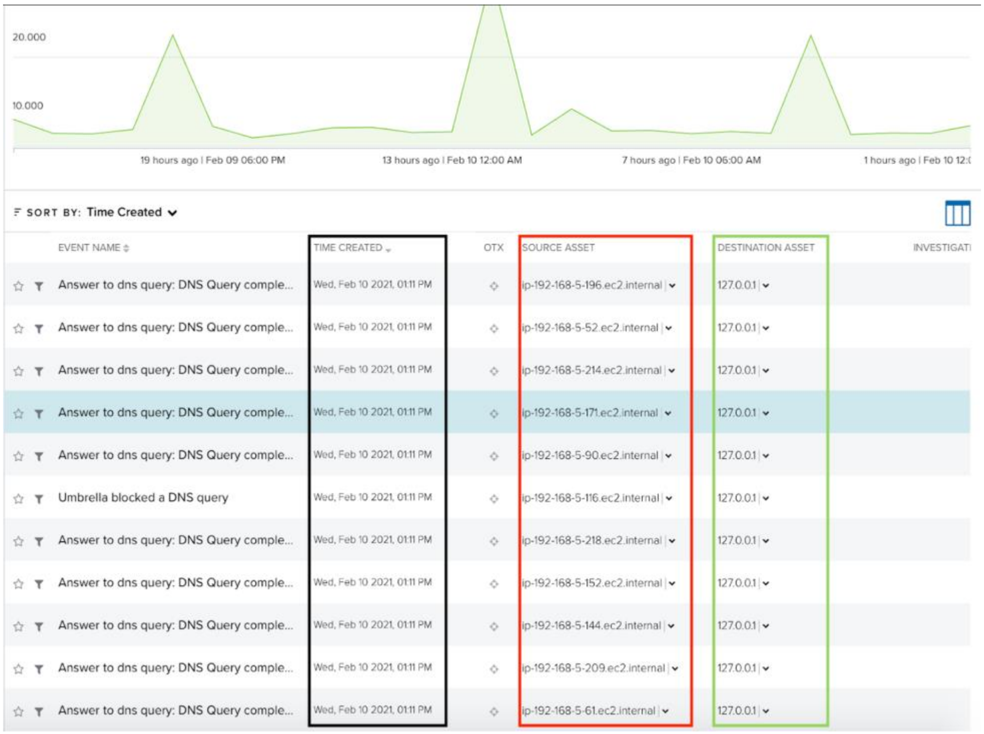


Figure 7 - The main dashboard of the AlienVault USM [31]

Figure 8, as mentioned shows the event page. In the figure, the source assets (red box) inform where the attack originates from – the IoC. The destination location (green box) is also shown; however, since it is a demo version, all the IPs are local. Furthermore, the timestamps are shown (black box). The events being shown on the event page are coming from the AlienVault Open Threat Exchange (OTX), which is a framework that makes it possible for all the users of AlienVault to upload threat reports, and in that way, share attacks just like the MISP platform [31].

Figure 9 shows the screen where users can upload attack incidents to share with the AlienVault community. When creating an indicator, input validation is implemented, it tells the user that the URL entered is on a whitelist created by AlienVault and should be re-validated, meaning that it is possibly a false positive.

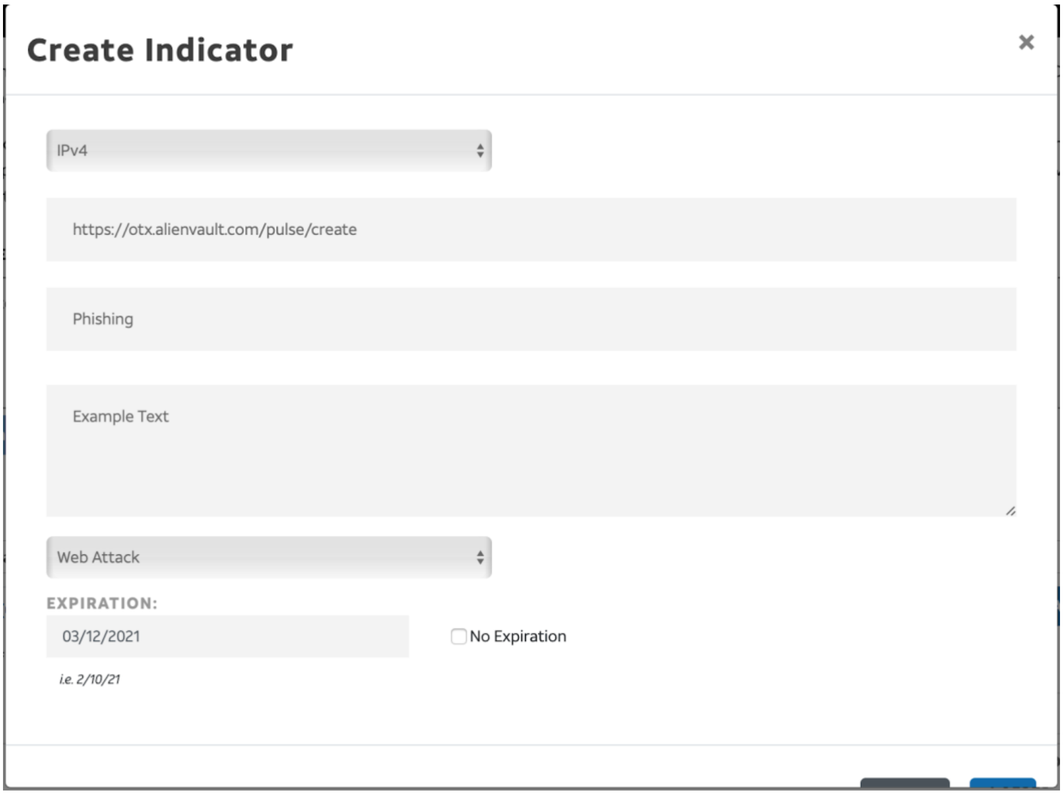
After a thorough investigation of their service and web page, looking through a vast amount of threat reports, no specific information about any data quality algorithms is mentioned and investigating the report, there is no evidence that such an algorithm should exist.


20.000
10.000
19 hours ago | Feb 09 06:00 PM
13 hours ago | Feb 10 12:00 AM
7 hours ago | Feb 10 06:00 AM
1 hours ago | Feb 10 12:(

SORT BY: Time Created

EVENT NAME	TIME CREATED	OTX	SOURCE ASSET	DESTINATION ASSET	INVESTIGATI
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-196.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-52.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-214.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-171.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-90.ec2.internal	127.0.0.1	
Umbrella blocked a DNS query	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-116.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-218.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-152.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-144.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-209.ec2.internal	127.0.0.1	
Answer to dns query: DNS Query comple...	Wed, Feb 10 2021, 01:11 PM		ip-192-168-5-61.ec2.internal	127.0.0.1	

Figure 8 - Event page, including threats and corresponding IoC [31].



Create Indicator

IPv4

https://otx.alienvault.com/pulse/create

Phishing

Example Text

Web Attack

EXPIRATION:

03/12/2021

□ No Expiration

i.e. 2/10/21

Figure 9 - The incident publishing page [31].

Sub Conclusion

After researching these organisations that provide threat intelligence information, it was quickly discovered that all three of the services were difficult to find detailed information about, such as algorithms, demos, or general information. The AlienVault service is the one that looks the most like the MISP platform, and a demo of that service was tried, which gave little information about the functionalities, but nothing about the algorithms. Looking at the data in the service, no evidence was found that a data quality algorithm was used to filter away poor events. The service allowed, just like the MISP platform, all their members to publish information, which again created an overload of data, and it seems that the platform has the same flaws as the MISP platform.

The two other services, Anomali and IBM X-Force are paid services, and only a simple demo were allowed to be evaluated, therefore the available information is scarce. It didn't seem like 'regular' paying members were allowed to publish data to the community, it was only their own researchers who were allowed to do so. Therefore, it is assumed that because publishing is restricted to researchers, the data quality would be higher, however this is only an assumption. IBM X-Force had no evidence of algorithm to filter poor data. Anomaly however reached out via email and said that they use a machine learning algorithm called 'Macula', however after researching it, no further information was found.

2.3 Research of Central Concepts

To get a deeper understanding of the important topics and concepts presented in this stage of the report, those will be investigated. Furthermore, the concepts will be investigated so that they can be used later in the project to create an architecture of a potential system and create a PoC.

2.3.1 Data Quality

One central and recurring concept in the report will be '*data quality*' (DQ); this paragraph will research that concept, find out how DQ is defined, and what factors enhances or lower data quality.

Data is mentioned in the Introduction of the report (section 1.1). It is a critical asset in many organisations and is considered very valuable; therefore, having high quality data is essential [73]. Many organisational decisions are also based on the data, and low-quality data can lead to wrong decision making. Clean data of high-quality produce other beneficial side effects, such as easily scalable solutions, expansion of systems and more. The term data quality is extensive and used in many different situations; therefore, creating a fixed definition of the concept can be tricky. Many studies have been conducted on that ground, attempting to define and refine it.

As already discussed, it is a well-studied field, and one thing that a majority of the studies can agree upon is that DQ is highly dependent on the user and usage of the data. This means that data must be shaped and cleaned for the specific requirement of usage. As mentioned, many attempts have been made to settle on a definition, and because of that and the complexity of the concept, this project will settle on the definition “fitness to use” [32]. The definition is very broad; however, it makes sense that it is the end user's requirements that determine if the data quality is good or bad, data quality might be structured and formatted correctly, but if it is not what the user is seeking, then it can be argued that the data is not applicable.

The concept of data quality includes what is in many studies called ‘the dimensions of data quality’ (DQD). It is worded and shaped in different ways depending on the study; nevertheless, the meaning and concept follow the same principles. Many studies have already been published researching the matter of DQD, in all of the studies, the dimensions, also called ‘cornerstones’, are almost identical. The dimensions represent metrics upon which the data quality could and should be measured. It has been a broad understanding for a long time that data quality should primarily be based on reliability; however, in modern studies it is acknowledged that the data quality must be measured upon all dimensions if possible.

Availability	• Accessibility • Timeliness • Authorisation
Usability	• Documentation • Credibility • Metadata
Reliability	• Integrity • Completeness • Accuracy • Consistency
Relevance	• Fitness
Presentation	• Readability • Structure

Figure 10 - Dimensions of data quality [32].

Figure 10, shows the dimensions of data quality, and in each dimension, a list of related terms that could and should be used to assess and evaluate data quality is attached. It is important to notice that data quality assessment should not be isolated to one dimension but should be seen as a whole [32]. The dimensions are coming from [32], however, it is worded differently in the figure, because of readability, however, the principle remain the same.

In [32], three different groups related to the data quality is defined.

- **Data Producers**, these are the ones that generate the data and populating it, this group is not restricted to humans, in modern societies smart devices collect enormous amount of data.
- **Data Custodians**, these are people who manage, store and process the data.
- **Data Consumers**, the people who use the data.

Each group plays a vital role in data quality life cycle. The definition 'fitness to use' is important here. It is widely accepted that it is the data consumers are the ones who decide whether the data is fit for usage, and inevitably decide whether the data quality is high or low.

Data Quality can be insufficient for several reasons, and many challenges regarding data quality exists. One of the main reasons data quality can be lacking is the fact that in some cases humans are interacting with the data. However, some of the other challenges today are the large number of different formats the data can be structured in, which often do not correspond to each other, and therefore, the data must be transferred to new formats to be interpreted. Another challenge is the amount of data on the internet, evaluating the incoming data can be difficult, much of the data is unstructured, and it takes many resources to clean, store, and obtain the data [33].

2.3.2 Reputation Systems and Trust

Transaction between online parties happen all the time, and some data exchange protocols do not take trust and reputation into consideration – For example, in an email service, the 'From' field is just some arbitrary email address, where no source of *true* identity is attached. In today's online societies and communities, many attempts have been made to create reputation and trust systems. The systems are great tools to moderate and control interactions and, in that way, enhance the quality of services. The primary purpose is, as stated in [43]: *"The purpose of trust and reputation systems is to strengthen the quality of markets and communities by providing an incentive for good behaviour and quality services, and by sanctioning bad behaviour and low-quality services"*.

The first concept to define is reputation systems. Reputation can be defined as *'to know someone or something because of what people say about them, rather than by having direct experience of them'* [35]. Reputation is often associated with a score, ranking or rating of some kind. This score will often be based on collective feedback from previous users of the entity. Reputation models are heavily used in E-commerce, such as Amazon or eBay.

The term trust is often closely related to reputation. The definition of trust according to Cambridge Dictionary [34] is *'to have confidence in something'*, meaning that an entity relies on something or someone even though there is no actual proof that the entity is trustworthy.

In the concept of trust, a few different information sources are defined; these are *'Direct experiences', 'Witness information'*, *'Sociological information',* and *'Prejudice'*. These different types of information are important to have in mind when creating a reputation and trust system because from where data originates is important to define, since this will have a significant impact on the trust [36].

The first kind of information is the most reliable information source, and it is quite self-explanatory. It is direct experience with a subject; however, this information source is divided into two, the first one is where experiences are directly obtained, and the second is direct observation of a group or community. The first one is mainly used and gives the most precise result since the second must be in a controlled environment and this can produce noise to the outcome.

The second information source is the most relevant in threat intelligence; it is the witness information, also called word-of-mouth. This type of information is based on experiences of others. This kind of information is often abundant and is more based on a quantitative approach. The witness information can be complex to use in trust and reputation models since the observations can sometimes suffer from manipulation of experiences.

The last kind two kinds of information sources: sociological information and prejudice are not relevant for this work and will not be considered further in the report [36].

Reputation systems are increasingly becoming very important to today's society. Reputation system is a mechanism to create collective trust between entities, interacting in an environment where information about the other party is insufficient; this can both be online and physical. Reputation systems are generally made for the party, which is buying a product or service since, in reality that party does not know what is received. On the opposite side, the seller knows that he is paid with some currency. The idea behind reputation system is to let users of online platforms rate each either by using score or reviews; this let the buying party see the reputation of the selling party by hearing thirds party's experiences. The examples are stated as economic examples, but these can be directly translated to exchange of data as well [55].

Discussing reputation systems, two different architectures of reputations systems exist according to [54]: centralised reputation systems and distributed reputation systems. The first mentioned is where information from user with direct experience about performance of a participant is collected, this information is shared publicly, so other users can decide whether to transact with the participant or not, after each transaction the rating of the performance is updated.

The latter mentioned is distributed reputation systems, where no central agent is collecting the score of reputation, instead there are opinion stores where users can observe other people experiences and it is up to the relying party to find these stores or obtain ratings from elsewhere from third parties.

Collaborative Sanctioning

A concept that was found interesting in the study of reputation systems, is collaborative sanctioning (CS) [54]. The act of having a community sanctioning a service provider for delivering a bad service.

If the quote in the beginning of the sub chapter is revisited, it was learnt that reputation systems are about giving an incentive to good data and services, and sanctioning the opposite, and the concept of collaborative sanctioning is good for exactly that. The concept is meant for an entire community to sanctioning poor services or data, by stop using it if their rating is poor, and the rating the done by the reputation system. Collaborative sanction and collaborative filtering (CF) are the same category but is not the same. CF is often used in streaming platforms to recommend items to users with similar attributes and can be different from user to user. CS is the complete opposite, it represents the opinion of a whole community, and in that way sanctioning or promoting an item.

Challenges of Reputation Systems

Creating reputation systems introduces some challenges. One main challenge causes many companies to suffer is referred to as incorrect reputations. One big actor in the reputation system market is TrustPilot. TrustPilot is an online platform where people can rate companies and give them stars based on how well the individual person thinks the interaction went. The reviewer can give from one star, which is the lowest up to five stars, for a great experience [44]. There are no requirements when creating the review, and it is entirely up to the individual to do the review.

One challenge that a platform like Trustpilot is experience is manipulated or untrue review, which can, consequently, hurt companies and their sales. To solve this problem, Trustpilot has implemented a solution, where the selling company can require a reference number from the reviewer to prove that the user has been a buying customer [45].

Another known challenge is the ‘cold start’ problem; the cold start problem is often introduced in relation to recommender systems, especially in collaborative filtering. The problem of a cold start is that the system cannot make any recommendations or give suggestions when there is little or no information about a user. The problem often occurs when a new user enters the system or when a new system is released [46]. In the study [46], different strategies to solve the cold start problem is introduced. The two most relevant strategies will be introduced here. ‘*Random Strategy’*. This strategy takes random elements and presents them to a user, who is asked to rate the suggestions and based on that, the system learns more about the users and their preferences. This strategy requires effort from the users. *‘Popularity Strategy’*. This is a strategy where ‘popular’ in the community are shared with the user, this strategy requires minimum users' involvement. The main problem of this strategy is the inequality of the elements in the database.

Data Quality and Reputation Systems

Data Quality assessment in regard to trust creation in TI is not a widely studied subject, however, it must not be neglected. A study *'Towards Improved Trust in Threat Intelligence Sharing using Blockchain and Trusted Computing' [37]* does try to shed some light on the topic, using the technology of 'Blockchain'. The study does not present any concrete solutions to solve the problem of low data quality, but some recommendations to improve the quality are mentioned, and these are: inform users by education, reputation systems, and an automated process of evaluation. The paper proposes a system, which can assess and evaluate TI sharing. The system will also create a reputation system, where users can rate TIs. The system's output will be a trust score and reputation score. The system is called 'TITAN' and as mentioned, it is built on a blockchain approach. The blockchain is not relevant for this project and is not explained further. However, the quality score and reputation score have high relevance. The quality score is not explained in detail but only TIQScore=f(Completeness, Freshness, Relevance). Where completeness score decreases for each missing field in the data. Freshness is a decay function where the score will decrease in proportion to how long time since the threat was uploaded. The last variable is relevance, which is a custom list with keywords, and if a threat includes this keyword a certain amount of points is given [37].

The previously mentioned study also presents an exciting concept referred to as *'Quality of Indicators' (QoI) [39]*. This term is introduced in 'Beyond Free Riding: Quality of Indicators for Assessing Participation in Information Sharing for Threat Intelligence', and is mentioned in the context of avoiding free-riders in TI. The term QoI covers four concepts, correctness, relevance, utility, and uniqueness.

These are just like in data quality, dimensions which should be accessed to find the actual quality. In the study [39], the authors are preparing what is referred to as a 'golden dataset'. The golden dataset is a set of data, which is considered the most optimal data set. They are using this data set to compare new incoming data sets to see how far the newly arrived dataset is from the golden one. Using this approach, the authors can see the differences. The first dimension is correctness; the dimension will indicate if the data is correct in terms of what is expected, is the data labelled correctly. The second dimension is relevance, is the data relevant to the community and is it contextually correct: The next dimension is utility. The utility is how relevant is it, but at a finer level, does this information tell more about the threat than other indicators? The last dimension is uniqueness, is the data seen before, or is it new in the community [39]?

2.3.3 Protective EU – Horizon 2020

This project will investigate particular two deliverables created in the Horizon 2020 [59], namely 5.1, Threat Intelligence Sharing: State of the Art and Requirements, and 5.4 Threat Intelligence. These two deliverables are the most relevant compared to this project's overall goal, and some of the problems are also experienced and to some extent solved in the Horizon 2020.

Deliverable 5.1: Threat Intelligence Sharing: State of the Art and Requirements

The 5.1 deliverable [41] as the name implies, is investigating the fairly new topic of threat intelligence sharing, the definition, the importance, and approach to do it. The study starts out by defining important terms and approaches of how to treat incoming data. The paper starts out by explaining three requirements threat data must fulfil to be classified as threat data, it must be **relevant, actionable** and **valuable**, this essentially means that the data must have been analysed or processed so it is relevant for the receiver of the data. Furthermore, the data must be actionable, this means that the data must be presented with such context that the data receiver can prompt an action or responds. Lastly the must create value for the organisation using the data.

The paper [41] differentiates between two central definitions, **Threat Data** and **Threat Information**. Both of the types carry information, however, the two must not be confused. The former refers to a low-level type of data, this means data that comes directly from the source without further processing, this is often data coming directly from sensors or network probes. The latter type refers to a type of data that has undergone analysing, this can be data from feed providers. The paper continues to provide interesting points of why threat intelligence is of such high importance, which will not be described further. The study starts to reflect on sharing concepts and factors, and state: *"Trust is important in TI sharing. On one hand, it can be related to the transmission of TI, e.g. assurance that TI is sent from and received by appropriate sender and recipient, respectively. However, it can also be relate to confidence in the TI quality sent" [41]*. The trust can be created by measuring different data entities, to ensure different level of confidence, this would create a confidence level to the sender of the data, based on the earlier records (reputation).

The challenges of TI sharing are also outlined, and here the emphasis is put on the human-interaction part. It is often humans who create TI, and this often create barriers, national and culture differences. Another barrier is that if an TI analysist get a threat which is unknown, the investigation will take much longer than if the threat was dealt with by a more experienced analyst. These challenges are only a few of the mentioned. As stated, earlier trust is major part of TI sharing, and the paper proposes five different trust model, created by NIST [41].

Validated trust: When one organisation obtains a body of evidence of another organisation (e.g., their security policies). The two organisations then know what level of security they each have.

Historical trust: Where one organisation is obtaining historical data about another; the data could be security decisions and activities.

Mediated trust: Where two organisations create trust based on assurances from a third party. The third-party could be an organisation one of the two organisations have used before and therefore could rely on the third party.

Mandated trust: When one organisation establishes trust with another mandated through rules, regulations, etc., from the third company in a position of authority.

Hybrid trust: Which simply means that some company has enough resources to implement more of the trust models and create a stronger trust with each other.

The last relevant concept investigated in the deliverable is TI sharing Quality, it is very briefly discussed, nonetheless high essential to this project. The paper emphasis that the feed quality must be of high quality, because critical decisions are based on it. Human operators are analysing each threat event and the number of feeds is growing, making the human operators a bottleneck. The automation of evaluating feeds, is the task that must be solved; however, the task is highly depended to the question of establishing enough trust between the actors to allow for fully automated integrated systems [41].

Deliverable 5.4: Threat Intelligence Community v3

Where deliverable 5.1, settled on a range of definitions, and pointing out the importance of threat sharing, deliverable 5.4, takes a more practical approach proposing an actual system to solve the problem of dealing with threat sharing, threat alerts, data quality and trust.

Deliverable 5.4 [40] starts out by introducing a long list of different modules and frameworks which is utilised in the system. The system proposed is very comprehensive and includes many technologies to support the task. What is most relevant to investigate is a figure presented (Figure) [40][41].

The model describes an algorithm to which is to determine the severity of the alert. This is done by investigate three independent factors:

Quality: The quality of the alert is determined by looking at the data quality parameter, namely completeness. Each field that is not filled, will reduce the completeness and provide a lower score, the completeness score is also combined with a decay function. The function calculates the relevance of the feed, meaning the freshness of the alert.

Certainty: This factor describes a certainty rating, the factors will count hits on a predefined list with keywords, the more hits the better, and the relevance will become higher, because the system is more *certain* that the feed is of the right category.

Source Trustworthiness: The last factor describes the trust to the data providing organisation, this is based on an average from previous alerts (reputation).

Score: This is the calculated score of the algorithm, this score will be stored in the reputation system.

The algorithm presented is highly relevant to this project, since it present one approach to solve the problem about trust and reputation. The model however does not really look into the specific piece of information delivered by a data provider, the algorithm looks more on the big picture, completeness, time and keywords.

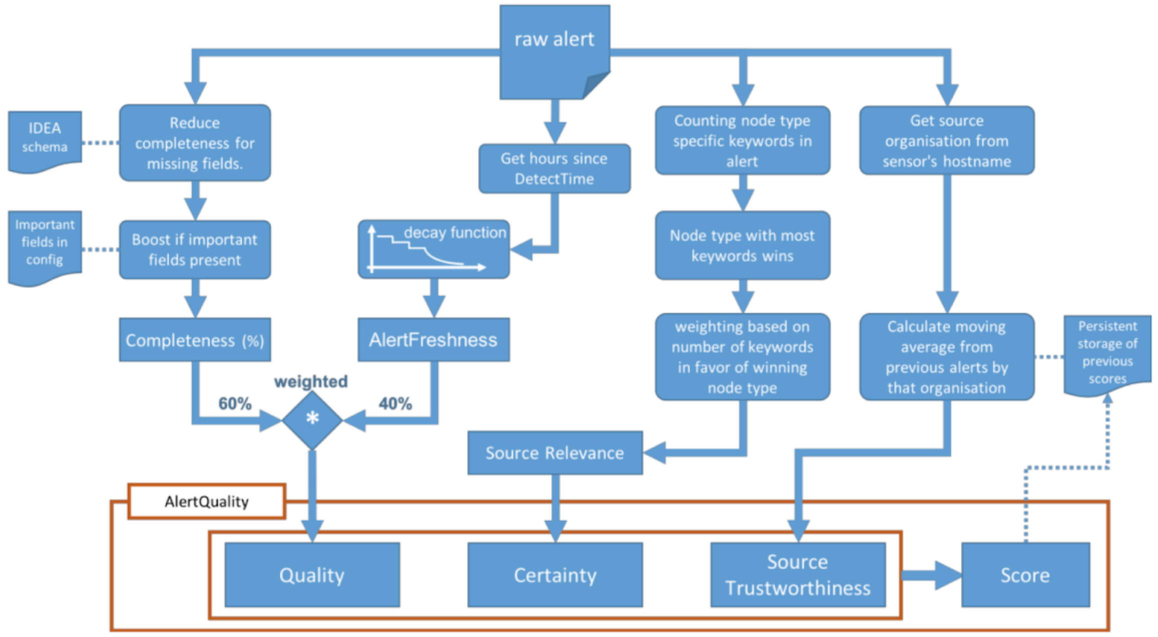

raw alert
IDEA schema
Reduce completeness for missing fields.
Get hours since DetectTime
Counting node type specific keywords in alert
Get source organisation from sensor's hostname
Important fields in config
Boost if important fields present
decay function
Node type with most keywords wins
Completeness (%)
AlertFreshness
weighting based on number of keywords in favor of winning node type
Calculate moving average from previous alerts by that organisation
Persistent storage of previous scores
weighted
60%
*
40%
Source Relevance
AlertQuality
Quality
Certainty
Source Trustworthiness
Score

Figure 11 - The final algorithm to measure data quality [40].

The rest of the deliverable explains more about other technologies, like authorisation and authentication, which is out of the scope of this project. Furthermore, it explains the actual implementation.

2.3.4 Metadata and Tagging

Metadata and tagging [30] are a vast and growing topic. In recent years, with growing amounts of data on the internet, this combined with new effective ways of handling structured data, metadata, and tagging are very effective, especially in data systems like recommender systems and similar. One study defined it as a way of organising things you do not have time to organise.

Metadata is in short terms data about data, it is a piece of information that describes the content of the primary data. It is often used in libraries or services like Netflix, Amazon Prime, or Facebook. Movies and books are often in digital services described with metadata such as genre: 'action', 'thriller', 'comedy, etc. The user can then just search for a term, like action and all items with metadata corresponding to that genre will be presented. Metadata is very well suited for automated processes because the machine is capable of 'understanding' the metadata and acting upon it.

It is a well-studied topic, and standards have been created in the field. One popular standard is the 'Dublin Core'. The Dublin Core includes 15 optional data elements, but these give a good overview of a data entity. The elements are title, creator, subject, description, publisher, contributor, date, type, format, identifier, source, language, relation, coverage, and rights [30].

In many online sources, metadata, tags, metadata tagging, etc., are used interchangeably, this report the focus will be on 'tags', and the definition will be close to the one described earlier as metadata. Tags will describe something about the context of the information; this can later be used in automated processes, look-ups, sorting and more.

2.3.5 Weighted Average Model

This brief section will introduce an approach of calculating the confidence score. The first approach which is investigated is using traditional mathematical formulars.

One widely used and well documented method is of calculating models is 'weighted average'.

The weighted average [47] is a model where the degree of importance is also accounted for, by dividing the number with a predetermined weight. The formular is a relatively easy model to comprehend and is not very complex, however, weighted average is used in many applications, such as stock trading, banking, accounting, healthcare. The model is often applied if the data model includes many variables, and those have different importance. One example where weighted average is often used is in statistical analysis, for example if a questionnaire has been conducted, and different age groups has been asked to answer questions, but for some reason one age group is very under-presented, giving this group a weight can provide more equality to the questionnaire.

It is clearly shown in the formula that, the model is almost identical to a normal average model, however, each number is multiplied with a weight, to balance the importance.

$$x = \frac{w_1 * x_1 + w_2 * x_2 + w_3 * x_3 \dots w_n * x_n}{w_1 + w_2 + w_3 \dots w_n}$$

2.3.6 Machine Learning Algorithms

Machine learning is in recent years becoming very popular because of it is very efficient pattern-seeking functionalities, which makes them highly suitable for analysis of data models. However, there are limiting factors to machine learning, which is introduced later on in the section. Because of those limiting factors, the hypothesis is that machine learning can be challenging to apply to the project.

Machine learning (ML) in today's world of digitalisation is a popular topic, and the technology is capable of processing many complicated calculations and seeing patterns not visible to the human eye. The topic of machine learning is large, and therefore it was decided not to research it all starting from one end to the other but instead seek help from a machine learning expert. The expert questioned is Ming Shen, an associate professor at AAU, who teaches machine learning courses at the master's level.

After explaining how a potential system would look like, Ming Shen was quick to propose 'anomaly detection', which is technique to identify statistical outliers compared to other data points. Therefore, the investigation of ML would start in the area of anomaly detection.

Anomaly Detection

Anomaly Detection is a widely studied field in statistics and machine learning and provides a way of detecting outliers in a dataset. Anomaly detection, also called 'outlier detection', 'novelty detection' and 'deviation detection', is defined as: "*An anomaly is an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism*" [42]. Anomaly detection is today used in many different kinds of software like fraud detection, intrusion systems, the medical and health industry and much more. It is considered by many industries to be an essential field of study because it is used to detect cyber-attacks, altering of data and more.

Anomaly detection can be defined in three ways, and these are: (1) Point anomaly, which is data that deviates from the normal pattern in a dataset, the study [42] uses as an example if a person usually uses five litres of fuel a day, but one day uses fifty litres, that would be considered a point anomaly; (2) Contextual anomaly, is when a data point behaves anomalously in a particular context. It is normal to expect that a credit card user will spend more money during Christmas; however, if the user would suddenly spend a lot during a non-festive month, something could be wrong; (3) Collective anomaly is when a collection of similar data points is behaving abnormally with respect to an entire dataset, [42] describes an Electrocardiogram (ECG) - one data point being low is entirely normal, however, if all data points are low, something might be wrong. These are different types of detection schemes; however, two different output formats should also be considered. The first output is scoring based output, where a data point can be scored on a scale, e.g., 0.0 – 0.1, based on how close the data point is to the 'normal'. The other output format is binary, so this simple means either it is 1 or 0, yes or no, etc.

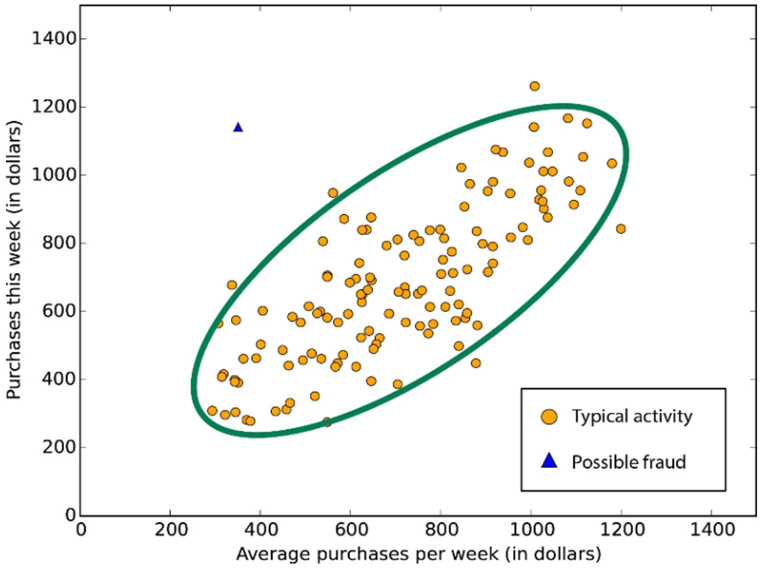


Figure 11 - An example of how anomaly detection works, showing an outlier to the entire dataset [55].

Figure shows an example of a fraud detection system. All the yellow dots in the chart are what is to be considered normal data, in the top left-corner, a blue triangle is shown. That data point is considered abnormal compared to the other data points and would in the case be flagged by the machine learning algorithm. How the model can be utilised in regard to data quality, will be investigated later in the report.

Classification

Classification model is a supervised learning model. The technique is label-based, which means that the model is able to predict whether a data point belongs to a specific category. This model, among other things, be used for image recognition. The machine would be able to predict what is in a given image, like detecting a dog or a cat - in that particular case the output would be the animal classification. The classification approach is often used to predict labels based on a list of features, and used in many situations, e.g., in spam filters, is the email spam or not? Or classification of items in a product list.

As mentioned, the model is supervised. Supervised learning is a method where a data set, is constructed and trained. The data set will include the features which should be learned, and it also contains the correct classification result. The model will then be trained with the data set and predicts the labels – categories of new data points, that it hasn’t seen before.

K-Nearest Neighbours

The Nearest Neighbours theory is a machine learning algorithm used in the classification area, and the idea is that the data point nearest to the given entry, from which a label is sought, will help to predict the label. The K is a value that defines the locality, this means that a smaller K will create local neighbourhoods and a larger K will create neighbourhoods where the minority is ignored.

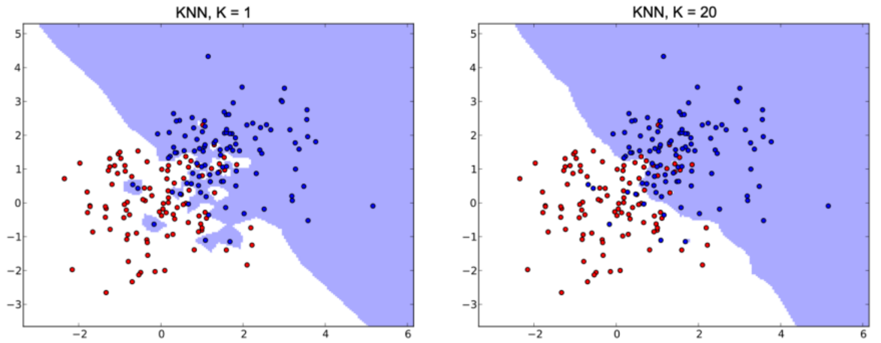


Figure 9 - The difference between a small and a big K-value [31].

Figure 12 shows the difference between a small and a large K-value. The figure is divided in two larger groups, blue and red. In the figure with a small K-value, small minorities arise in the blue area, this is because of the K-value. In the other figure to the right, smaller minorities are ignored, and the regions are not as local [31].

Support Vector Machine (SVM)

Support Vector Machine is another kind of machine learning algorithm. It is widely used and is especially efficient in pattern recognition, for example in digit or image recognition. SVMs are built on the theory of hyperplanes. A hyperplane is a plane dividing two or more different groups of data points. The algorithm will calculate the margin to the hyperplane, and by that finding how far away are the points in the two groups from each other. The further away the points are from the hyperplane the surer the algorithm is that the point belongs to the specific group.

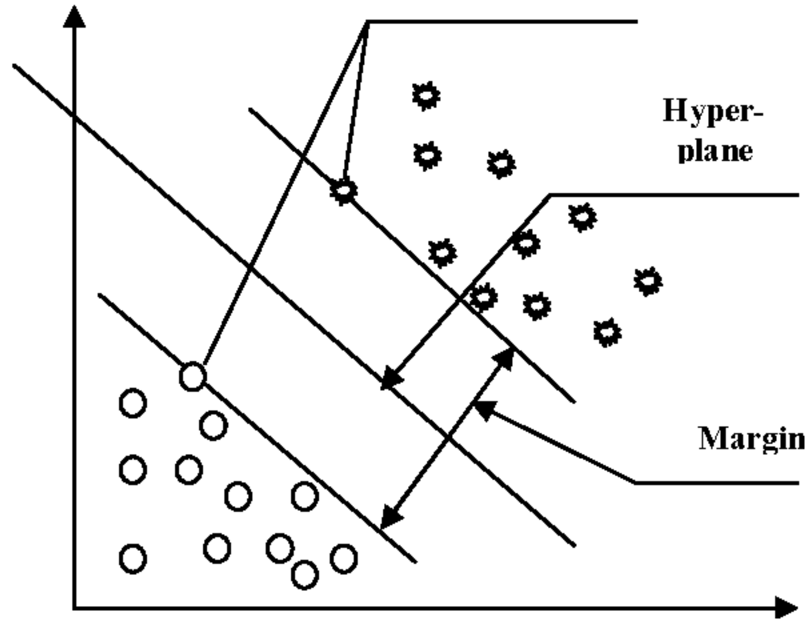


Figure 10 - The hyperplane and the margin to the two separate groups of data [51].

Figure 13, shows a diagram explaining the SVM. The hyperplane is the line centred in the figure, also showing the margin to the nearest data point. The idea is to find the hyperplane with the most optimal margin [51].

Machine Learning Data Sets

One important factor, that can also be limiting for smaller not well-studied projects can be the data sets. All machine learning models require a well-structured and organised data set. Often the data sets must also be quite extensive to get the best results and model accuracy. Getting the data can be difficult in itself; however, another major problem is formatting and structuring the data. As discussed in 2.3.1, data comes in a wide variety of different format, and for the machine learning model to work, the format must be unified. Furthermore, the data must be very specific to the problem that is being solved, unnecessary data or unprecise data can lead to untrue predictions [56].

Sub Conclusion

Machine learning has a wide range of benefits, when it is being used in data analysis models, however, it presents some challenges. The main benefit is the capabilities to comprehend data pattern, that is not visible to the human eye, and would requirement vast amount of resources to calculate. The two machine learning approaches investigated in this chapter is anomaly detection and the classification method. The anomaly detection model is beneficial in statistical analysis, finding data points that are abnormal compared to the overall dataset. The classification model is mostly used in prediction schemes, where the result is known for the training set, namely supervised. The approach is capable of labelling items, based on known features. The theories investigated will be further explained in regard to data quality and the proposed data model, at a later stage of the report.

3 Analysis

The State of the Art chapter laid the foundation for the continuation of the work. The intended purpose of this analysis section is to utilise the techniques researched in the earlier chapters hereafter discuss and evaluate them. The outcome will be a list of usable techniques, which would be applicable for the intended proposed evaluation system, but also to be applied in a proof of concept (PoC). Furthermore, a couple of realistic validated scenarios will be formed; this will give the reader an idea of how the system should be put into usage in real-life situations. The final outcome of the analysis chapter will be a detailed and comprehensive requirements specification that will later be used to create diagrams, architectures, and eventually a PoC.

3.1 Interviews

One crucial part of the project is the many interviews conducted as discussed in section 1.7.1. These are valuable since threat intelligence and sharing of such is a relatively new topic, and consequently information and studies on the topic are very scarce. Besides the fact that the topic is new, the MISP technology is highly specialised, and information sources are primarily blogs and forums, therefore getting information from experts is very giving in terms of reliability.

The interviews are not in traditional, sit-down style, where questions are prepared, and then answered by the experts. Instead, they interviews are conducted over a more extended period of time and in a more casual conversation way. The relationship between the author and the experts is a bit of collaboration, however not extensive, since the experts only provides inputs, suggestions and threat event information.

The following chapter will present short résumés of the interviews, where the central points are presented. Several of the interviews can be found as audio files attached to the project or as email correspondence in the appendix.

3.1.1 Dennis Rand, MISP Expert

As stated earlier in the introduction, the interview was not conducted in single sit-down session; multiple interviews were conducted throughout the entire project, from the starting phase to the final testing session. Dennis Rand played a significant role in the project since it was with him that the initial contact was made. The sessions were conducted online together with the supervisor of the project, Peter Anglov.

Dennis Rand is the founder of eCrimeLabs [58], a company specialising in cybersecurity consulting and hosting of MISP instances. Moreover, eCrimeLabs also offers a SOAR

(Security Orchestration, Automation and Response) [63] system, CRATOS, which has MISP included. This combined, gives Dennis extensive knowledge of the MISP universe.

The main agenda was to discuss how the MISP platform could be improved, in the eyes of someone working with it daily.

Dennis' answer was that MISP is full of events and threat information, and therefore two problems arise. The information is often of very low quality and the amount of data is enormous. Therefore, a solution to evaluate the events based on its quality and in large bulks, is needed. The solution should evaluate the data somehow. Based on the evaluation, it should calculate a confidence score and give the event a suiting tag, so other systems could automatically work with that tag, knowing the state of the data quality. The solution should preferably work in real-time.

Today, users of the MISP platform and the network forensics are experiencing a high data rate, and the problem is that with the amount of data entering the system, it is almost impossible to keep up with and analyse each threat event. This may cause a waste of time because if the event information is of low quality, and no actionable decision can be made.

The intention of the solution and its main contribution will be to aid the network forensics make a professional action to a threat better and faster.

Validation of IoCs in the MISP has been tried before, it could for example be to look up the IP address or domain name from where an attack originated. However, this is not a viable solution because malicious persons can easily change these parameters and make the validation incorrect.

One challenge could be to try to include machine learning in the algorithm. However, it is unknown if it is even possible, and a calculation algorithm would probably be sufficient to get the job done. MISP has a technology called 'Zero MQ' implemented: it is here that the events should be intercepted and be read in real-time. That is possible, because everything that happens in the MISP instance is recorded in the ZMQ.

Dennis states that he believes that a system to evaluate large amount of data, its trustworthiness and by that reducing the threat in network landscape is needed, and would be valuable for the MISP community: *"With this in mind the project the author is doing is essential to any organisation recieving arbitrary threat data from external parties, to help them assess the data recieved, as a guideline for the analyst on how potential trustworthy the data is, and with that be able to handle a larger amount and thereby reducing the threat landscape based on knowledge"* [Appendix 10.2.3].

3.1.2 Alexandre Dulaunoy, CIRCL, MISP Expert

CIRCL is a big organisation in the threat intelligence market, and they have dedicated researchers and developers, therefore it was decided to contact them for more expert knowledge. Alexandre Dulaunoy is a network forensic working in the field of incident response, and he is also a developer on the MISP platform. Dennis Rand made the initial connection, since he has direct contact within the organisation. He explained the idea, and the response from Alexandre was, "*This sounds like a very promising project. This could be indeed very useful to evaluate the events shared among a community or even within a private/local community* ". Alexandre also provided some ideas for threat event features that the system should be able to rate upon, presented later in the chapter. (The email correspondence can be seen in appendix 10.2.2).

Alexandre states their apparent interest in such feature in the core project "*As we are interested to have such kind of feature in MISP directly, I'm wondering if we could not improve the integration of the project in the core MISP*" [appendix 10.2.2]

Based on the received email, it was assumed that CIRCL has a pretty good idea of how a good dataset should be structured and therefore, Alexandre was asked if he could provide a list of data sets, which they imagined would be seen as excellent data sets. Requesting well-structured data sets was done to follow the same approach as in section 2.3.1; in that study, they use a 'golden dataset' and build the solution around it.

3.1.3 Persons of Danish Defence Intelligence Service

By personal experience from the supervisor, it was known that the Danish Defence Intelligence Service uses the MISP platform to some degree. It is assumed that because the service is dealing with large amounts of data, they could benefit from a service capable of quantifying the data quality. Due to these reasons, it was decided to reach out to them and hear if they would be interested in making a small collaboration, giving their take and input on how such a system should look like, and what requirements would be needed if it were to be used in their systems.

As a result of reaching out to the organisation, an invitation to come see their system setup and hear more about the organisation in person was received and accepted. Since the service is operating in full discretion and the organisation is highly confidential, recording of interviews and conversations was not allowed and therefore no documentation can be provided.

After a couple of conversations with the contact person, which possesses a relatively high position within the organisation, it was learned that an algorithm to measure data quality is needed, so the work of manually evaluating each threat event could be minimised, are needed. FE are getting threat events from many different services, not only the MISP platform, therefore it is essential that the system could evaluate data from multiple sources and not only the MISP data, so in other terms the solution must be very generic and adaptable.

FE also proposed that if a functioning PoC was to be created, unclassified data from their system could be provided to test the PoC with ‘real data’.

3.2 Scenarios

This brief subchapter includes three near real-life scenarios that will explain how the system could be put to use. The scenarios will describe three different situations where the system is utilised and considered most efficient. The scenarios will be different and will be based on the interviews conducted with the experts throughout the project; therefore, the scenarios are considered validated because the experts were involved in their creation.

3.2.1 The MISP, New Feature, and Night-Time Protection

Introduction

The first scenario described will explain where the system will contribute to the MISP platform, especially at night-time, when organisation staff is usually at home, not evaluating the threat events that are incoming continuously.

The scenario

Casper works in an organisation as a network forensic; they have recently implemented a MISP instance to be a part of a cyber threat sharing community. They decided to do so because they realised that because of their growth, the organisation is becoming more and more of a target for hacker attacks, and they just recently dodged one. The hacker attack they dodge started at night-time, and there were none at work to counter the attack. Casper clocked-in the morning after and realised an attack was going on, and because of his quick thinking, he countered the attack, and no information was lost or stolen.

With the MISP instance and the newly added feature called ‘Event Data Quality Evaluation’, Casper is not so worried anymore when he leaves the office because he knows that the service is running all night long.

One morning, Casper came into work, and just as every day, he goes through his security logs and sees what has happened in the MISP instance.

He can then see that the MISP instance and the new feature has evaluated many threat events during the night by sorting and tagging them. The evaluation of the MISP events is based on the configuration Casper has specified. He can see that some events with poor data quality have been sorted out and ignored since they were probably not so ‘dangerous’; however, not all of the events have been ignored. One event which scores a relatively high confidence score has been spotted. The MISP instance is then, on its own, capable of blocking all network traffic coming from that event because it is connected to the organisation’s firewall.

Casper is happy that the new system is working; in a typical instance, Casper has to manually go through every event, evaluate them and decide whether to block the traffic or not, and now this system is capable of doing precisely that, but much faster and with higher precision.

3.2.2 Amount of data make the evaluation impossible

Introduction

This scenario will describe a situation of overload of threat data coming into the MISP platform because more and more people are joining these collaboration systems and share data. Furthermore, it describes how the evaluation tool helps lower the burden of work for network forensics.

The scenario

Casper is working as a network forensic in a large company. A long time ago, they installed a MISP instance. Casper is responsible for going through the MISP instance, searching for relevant threats, and deciding what action must be taken against the malicious activity.

When Casper is out talking to friends, who work in similar positions, Casper notices a growing trend for using MISP, and that same trend is also evident in the amount of data in the MISP. More and more people are joining and contributing to the community with threat intelligence.

Casper has subscribed to many feed providers, and they are releasing a large number of events each week, and it is Casper’s job to go through each event and all the attributes, evaluate the threat and decide whether he should block the traffic or ignore the threat because it is not relevant, or does not provide any context.

Because of the trend of using MISP, Casper cannot keep up anymore with the amount of information he has to go through, and this can have severe consequences for the company as the risk rises.

One day, MISP releases a new feature, a data quality assessment feature, which utilises a tool called ‘ZeroMQ’. The ZMQ is a message queue, and each event has to go through it; at this point, the event can be captured and evaluated. Casper will then specify a configuration file that the event has to be evaluated upon. The system can create a confidence score from 1-100, where 100 is the most optimal data quality. Casper can then choose a threshold, e.g., scores under 75 should just be ignored since the data quality is so low or without context. Nevertheless, if the score is above the threshold, the system must block traffic from that event.

After this feature has been implemented, Casper now has a lot more time to go through the events with the highest scores and think more about suitable actions to sanctioning the events.

3.2.3 Full Customisation Possibilities

Introduction

The last scenario will describe how the algorithm works and its large number of customisation options, allowing the system's users to evaluate what they find most suitable for the organisation.

The scenario

Casper from the large company sets up the new features from MISP where the incoming events are evaluated upon their data quality. When setting up the new functionality, Casper has to decide on what factors the system shall evaluate the event upon and the way that is done Casper find intuitive. In the MISP instance, a configuration file is placed, which works just like a programming IDE, and here Casper quickly reads the documentation and the small 'library'.

Now Casper is ready to set up the features to evaluate upon. In the configuration file, Casper is shown different event features that he can turn on and off as he likes, this makes is possible to measure on features that are important to the company. He can further customise how many points should be granted towards the confidence score, if the requirement is fulfilled. A functionality to balance the turned-on features based on importance can also be used. This feature allows users to e.g., set importance of tags higher than usage of galaxies, this will have an influence on the confidence score. In extension to the data importance of features, Casper can also set different threshold, so e.g., the threshold of tags can be that an excellent event should contain four tags, so if the event contains two tags, only half the number of possible points should be given.

These customisation options are suitable for the organisation because Casper has multiple evaluation tools running, and using these configurations, he can focus each evaluation on something different.

3.3 Analysis of Technologies

This chapter will investigate how the frameworks and technologies, which were researched in the State of The Art (Section 0), can be utilised in a new system and in an actual implementation of a PoC. These technologies will together propose a solution to answer the problem formulation. This chapter will serve as the backbone for the development and design chapter.

3.3.1 Zero Trust in MISP

The main issue of the MISP and data quality is that there are no reputation systems or trust-creating mechanisms that can help identify trustworthy sources of information.

Using online technologies where different parties interact, the concept of trust will always be involved. Doing trade in the physical world or with a website, a person will most often research the other party before buying or selling an item to ensure the other party has no intention of harmful conduct. The same concept stands when sharing of data is happening.

Trusting unknown parties in the digital world can be very complicated since entities are not obligated to expose their true identity.

As of the current and latest MISP version, 2.4.138 at the time of writing, there is no mechanism to keep track of data or feed providers and their true identity; however, the users can see earlier data published by the provider, which is the only clue to how good data provided is.

When a security incident happens, larger feed provider organisations have dedicated researchers to explore, evaluate and describe the attack in the MISP platform. The feed providers will create an event and fill in relevant information about the attack, such as objects, galaxies, report events, as well as other relevant data, and attach attributes, which are the actual attacks examples (IoC). The event, along with all the data, is then published to the community (in this project, it is the feed community), and the receivers can do with the information as they like.

The above-described process very well, but one problem arises at the receiving party. How do the receivers of the threat data know that the feed provider is not a malicious group trying to spread false information in the community? Another example can be, that even when the event is published in good faith, it is of such low quality that the data is not useful.

As stated, in the current MISP version, there are no reputation systems to keep track of how good data the feed provider is, in terms of earlier delivered data, and there are no algorithms to calculate how high quality the data is. The feature that comes nearest to such algorithm is the box IDS, which can be flagged by the feed provider. The IDS box means that the feed provider states that the data is so good and clean, automatic services can process it without human interaction. Validation of IDS is not implemented; therefore, trust is essential when using this IDS feature.

Discussing the issue with experts in the field, there is a broad understanding that all events must be manually evaluated before making any decisions because such mechanisms and algorithms do not exist. This way of thinking is called ‘zero trust’ [64], the concepts is simply that no entities are trusted by default, and everything must be assessed manually.

3.3.2 Creating trust

Now that it is established that trust and reputation creating mechanisms are not implemented in the MISP platform, what must be discussed next, is how can trust then be created. Trust on online platforms can be created in many ways, and many standardisation organisations have principles and guidelines to build trust. One standardisation organisation mentioned in section 2.3.3 is NIST. NIST operates with five different models to obtain more trust in communities. Most of the models from NIST use third parties for validation, which to some degree, must verify one of the parties; this could also be done in the MISP platform. One method is the mediated trust, where one organisation would rely on a third party's experiences, e.g., a reputation mechanism. The reputation mechanism could be implemented so each organisation would keep a list of trusted feed providers based on their data quality. Another approach would be that each user of the MISP platform would keep their own private list of trusted organisations. Keeping a private list would also make it possible to manipulate the list for more customisation. This could be beneficial if the experience with the organisation has been obtained elsewhere than the proposed solution. For example, if a MISP user personally knows a feed provider, they could then tell the system that feeds from this provider must be trusted.

The solution mentioned above, using third party experiences, could be utilised well in a reputation system, where each feed provider is rated based on the quality of the information provided.

As described in section 2.3.2, reputation systems have many benefits; however, different challenges with reputations system must also be considered.

One problem that quickly arises is the 'cold start' problem. The problem which is described in Section 2.3.2, occurs when the system is initially launched, or a new user is created. When the system initially launches, there are no previous interactions with any feed providers; therefore, the system cannot know how trustworthy the feed provider is because there are no data about them. The same problem is present when a new user is created; the new user has no earlier experiences with the feed providers.

As described in Section 2.3.2, two popular solutions exist to overcome the problem: the 'Random Strategy' and the 'Popularity Strategy'. The best strategy to use in this case is the Popularity Strategy; this strategy suggests looking into the community and see how the feed providers are rated elsewhere, and the new instance can utilise this data. The Random Strategy is not applicable, because presenting the user for random feed provider ratings would not make sense at all. The Popularity Strategy is very applicable together with the collaborative sanctioning described in section 2.3.2, because Collaborative Sanctioning will rank data providers low if they present data of low quality. Using the Popularity Strategy will further sanction feed providers of low quality.

The Collaborative Sanctioning is important to use, so feed providers that provide information of low quality are sanctioned, and other users of the platform are warned. On the other hand, if feed providers deliver information of high quality, they should be rewarded, by having more people utilise their information.

Figure 14, shows the proposed solution. It depicts a new user entering the system. At the point of entry, the user has no information about feed providers in the community. The idea is that the new user, can request from the community a shared list of rated feed providers. The seasoned users of the system could then share their list of feed providers with the new user. This approach could, to some extent, solve the cold start problem.

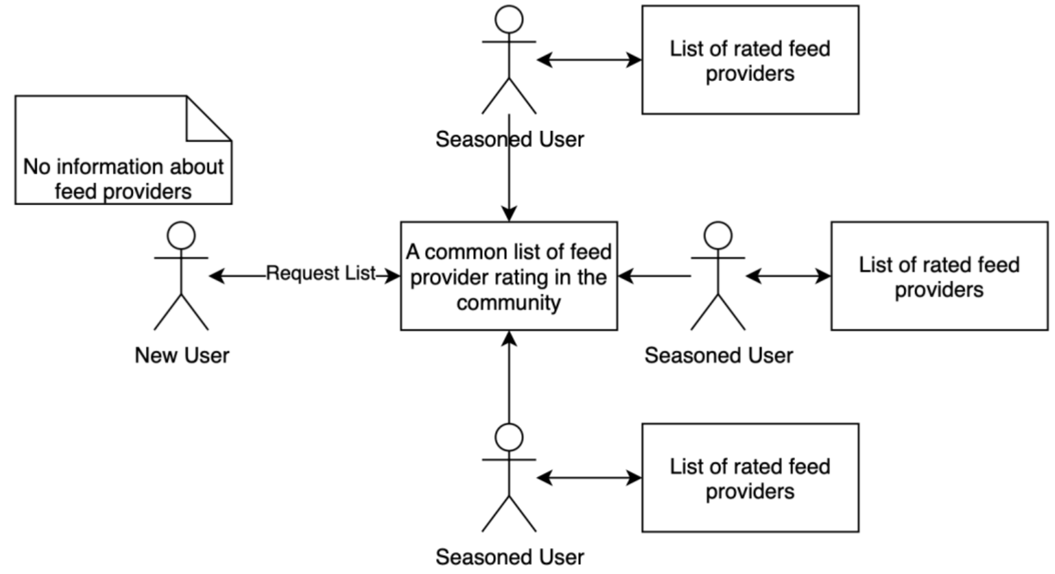

List of rated feed providers
No information about feed providers
Seasoned User
New User
Request List
A common list of feed provider rating in the community
Seasoned User
List of rated feed providers
Seasoned User
List of rated feed providers

Figure 11 - How the community could share a list of feed provider ratings to a new user.

Another challenge that reputations systems are facing is of information sources. As explained in Section 2.3.2, different information sources exist. In this work, it is primarily the ‘Direct Historical’ and ‘Witness Information’, which are considered. The problem with threat intelligence information is that it is generated by humans and it is often Witness Information that can be exaggerated or untrue. As already discussed in Section 2.3.2, Trustpilot, has implemented a method where the selling party can ask for an order or reference number to validate an actual transaction; however, this technique is not applicable for this solution. False positives in threat intelligence sharing are a well-known problem, and the MISP platform has already created a minor solution to help users identify false positive. They have created long lists with known, white-listed IPs and domain names. If one of those IPs or domains is used in an event, the entire event will be flagged with a big ‘False Positive’ warning. Besides the warning, they have also created an API that can check for false positives. One solution is to implement the API to check for false positives in the proposed system, and if a feed provider organisation keeps posting false positives, put them on a watch list. Being on the watch list would mean that none of their events would be processed automatically until they are removed from the list again. The Direct Historical experience would be great to utilise, but it would be challenging since it would require dedicated researchers for each private MISP instance, to directly obtain the data; therefore, this could be solved by creating threat events in-house, but this is a feature for the future when more resources are available.

The MISP community is significant, and many blogs and forums exist. Therefore, it was decided to search the forums to see if any smaller communities exist where feed providers' ratings are shared. This would help when creating the system so that the cold start problem could be avoided. After more comprehensive research, there are no indications that such communities exist, and therefore it is assumed that such feature in the system is reasonably untouched.

3.3.3 Data Quality in MISP

In the State of the Art analysis, Section 2.3, where research of central concepts is conducted, it was apparent that data quality is an essential topic in threat intelligence sharing and other data-driven systems. The data is often used to make critical decisions, and therefore neglecting it could lead to severe consequences, as explained in Section 2.3.1. The least damaging consequences could be making wrong decisions, and lowering trust in communities, while worst ones could be not blocking network traffic that should have been blocked, which in the worst case could lead to intrusions into the network infrastructure.

The data quality in the threat intelligence platforms is of high importance; dedicated network forensics base their professional decisions on the data quality; therefore, data must always be relevant and actionable. Trust also plays a significant part in decision making. If network forensics have a good trust relationship with the feed provider based on previous threat events, the network forensics can trust that the data is reliable, and to some extent, make the decision more automatic, so the system itself can block events of high data quality.

To create an evaluation system, it is vital to understand the data flow in the platform. In the current version of MISP, no mechanism is implemented to help network forensics or regular users make decisions upon threat events.

Figure 15 shows the setup of incoming threat events. A feed provider will be noticed about an ongoing or finished threat. The feed provider will then analyse the attack and collect all possible information. When the feed provider has enough information about the activity, an event will be created. The information about the activity will be entered in the newly created event, and attributes will be attached (actual attacks). The event will then be published to the MISP platform (core project) and shared with the community. Users and forensics of the MISP will then be notified that a new event has been published; it is up to the individual forensics to decide whether to block the traffic or ignore it.

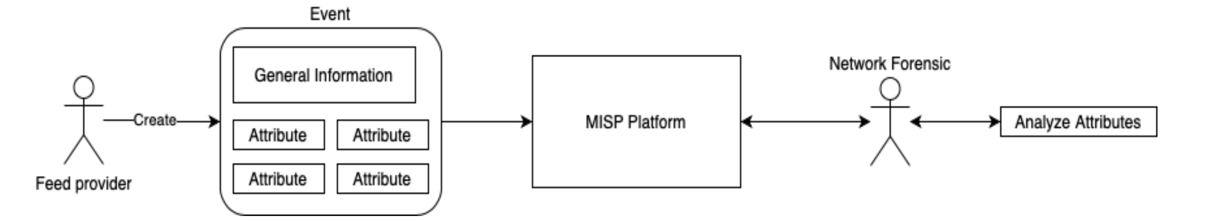


Figure 12 - The current flow of events in the MISP platform.

This setup is working as intended; however, putting all the responsibility on the forensics can have negative consequences, such as poor decision making, or losing focus. It is here that the main contribution to the MISP community will be. This process must be optimised because humans can make mistakes, and with the growing amount of data in threat intelligence, it is impossible for forensics to analyse and evaluate all the information.

The exact point in the current flows of events where the system can be improved is between the MISP platform and the network forensic. The idea is to introduce an algorithm, which saves time for the network forensics by pre-analysing the event data and filtering out low-quality data events.

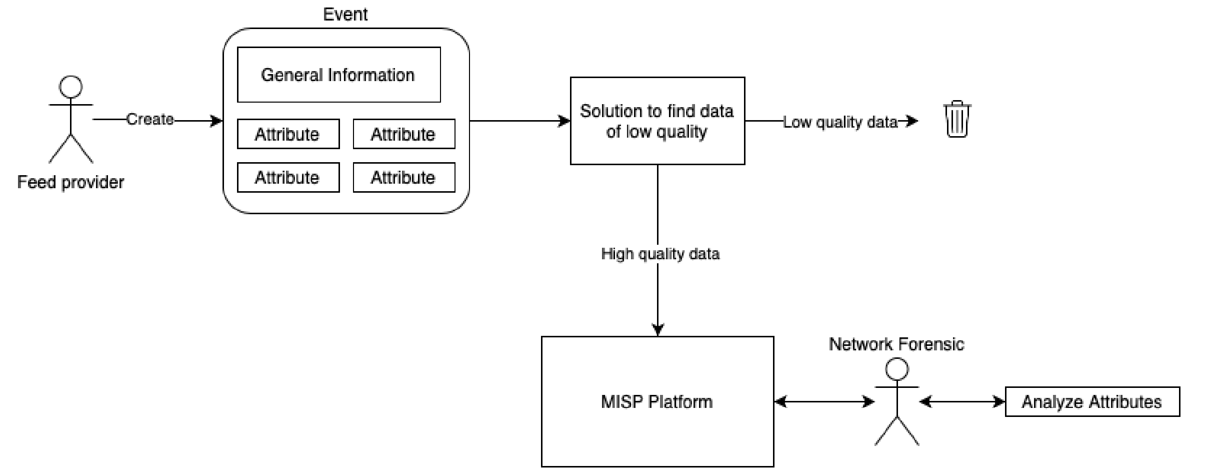


Figure 13 - Shows where a potential solution should be placed in the MISP landscape.

Figure 16, displays a possible solution to how a system landscape could look if a module to filter the events was implemented; the module should, as discussed in Section 3.3.2, include a reputation system and a sorting algorithm to filter out data of low quality. The reputation system will allow forensics to see how well feed providers are performing and by that built trust with the feed providers. The figure shows the solution as a black box, illustrating where the solution should be placed; the black box will be investigated later in the report.

Features to Evaluate

The studies discussed in Section 2.3.3 and 2.3.2, both describe what they find relevant to evaluate, which is the general data of threat event, the completeness, relevance and freshness. These dimensions are all present in MISP event and will be evaluated, however, this information alone is not sufficient enough to determine the data quality of the threats, the evaluated features must be more detailed. Important features are selected in Section 2.1.1, and to find the most relevant ones, Dennis Rand was once again asked, since it requires expert experience to know what features are most important to the forensics.

To get even more knowledge, it was also decided to contact Alexandre Dulaunoy from CIRCL (Section 3.1.2) and ask him for advice on how to rate the information. Alexandre Dulaunoy put together a list of features that could be evaluated (Appendix 10.2.2).

Suggested Event Features
Are objects used?
If yes, are those connected with relationships?
What’s the ratio of relationships on this event? Versus the over ratio of the community?
What are the relationship used? Versus the over ration of the community?
Are the comments fields used or not?
Does the comments field contain sentences? Or duplicate information form the attributes/objects?
Are the first-last-seen field set?
Are the type of object used common? Or uncommon among the other events?
How many tags are used? Versus the median of tags used in the community
Are those tag for taxonomies? Or not?
Are galaxy clusters attached? Ratio/community?
Are event reports used?
The threat level
Are event extension used?
Are TLP used?

Figure 14 - List of event features to measure upon, made in collaboration with Dennis Rand and Alexandre Dulaunoy [10.2.2].

Figure 17 shows the combined list of what factors an algorithm could measure upon, proposed by Alexandre Dulounoy and Dennis Rand.

As discussed in the State of the Art Section 2.3.1, data quality should be measured on all of its different dimensions. The dimensions define the data quality as a whole, and therefore it was decided to compare the potential event features that should be measured, listed in Figure 17, with the data quality dimensions to see if the feature cover all of the dimensions. It was learned, that it is complicated because some factors can easily fit into more than one dimension, and some dimensions are more difficult than others to measure.

The dimensions and evaluation factors can be seen in Figure 18.

Factor to evaluate upon:	**Data Quality Dimension:**
Are data fields filled?	Readability (Completeness)
What is the threat level?	Relevance (Fitness)
Are comments full sentences or bullet points?	Presentation (Structure)
Are objects used?	Usability (Documentation)
Are galaxies used?	Reliability (Accuracy)
Are tags used?	Reliability (Accuracy)
Are the tags taxonomies?	Reliability (Accuracy)
Are TLP used?	Availability (Authorisation)
Does the event extend another event?	Reliability (Consistency)
Is the event on the warning list?	Reliability (Integrity)
How many tags are used?	Reliability (Accuracy)
Is the report function used?	Presentation (Readability)

Figure 15 - The event features and corresponding data quality dimension.

As stated in [32], it has been a general misconception for a long time in many communities that the quality of data is only concerned about reliability; therefore, it is vital that the event features to measure generally fall into most of the dimensions. Figure 18 shows the features to evaluate, and their corresponding dimensions. Not all features from Figure 17, are shown in the list of dimensions because they are not considered to be used since it would not be possible to automatically evaluate those from the MISP threat events.

Quality of Indicators

During the interviews with the experts, it is clear that not all factors are equally important for data quality evaluation, meaning that not all factors should count equally to a potential rating score. One way to give individual score is to divide the categories into different groups of importance, like a kind of Quality of Indicators, as discussed in Section 2.3.1. The section discusses how to use the QoI. It consists of four dimensions: correctness, relevance, uniqueness, and utility. The first two dimensions are already covered in the data quality dimensions. Uniqueness of information in the threat event, is very difficult to measure, because of the simple reason that the MISP events are not stored in the ZMQ, and the system must run in real-time, therefore comparison of old and new events is not a possibility. The last dimension, utility, or how the data is used, is applicable. If the event feature says more about the threat than the other features, then it should be counted higher in the rating mechanism. It is decided to divide the features into three groups of utility/importance:

- **1st Degree** of QoI is the group that includes the lowest importance factors to evaluate; these do not define the data quality to a large extent; however, these are fundamental and should still be filled in by the data provider.
- **2nd Degree** of QoI is considered to be of higher importance. These are classified as more defining of the data quality; however, they can be filled out without deeper knowledge about the threat event.
- **3rd Degree** of QoI, refers to all the factors that requires extensive knowledge about the threat event.

To define the factors that should be in each group, the list from before delivered by Alexandre was analysed, and they were divided into the three degrees of QoI and can be seen based on their importance in terms of data quality. The QoI and the features can be seen in Figure 19.

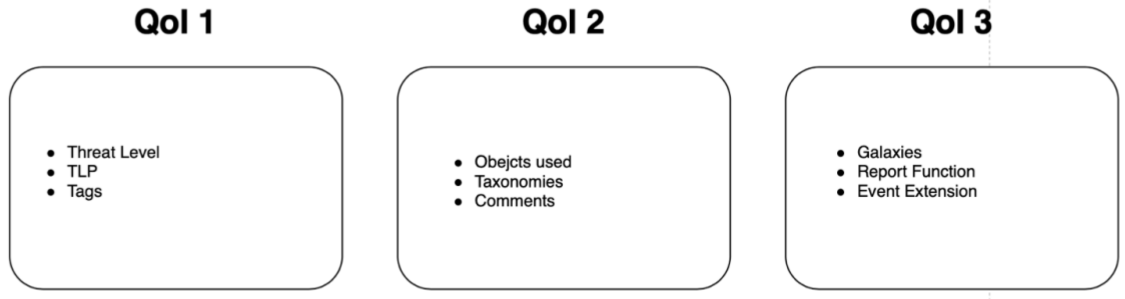


Figure 16 - The quality of indicators and events features, prioritised after importance.

Data Structure Challenge

One of the requirements from Section 3.1.3 is that the algorithm must be generic, meaning that the data quality evaluation model must not only be bound to the MISP platform only. It should be able to evaluate threat events from other platforms too. Many of the platforms use JSON format; however, the JSON structures are not standardised, and this could cause a problem for the algorithm. When coding, a JSON analysis must take place, meaning that to access the data in the treat event, the name of the JSON object must be known because the data is paired as key:value in JSON e.g. Galaxy:PhisingEvent. Therefore, it is a must that the key name must be known beforehand, however, data coming from another provider could use different names for the key or not even include the same data; therefore, the algorithm must be somehow robust and generic to withstand this challenge.

It will be challenging to overcome this problem. However, one approach is to formalise the data: taking the JSON from different origins and creating one unified JSON structure with the data, so all objects have the same names. Then the newly created JSON should be analysed instead of analysing different JSON structures. This approach will not solve the threat events including different pieces of data; for example, looking at the STIX 2.1 [65] example JSON file, they do not include tags and galaxies, which are essential for the MISP analysis and confidence score. Based on this knowledge, a conclusion must be made that the only way to overcome the problem is to create a separate analysis model for each JSON structure passed into the evaluation model, which will be very consuming in terms of resources.

3.3.4 The Rating Algorithm

The scoring of the event will be the central to the algorithm; this calculation model should be reliable and consistent. It is decided to investigate the potential of applying machine learning algorithms, but also a more traditional mathematical calculation approach.

Machine Learning and Mathematical Formulas

Machine learning, as discussed in the State of the Art section (Section 2), is a prevalent technology and used in many applications, and therefore it must also be considered and explored in this project. Machine learning comes with many benefits; however, it can be complex to implement, and it relies on suitable datasets that can be difficult to obtain. Making machine learning models on inadequate datasets can give erroneous or biased results, which will produce wrong outputs.

Machine learning is not a part of the MISP platform or community yet, and therefore finding suiting datasets for an implementation model was not possible. Therefore, the dataset must be created before implementation, or another approach must be developed.

As mentioned in section 0, contact was made with Ming Shen, professor at Aalborg University, who proposed anomaly detection, discussed in Section 2.3.6.

After researching the anomaly detection technology, it is realised that it may be suitable. One approach could be to use the acquired ‘golden datasets’ from CIRCL as an anchor point, and when new data is coming into the MISP platform, simply compare the new data to the anchor point, because it is known that the anchor point should be of the highest quality possible. The anomaly detection model can produce a confidence score based on 'how far' the new data set is from the golden ones. Figure 20, shows the model for this solution, including anomaly detection. A new threat event is captured in the ZMQ and is then compared to the golden datasets, which will produce a confidence score, that will be stored in the MISP instance.

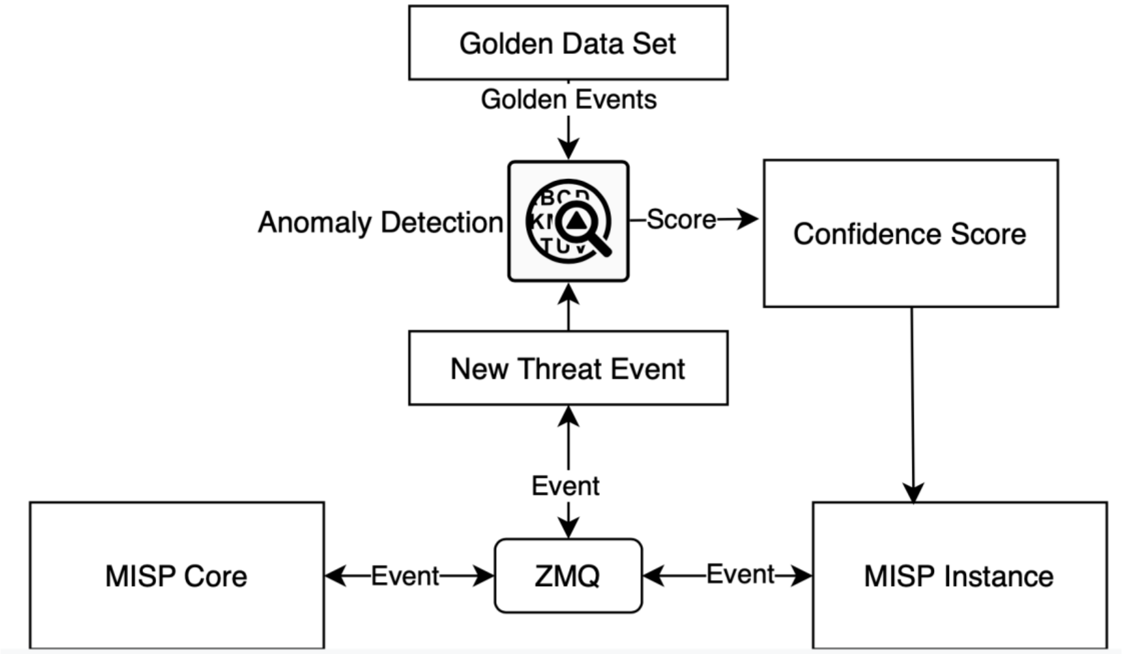


Figure 17 - The anomaly detection model, comparing new threats events with the golden dataset.

Since no data sets were available, another solution was necessary. Another approach could be to use a mathematical formula, calculating a confidence score for each threat event. The formula, which is relatively simple and very well suited for this operation, is the weighted average model, described in Section 2.3.5. As discussed, the model is often used in systems with many variables and systems where variables need to be balanced in terms of importance. One requirement for this project is that the MISP platform is highly customisable, and therefore the user should be able to set how much each event feature should count towards the confidence score. This means, when using this model, the user of the proposed solution can set the number of points each event feature should count, but furthermore, they can also set the weights of each QoI, as discussed in Section 2.3.2, and in that way make one QoI count more than the others.

The weighted average model will score the threat event from 1 to 100, which indicates how good the data quality in the incoming threat event is. The first formular shows the first part of the weighted average model.

$$Confidence\ Score = \frac{\left(\begin{matrix}(Point\ of\ QoI_1 * QoI_1 Weight) + \\ (Point\ of\ QoI_2 * QoI_2 Weight) + (Point\ of\ QoI_3 * QoI_3 Weight)\end{matrix}\right)}{AmountOfChecks}$$

To find the points of each QoI, data quality checks are implemented. These will score each event feature and to what extent they satisfy a threshold specified by the user. For example, the user can set a threshold value, stating that if an event includes six tags, 10 points should be given towards the confidence score. The model will then also calculate how many points should be given if for example only three tags are provided, which should be five points, the following formulas explained further:

$$Y = P\% * X$$

$$Points = \frac{\#Tags}{Optimal\ Tags} * Optimal\ Points$$

The last part of the model is to calculate the most optimal point score, the largest amount of points the event could get if all data quality check where fulfilled. This follows the same principle as the confidence score:

$$Optimal\ Score = \frac{\left(\begin{matrix}(Optimal\ Point\ of\ QoI_1 * QoI_1 Weight) + \\ (Optimal\ Point\ of\ QoI_2 * QoI_2 Weight) + (Optimal\ Point\ of\ QoI_3 * QoI_3 Weight)\end{matrix}\right)}{AmountOfChecks}$$

To see how good the confidence score is, the confidence score is held up against the optimal score, and the outcome will be the outcome from 1 to 100%, where 100% would be if the event satisfies all of the data quality checks.

$$Final\ Score = \frac{Confidence}{Opmital} * 100$$

Hybrid Model

After a period of research, a third approach was investigated, where the two models are combined. The anomaly detection model is limited because no dataset exists. However, the second mathematical model evaluates each incoming event, and therefore, the model could for each event store the features and the outcome in a file. Eventually, when the mathematical model has evaluated a vast amount of events, an anomaly detection model should be powered. After that point, two ways to continue the model is possible; either the mathematical model is shut off since it is not needed, or it is continued, and the confidence score of both models is compared and shown to the user. The second approach will require more computing power but is not assumed to be overwhelming.

3.3.5 Tagging of Events

The intention of the system is that it should function without any human interaction (besides for events that should be manually evaluated). It is therefore crucial that the system knows when an event is ‘secure’ enough to be processed automatically later in the system landscape. MISP is utilising the great potential of the tagging mechanism, as explained in Section 2.3.4.

MISP utilises the mechanism of tags heavily, and it is believed that this approach would also be applicable for the system proposed in this report and in the PoC. PyMISP (section 0) makes this possible. PyMISP has a function that lets the user enter his tags both locally and globally using the API; furthermore, the API allows the user to create custom tags.

One possible solution would be, after evaluating the event, and when the confidence score has been assigned to it, a tag must be attached to the event. The tag must be based on the score given earlier; this allows for further automation of the event analysis, later on in the system landscape - in a potential SIEM systems or other security software. The part of later integration will not be covered in this project but is considered.

The tagging of the event is essential; the tag must be very descriptive, and, e.g., tell how high the confidence score is; this would allow the system to know what action to take based on the confidence score entirely automatically. For example, three different thresholds could be made. All events scoring below the lower limit should be ignored or deleted. Events that score between the lower limit and the upper limit, should be manually evaluated and lastly if an event scores above the upper limit, it could automatically be processed in the system.

3.4 Requirements Specification

This section is the last subchapter of the analysis, where all of the information will be put together, forming a unified outcome, namely the requirement specification. It will serve as the guideline for the implementation of the proposed system. The requirement specification will be divided into functional requirements, which will describe the functions and behaviour of the system, and non-functional requirements, which are more performance-based. The functional requirement table will be divided into three columns, the actual requirement, the origin - where it originates from in the report, and a description of the requirement. Furthermore, the requirement will be prioritised; this is done to limit the scope and be realistic of what is possible in the span of this project. Some features are maybe desired, but because of the limited time the project is limited to, some will not be implemented.

To handle the prioritisation, it was decided to utilise the MoSCoW method; this method is based on the M - Must, S - Should, C - Could and W - Won’t [49].

It is important to notice that the requirements specifications, is representing the PoC, and not a fully functioning system.

Functional Requirement (Priority)	**Origin**	**Description**
The system must be able to read incoming threats from the ZMQ. (M)	3.1.1	The data must be real-time; to achieve this, the system must intercept the data in the ZMQ.
The system must be able to evaluate the threats based on different event features. (M)	3.1.2	The features to evaluate could be limited to tags, objects, galaxies, event report, threat level, TLP tags, taxonomies, information, event extension, trusted organisations and warning list.
The system must calculate a confidence score. (M)	3.1.1	The algorithm will be based on how good the data quality is, calculate a confidence score.
The system must provide a tag to the event based on the confidence score. (M)	3.1.1	Based on the confidence score calculated by the algorithm, the system will provide a tag for further automation implementation.
The user must be able to say which factors the evaluation should include and not include. (M)	3.1.3	For customisation purposes, and because the system must be generic, the user must be able to set the factors to evaluate upon
The user must be able to decide how many points each factor should give to the confidence score. (M)	3.1.1	For customisation purposes, and because the system must be generic, the must be able to set how many points each factor fulfilled should provide
The system must use QoI of different degrees and weigh them. (M)	2.3.1	To let the user, decide the weight of each QoI, the importance can be set of the QoI
The system must include a weighted average model to calculate the confidence score. (M)	**Error! Reference source not found.**	A weighted average score was decided to match the needs for such system. The model will calculate a confidence score and the optimal score, the ratio will indicate of how high the quality is.
The user must be able to upload a JSON file. (W)	3.1.1	To find factors in a JSON file, the user must be able to upload such and then the system should be able to locate them.
The system must have a machine learning algorithm implemented to define QoI weights. (M)	3.1.1	To guide the user in setting the weights of the QoI, a machine learning algorithm must be implemented.
The system could have a dashboard. (W)	3.1.1	The system could have a GUI that shows a dashboard of key indicators that have been evaluated.
The system must allow the user to set threshold values for DQ checks. (M)	3.1.1	It is important that the user can specify to the system when a data quality check is fulfilled. Points should be given depending on how much the check is satisfied.

The system could use anomaly detection to evaluate threats. (C)	2.3.6	The system could use an anomaly detection scheme from a golden set provided by CIRCLE
The system could put feed providers on a watchlist. (C)	3.1.1	The system could put organisations on a watchlist if they keep posting false positives.
The system won’t include a mechanism to share rating. (W)	2.3.2	The system won’t include a mechanism to share rating list to overcome the cold start problem
The system must include a reputation system. (M)	3.1.1	The system must have a reputation list, e.g., Excel to represent the feed provider ratings.

Figure 18 - The functional requirements

The following table lists the non-functional requirements; these are requirements that say something about the system’s performance.

Non-Functional Requirement (Priority)	**Origin**	**Description**
Must be real-time (M)	3.1.1	Since threat intelligence is time sensitive, it is important that the system operates in real time.
Must run at all-time on the server (M)	3.1.1	Threats are constantly entering the MISP platform, and therefore the system should be running non-stop.
Must use PyMISP (M)	2.1.5	The MISP platform has a service called PyMISP which has different APIs implemented.
Must have a running MISP instance (M)	2.1	A running MISP instance is needed, to get access to threats and the ZMQ.
The system must always handle the newest version of MISP (M)	3.1.1	Because MISP is constantly developed and is open-source new versions are often released. It is important that the proposed system can handle this.
The system must handle data from other threat intelligence platforms (W)	3.1.3	One requirement from FE, is that the system must handle data from other platforms than MISP.

Figure 19 - The non-functional requirements

Sub Conclusion

It was found that many concepts from the State of the Art analysis could be utilised in the system. The chapter started with the interviews, where the general direction of the report was established, and essential requirements were discussed, but furthermore, the direction was confirmed by the Danish Intelligence service that this was, in fact, something that could be used in ‘real-life use cases. However, the solution should be generic and not explicitly bound to the MISP platform.

It discussed in the analysis section that trust in the MISP platform is limited, and by experts' statements, it can be concluded that a 'zero trust' approach is used, so all events must be evaluated. A solution should be developed, and the solution could be based on a reputation-based approach to build more trust in the community. The reputation system will create a list of feed providers where their rating is specified. The list should be private to the MISP instance; however, with the possibility to share the list with new users of the system to avoid the cold start problem.

The reputation system will be based on a calculation algorithm, which can calculate a confidence score. The most optimal solution will be a hybrid combination of a weighted average mathematical model and anomaly detection, a machine learning model. The model will produce, as mentioned, a confidence score and based on the score, a tag will be given to the event, so other frameworks in the infrastructure can atomically process this event and its information.

4 Design

The analysis chapter laid the foundation for a set of functionalities a system must possess to solve the problem formulation; it also gave inspiration to functionalities for a proposed PoC, where some basic features are included to confirm the concept and idea. The next stage of the project is the Design chapter. This chapter will first provide a minimal prototype of the system, showing how a potential system should look. Also, the architecture will be illustrated and explained in detail with the necessary UML diagrams to provide a visual explanation. The chapter will finally propose a proof of concept and the implementation of such, including snapshots of the most central parts of the code.

4.1 Prototype

The intention of this sub chapter is to provide the reader with an idea of how such potential system could look like, if a full implementation of the system was done. The project is mostly concerned about the backend algorithm to quantify the data quality of the feed providers, however a GUI (Graphical User Interface) should also be considered for potential development in the future.

4.1.1 Graphical User Interface

Since the system would be able to categorise the trust sources after how well they perform in terms of data quality, and the algorithm can measure different parameters or indicators, it would be useful to design a GUI to show the user a data quality report. Before moving forward, it is important to notice that the following figures do not represent actual implementation but is conceptual designs. Figure 23, shows a dashboard that the system users could be presented with; the idea is, that the dashboard can display different indicators of data quality that the system has evaluated. The graph in the middle would show how the general performance was over all of the subscribed feed providers. Under the graph, the user would be presented with different indicators of the algorithm, and to the left, a top score list of the feed providers should be shown. These components combined would give the user a good impression of the general quality of data that is fed into the system or if any changes should be made to enhance the data quality. To the left in the dashboard, a configuration menu is presented. In this menu, the user can customise the algorithm to fit the needs of the organisation.

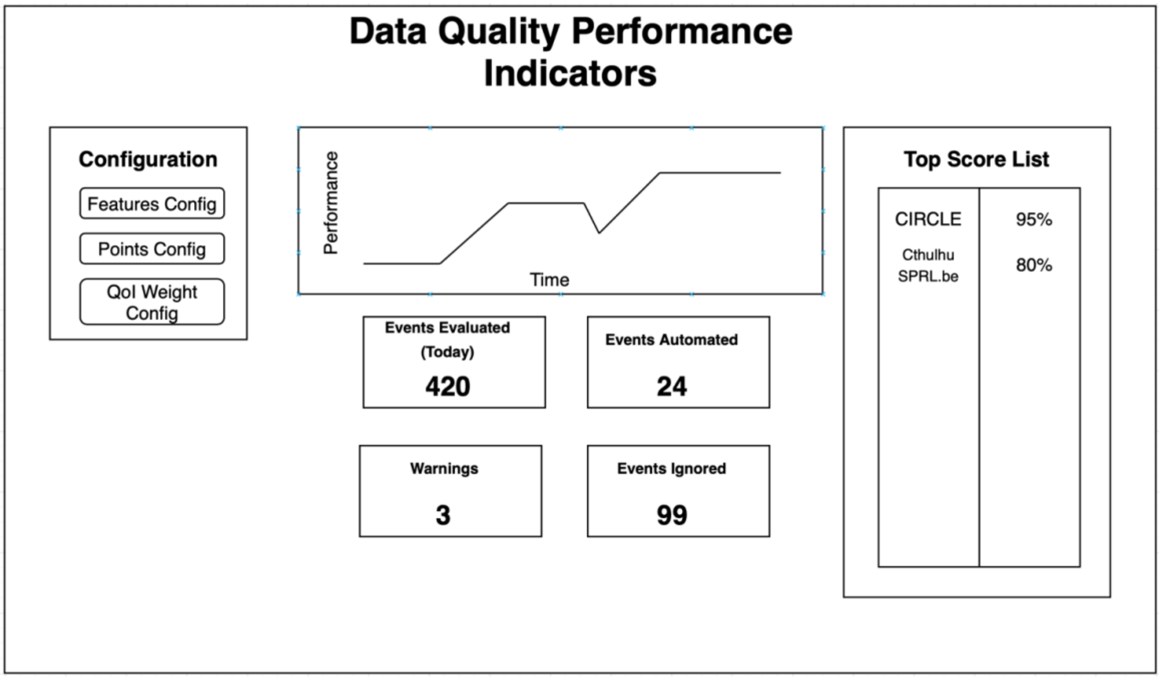


Figure 20 - The main dashboard, where performance indicators are presented to the user.

To configure the event features the system must measure, the user should press the button ‘Features Config’. The user would then be presented with the following options (Figure 24). If the button ‘find features’ is clicked, the system would then search through the JSON structure and find all entities in it, and show them in the list ‘found features’. The users would then be able to drag-and-drop the features into the QoI buckets on the right, which represent how important they are. The enables customisation of the algorithm; if none were dragged, the system would then put no weights on the features.

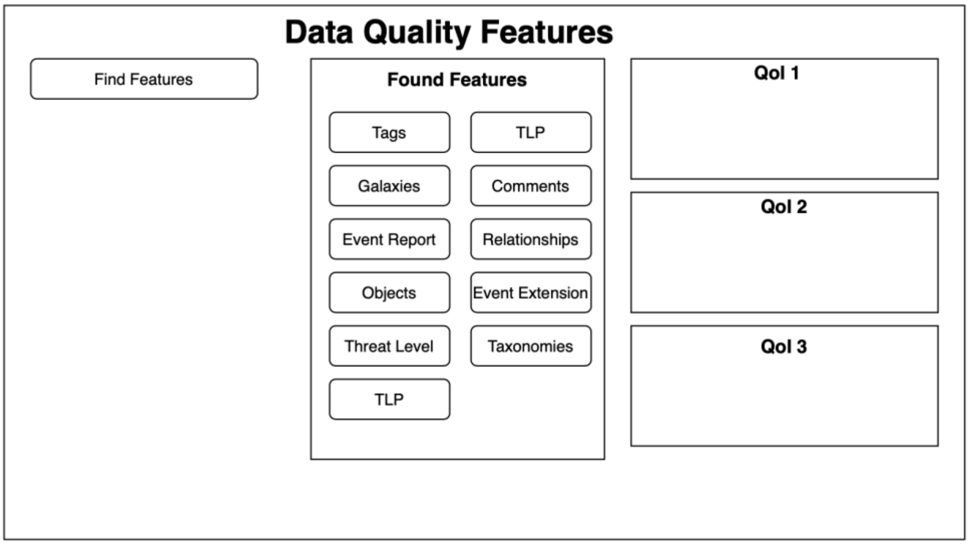
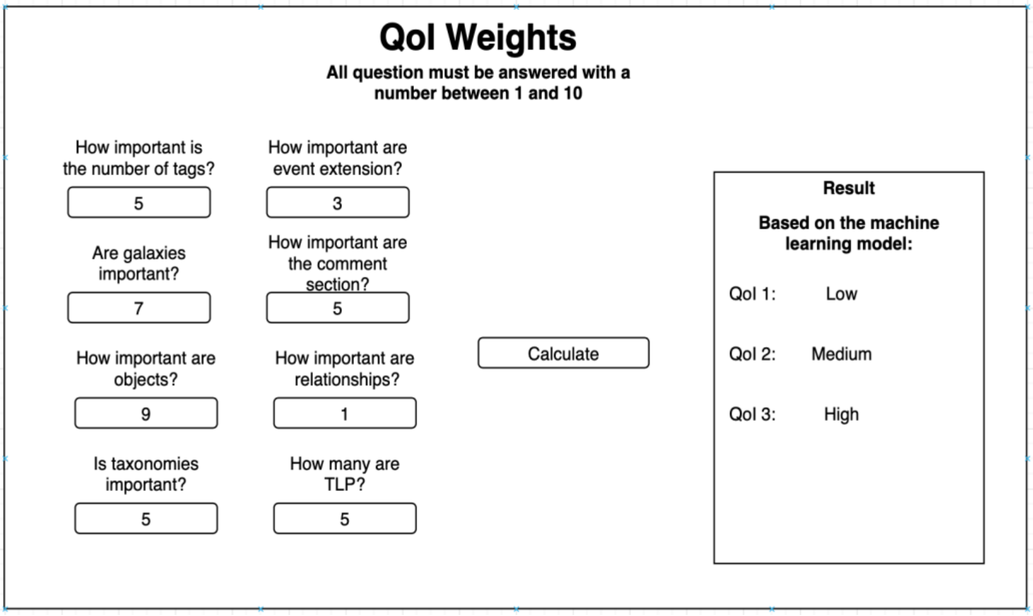


Figure 21 - A dashboard showing the events features which can be used to evaluate upon.

When all of the features have been divided into the different levels of QoI, their weights could be specified, either done manually or by using a machine learning algorithm. Figure 25, shows the features' weights dashboard; here, the user can manually set the weights by entering the weights' priorities. This is done in the right-most box. If the user has trouble deciding on the weights of the event features, a simple machine learning algortihm can help. The user is prompted to answer a series of questions. The user can then press 'calculate', and based on the answers, the system will propose a weighting scheme, which can be high, medium or low.



QoI Weights

All question must be answered with a number between 1 and 10

How important is the number of tags? 5

How important are event extension? 3

Are galaxies important? 7

How important are the comment section? 5

How important are objects? 9

How important are relationships? 1

Is taxonomies important? 5

How many are TLP? 5

Result

Based on the machine learning model:

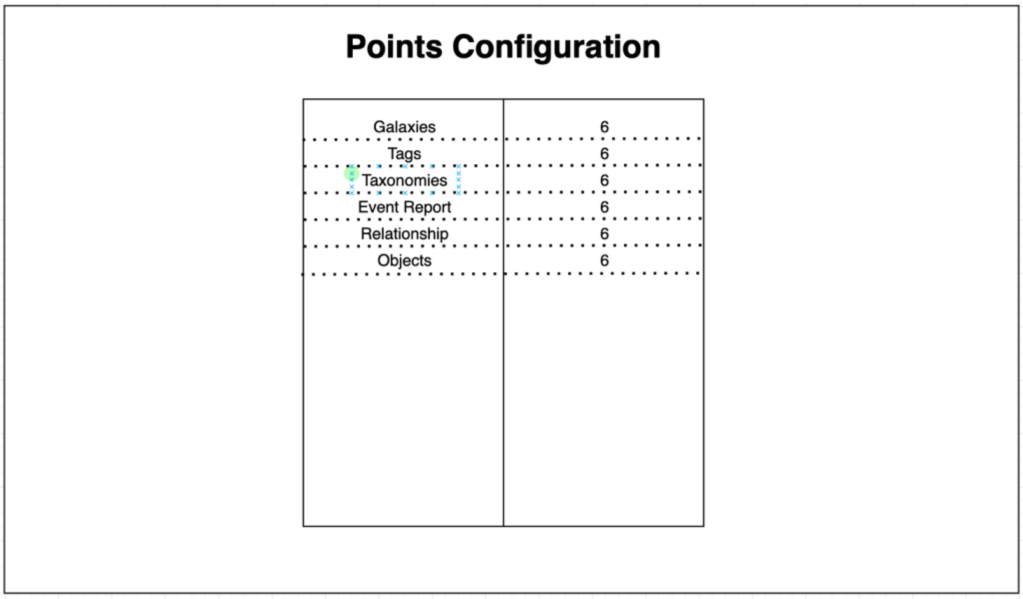
QoI 1: Low

QoI 2: Medium

QoI 3: High

Figure 22 - Dashboard where the user can either manually set the weight of the QoI, or use the ML algorithm to provide a suggestion.

The last option the user has on the main dashboard is 'Points Config' (Figure 26). This selection will present the user with a page where the point of each event feature can be specified; the user can also prioritise how much each feature should count towards the final confidence score. The user can set points between 1-10, where 10 is the highest amount of points, and should only be given if the feature is vital for the fitness of usage.



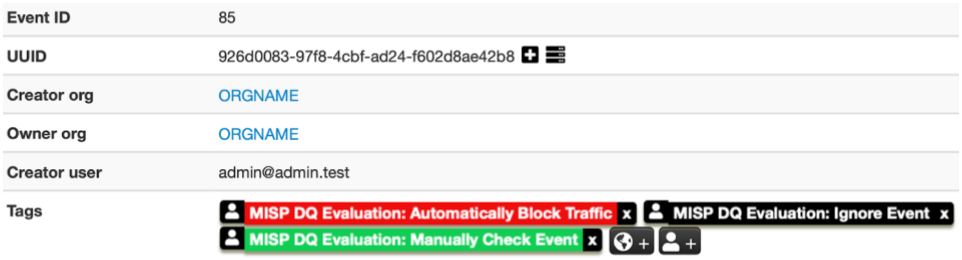
Points Configuration

Galaxies	6
Tags	6
Taxonomies	6
Event Report	6
Relationship	6
Objects	6

Figure 23 - The event feature points configuration menu, where the user can specify the amount of points each feature must give.

After all the measurements and calculations of the confidence score are completed, one essential discussed part of the system comes in play, which is the tagging mechanism. Figure 27 shows the final output of the system, namely the tag. When the user later inspects the event that has been evaluated, the event will have been tagged with one of three tags. The three tags will be:

- **Automatically Block Traffic**
- **Ignore Event**
- **Manually Check Event**



Event ID	85
UUID	926d0083-97f8-4cbf-ad24-f602d8ae42b8
Creator org	ORGNAME
Owner org	ORGNAME
Creator user	admin@admin.test
Tags	MISP DQ Evaluation: Automatically Block Traffic x MISP DQ Evaluation: Ignore Event x MISP DQ Evaluation: Manually Check Event x

Figure 24 - The three tags the system will be able to provide the threat event.

4.2 Architecture

A critical aspect of the system is the architecture, and it must be thought through so no time is wasted implementing solutions that are not viable. The analysis discussed many different technologies, why they are valuable and how they could be used. The requirement specification gathered them and outlined what the system must, could and won’t include. With the gained knowledge it is now possible to start designing a potential system, and a proof of concept.

Section 3.3.2, discusses the importance of knowing the current data flow of the MISP platform and how it can be improved. The conclusion is that a module must be inserted after the feed community, which should be capable of evaluating each and every threat event that enters the system. This must be done in real-time or as close to it as possible. In the previous mentioned section, the solution was just a black box, but with the knowledge gained throughout the project, the black box should now be investigated and settled upon.

Figure 28, shows a high-level model of how the calculating unit in the algorithm works. The figure shows the different QoI of importance, there are three degrees of features, where the 3rd Degree is considered to be the features that tell the most about the data quality of the threat event. The QoIs will be balanced using a weight, this approach makes it possible to give the QoI more weight towards the confidence score. The figure also presents a list of trusted organisations. This list will be continuously updated, as new events arrive to the MISP platform.

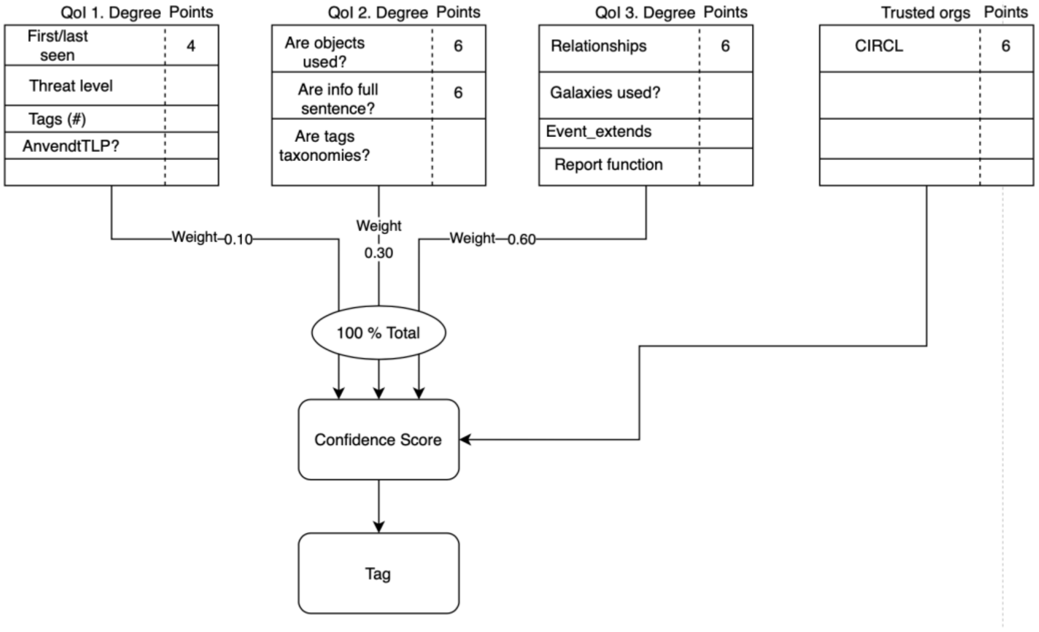


Figure 25 - The features and calculation data flow.

The confidence score mechanism is the heart of the algorithm and is discussed in Section 3.3.4. The concluded approach is a hybrid model, which is a combination of a weighted average mathematical model, and an anomaly detection algorithm. Figure 29 shows the final data model. A threat event will enter the system, immediately after a JSON analysis will begin, here the relevant features will be examined and stored in new variables.

The variables will then be evaluated based on the configuration the user has specified beforehand. The evaluation will then produce a confidence score, where the final output is a tag which will be attached to the threat event. However, another process is also started, each event evaluated will be stored in a separate CSV file: this will be done a prespecified amount of times, to get a comprehensive data set. When the CSV file contains enough data, the anomaly detection model will be activated, and hereafter the two models will compare results, and based on the comparison produce a tag for the event.

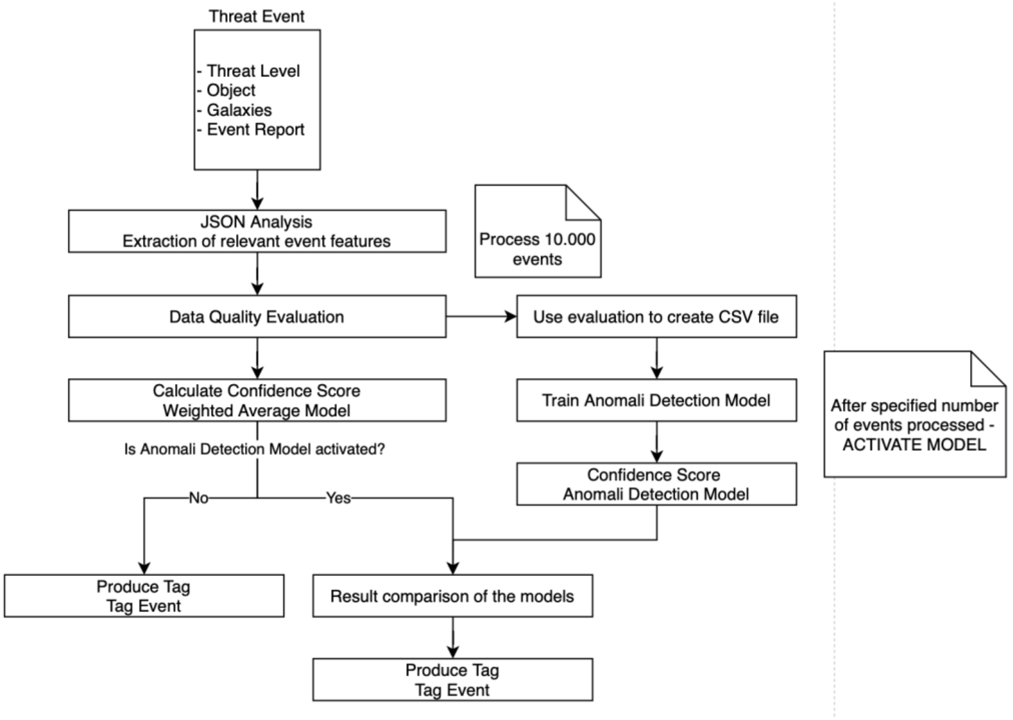

Threat Event
- Threat Level
- Object
- Galaxies
- Event Report
JSON Analysis
Extraction of relevant event features
Process 10.000 events
Data Quality Evaluation
Use evaluation to create CSV file
Calculate Confidence Score
Weighted Average Model
Train Anomali Detection Model
After specified number of events processed - ACTIVATE MODEL
Is Anomali Detection Model activated?
Confidence Score
Anomali Detection Model
No
Yes
Produce Tag
Tag Event
Result comparison of the models
Produce Tag
Tag Event

Figure 26 - Data flow of the calculation model, including the weighted average model and anomaly detection.

Figure 30 shows the full architecture of the entire developed system landscape, it illustrates its full setup and data flow of the system. The included components are feed providers, which generates events, the MISP project, that delegates the events to the correct subscribers, the ZMQs, which are the message queues and the evaluation module created in this project and visualised in Figure 29. This is considered to be the main contribution to the MISP platform, and is believed to be valuable to the users of the MISP instance, in terms of decision making, and enhancing the trust in the community.

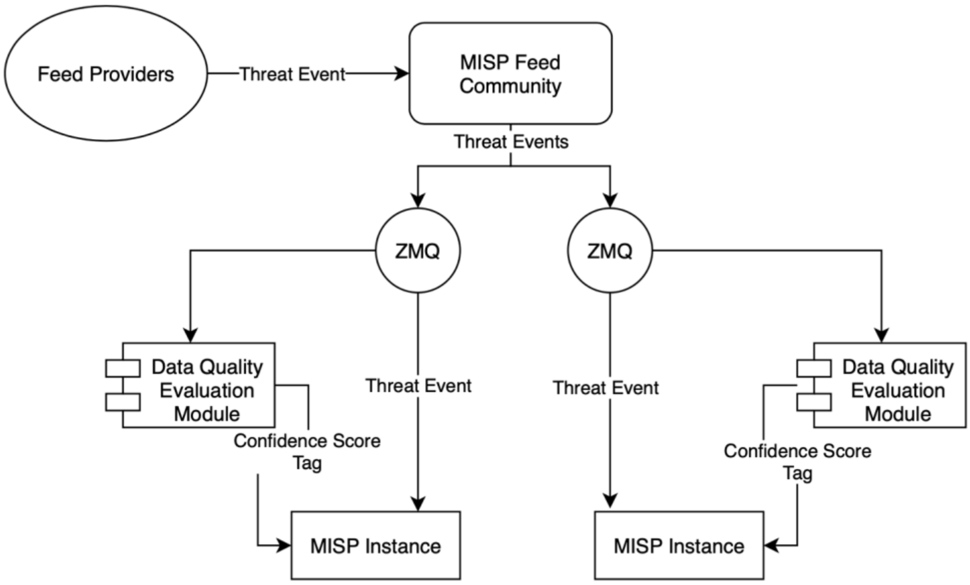


Figure 27 - The main contribution to the MISP community, including developed evaluation model.

4.2.1 Sequence Diagrams

This brief subsection will present a couple of different sequence diagrams to give a broader understanding of the processes that take place during the run-time of the system. Sequence diagrams are a tool to understand abstract systems because they give knowledge about in what sequence the data is flowing. Furthermore, they tell what actors and objects exist in the system and how they interact.

The first sequence diagram to be introduced illustrates the entire MISP project. Feed providers will investigate and research attacks and threats to alarm other people; they generate an event in the MISP instance. They will then publish that event to the community. Everybody with a MISP instance can subscribe to specific feed providers that they find trustworthy. When a feed provider publishes the event, everybody who has subscribed to them will be notified and can see the newly created event and its information. If the organisation that runs the MISP instance has the resources, they would have hired a network forensic, whose job is to evaluate the threat and come up with an actionable decision; this could be to block the network traffic.

Figure 31 illustrates the system landscape of the MISP project.

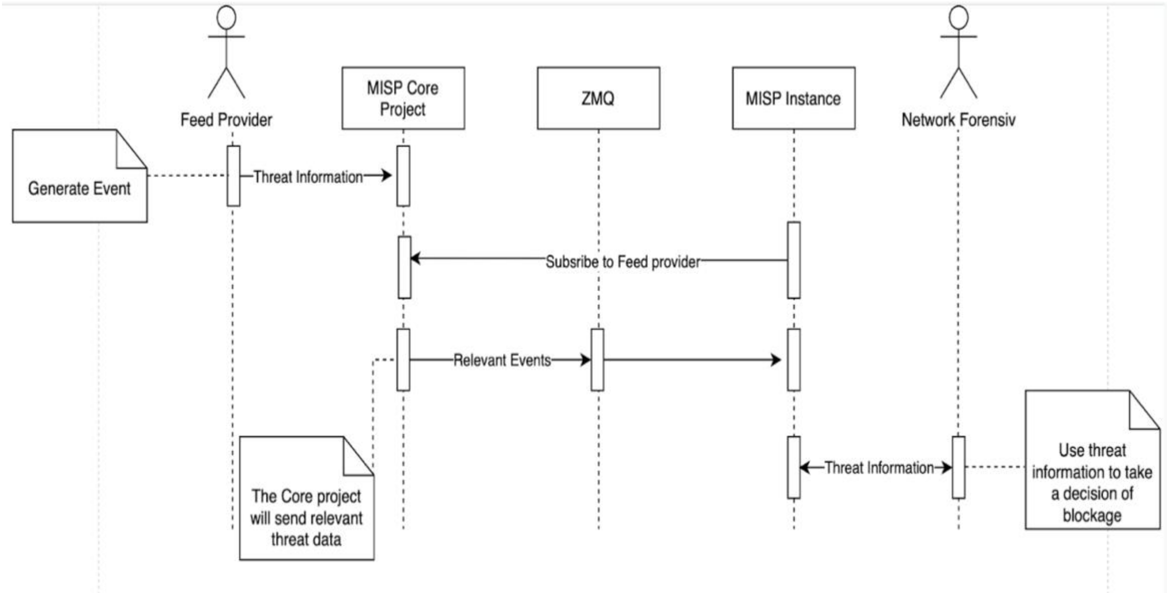


Figure 28 - Sequence diagram of the current MISP landscape.

The contribution to the MISP community is illustrated in Figure 32. As discussed in Section 2.1, a data quality evaluation model to help forensics using the MISP platform is needed, and it will be added just after the ZMQ in the model. The data quality evaluation model will intercept all events going into the ZMQ. The model can then extract the correct data and base the evaluation on the event features specified by the user in the configuration file. The user will also specify three different threshold values. In the case below, the user has specified that event with a rating below 70% should be ignored since the data quality is so low that no context is provided. Events rated between 70% and 90% should be manually evaluated by a forensic to decide upon an action to the threat. The last threshold includes events that are rated above 90%; these events should be clean and have so much context that the system securely should block the traffic without creating any further complications. The threshold values are not validated, they are visualised for demonstration purposes.

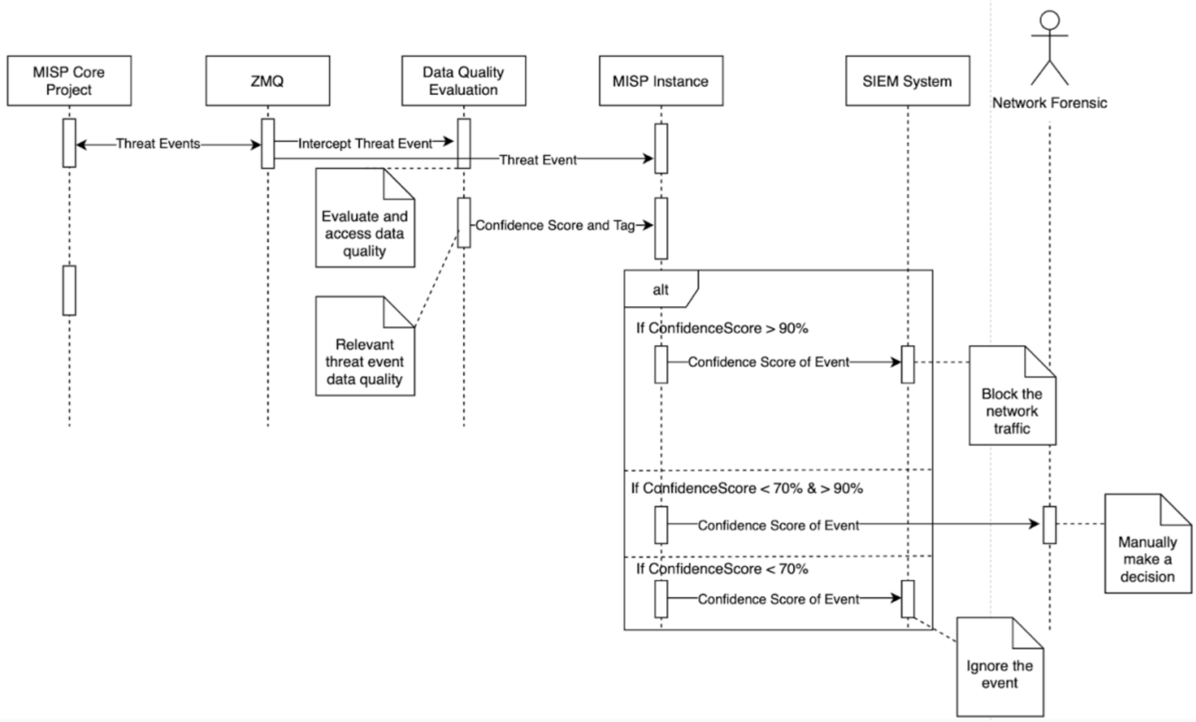


Figure 29 - Sequence diagram of the processes in the evaluation model.

5 Implementation

This chapter will describe the implementation phase of the project. This is the phase, where the actual code for the PoC is written. It is divided in two parts: the initial setup and the coding of the proposed solution.

5.1 Initial setup

Before being able to start writing the code for the proposed system, an initial setup must be done. As stated in the requirements specification (Section 3.4), the system must have a running MISP instance, b to get access to the threat events along with the installation of instance the ZMQ is also installed. To fetch the threat event the ZMQ must be subscribed to: how that's done will be explained later.

The MISP instance must be installed on a Linux based entity, and preferably accessible everywhere, therefore it was decided to host the MISP instance on a cloud solution. The cloud solution that was chosen is DigitalOcean.com. On the cloud market, there exist many actors, such as Amazon, Microsoft, Google, etc. However, the one chosen in this project is DigitalOcean. That is because of the very easy setup of the solution, and relatively cheap pricing. Another argument for using DigitalOcean is, that it was recommended by Dennis Rand as being intuitive and stable. When the server was configured and running, the MISP instance could be installed; this was done by following some relatively simple tutorials.

5.2 System coding

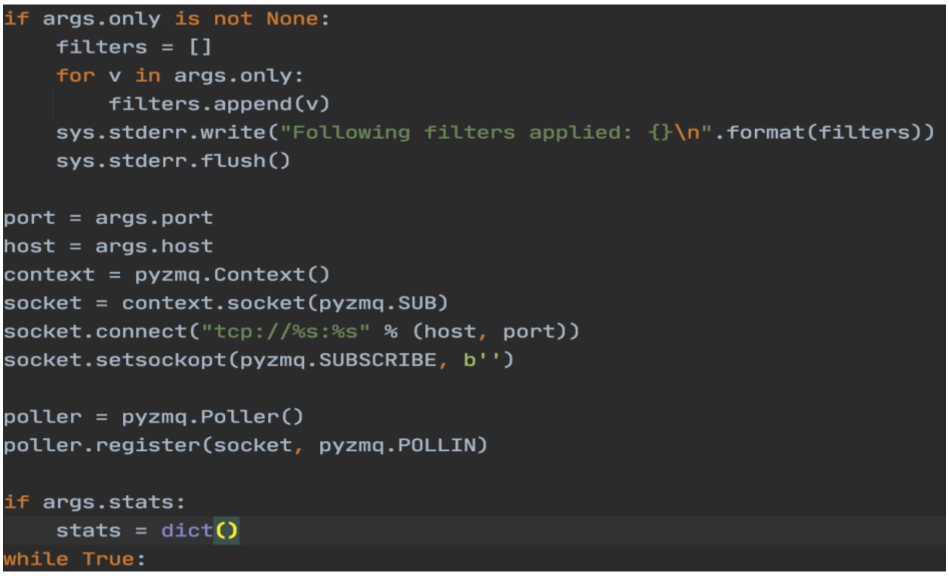
This subchapter will explain what was done in the actual implementation of the system. The most central parts of the code will be explained and visualised with screenshots. This subchapter heavily refers to the requirement specification and the MoSCoW prioritisation (Section 3.4). All of the requirements marked with the M, must be implemented, this is due to the critical function that they have, and without them the system would not function. The coding work done by the author, except one of the examples, which is clearly marked and cited.

System Specification

The system is coded in Python, the current and used version is Python 3.8. The decision of using Python is very dictated, since the coding of the MISP instance is done in Python, and the API client, PyMISP, is also Python based. To upload the Python script, containing the code, to the server, FireZilla is used. FireZilla is a software capable of uploading files to a Servia via FTP (File Transfer Protocol).

ZMQ

Before being able to evaluate the threat events coming to the MISP platform, a connection to the MISP instance's ZMQ must be established. A few examples are presented in the documentation of MISP platform, and these served as a guide for the setup. The setup is relatively simple. Port and host of the server are specified for the socket TCP connection. Also, the channel that must be listened to is specified. This PoC is listening on channel "B", which is a channel created by the MISP. After the connection has been established, a 'while true' loop is initiated. The loop will continue executing until the user stops it. Figure 33 shows the specific code.

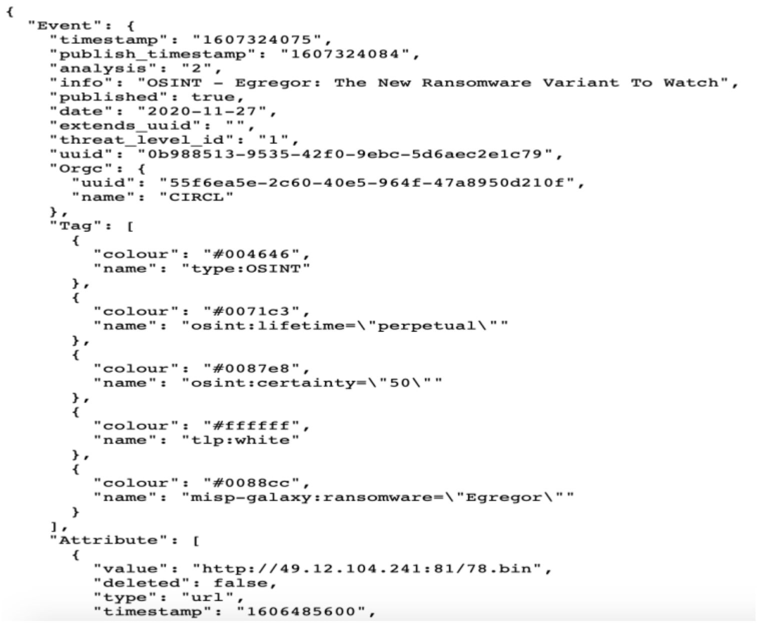


```
if args.only is not None:
    filters = []
    for v in args.only:
        filters.append(v)
    sys.stderr.write("Following filters applied: {}\n".format(filters))
    sys.stderr.flush()

port = args.port
host = args.host
context = pyzmq.Context()
socket = context.socket(pyzmq.SUB)
socket.connect("tcp://%s:%s" % (host, port))
socket.setsockopt(pyzmq.SUBSCRIBE, b'')

poller = pyzmq.Poller()
poller.register(socket, pyzmq.POLLIN)

if args.stats:
    stats = dict()
while True:
```

Figure 30 - Python code. Setup of connection to the ZMQ.

Extracting features

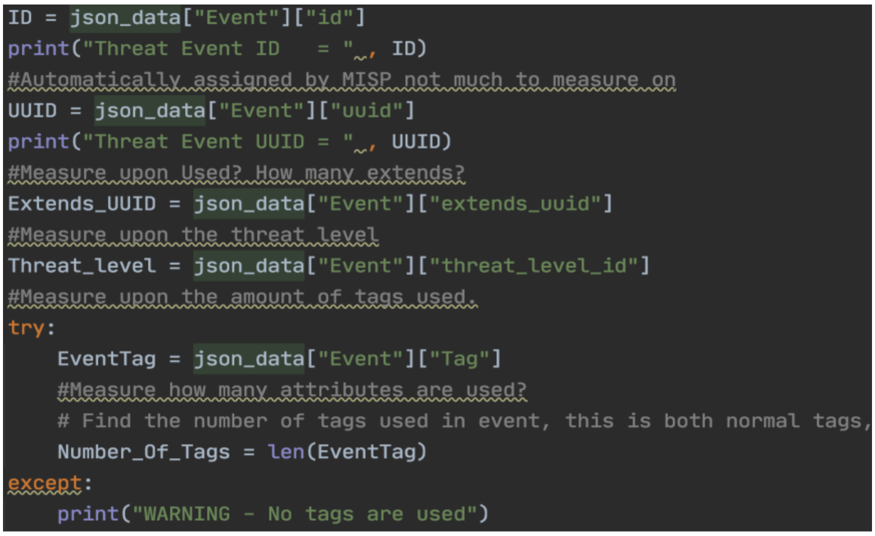
After the subscription to the ZMQ was established, listening for threat events could begin. As the threat events enter the MISP, different event features must be extracted. The relevant features that are deemed suitable are described in Section 3.3.3. To extract features from the event, it is important to understand its format and structure. The format of the events are a JSON structures, and all of the events follows the same structure and naming convention. Figure 34 shows a short example of the JSON structure of a single event and the naming of the different features.



```
{
  "Event": {
    "timestamp": "1607324075",
    "publish_timestamp": "1607324084",
    "analysis": "2",
    "info": "OSINT - Egregor: The New Ransomware Variant To Watch",
    "published": true,
    "date": "2020-11-27",
    "extends_uuid": "",
    "threat_level_id": "1",
    "uuid": "0b988513-9535-42f0-9ebc-5d6aec2e1c79",
    "Orgc": {
      "uuid": "55f6ea5e-2c60-40e5-964f-47a8950d210f",
      "name": "CIRCL"
    },
    "Tag": [
      {
        "colour": "#004646",
        "name": "type:OSINT"
      },
      {
        "colour": "#0071c3",
        "name": "osint:lifetime=\"perpetual\""
      },
      {
        "colour": "#0087e8",
        "name": "osint:certainty=\"50\""
      },
      {
        "colour": "#ffffff",
        "name": "tlp:white"
      },
      {
        "colour": "#0088cc",
        "name": "misp-galaxy:ransomware=\"Egregor\""
      }
    ],
    "Attribute": [
      {
        "value": "http://49.12.104.241:81/78.bin",
        "deleted": false,
        "type": "url",
        "timestamp": "1606485600",
```

Figure 31 – Shows part of a JSON structure and naming convention in a MISP threat event.

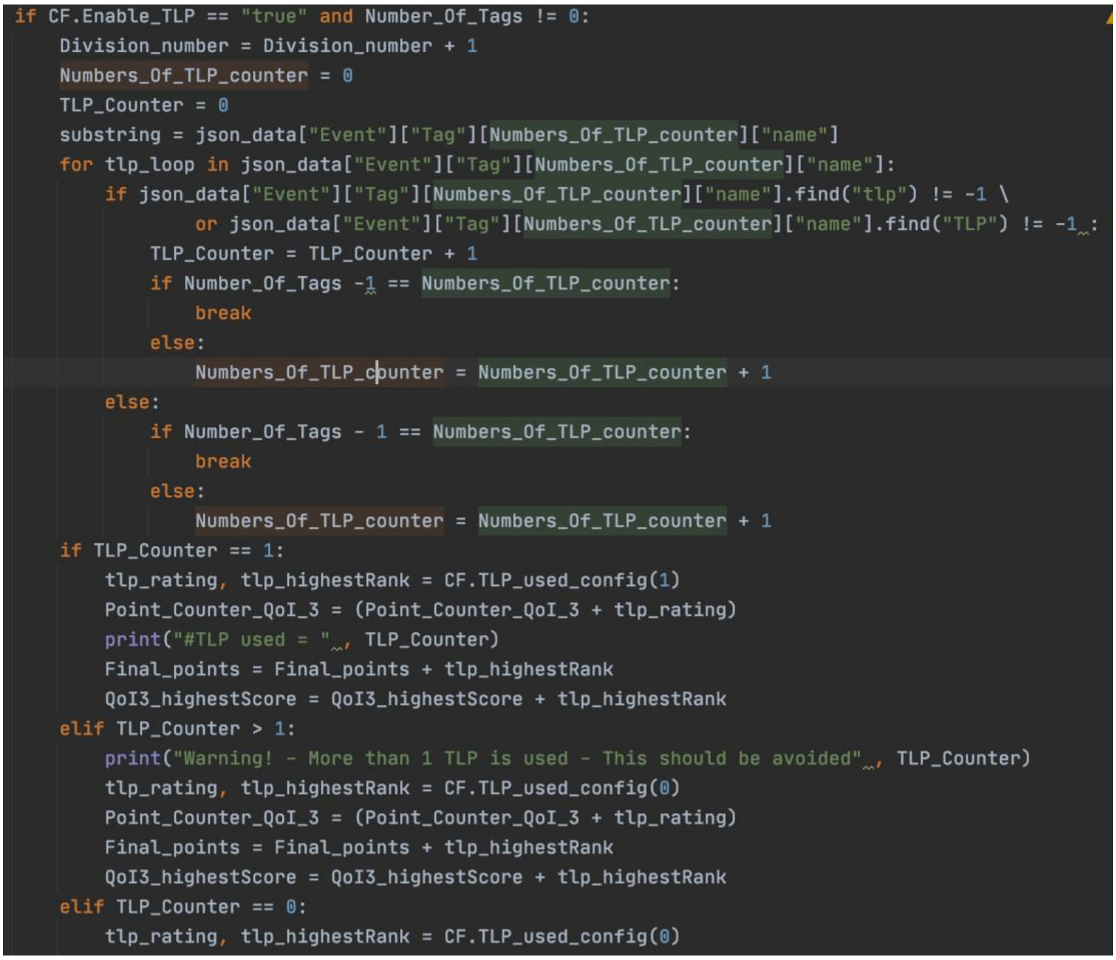
After the structure and name of the objects is known, a mechanism to extract the relevant features is created. This is done by the function JSON.load. This function converts the JSON structure into a Python dictionary, and hereafter the data can be extracted and stored in temporary variables. Figure 35 shows the mechanism: the system will issue a warning if the event has no tags attached - this is first and clear sign indicator that the event is of low quality.



```
ID = json_data["Event"]["id"]
print("Threat Event ID   = ", ID)
#Automatically assigned by MISP not much to measure on
UUID = json_data["Event"]["uuid"]
print("Threat Event UUID = ", UUID)
#Measure upon Used? How many extends?
Extends_UUID = json_data["Event"]["extends_uuid"]
#Measure upon the threat level
Threat_level = json_data["Event"]["threat_level_id"]
#Measure upon the amount of tags used.
try:
    EventTag = json_data["Event"]["Tag"]
    #Measure how many attributes are used?
    # Find the number of tags used in event, this is both normal tags,
    Number_Of_Tags = len(EventTag)
except:
    print("WARNING - No tags are used")
```

Figure 32 - The event features extraction mechanism.

The threat event features have now been fetched and stored. This allows for the main part of the system to be implemented, the data quality checks that will provide points (specified by the user) to the confidence score. A data quality check is done for each set of event features, and more features are evaluated with more than one check. As explained in Section 3.3.4, points, which are specified in a configuration file by the user, are given if a feature satisfies the data quality check. A threshold must be specified also by the user, and the points will be given based on to what extent the features fulfill the quality check.



```
if CF.Enable_TLP == "true" and Number_Of_Tags != 0:
    Division_number = Division_number + 1
    Numbers_Of_TLP_counter = 0
    TLP_Counter = 0
    substring = json_data["Event"]["Tag"][Numbers_Of_TLP_counter]["name"]
    for tlp_loop in json_data["Event"]["Tag"][Numbers_Of_TLP_counter]["name"]:
        if json_data["Event"]["Tag"][Numbers_Of_TLP_counter]["name"].find("tlp") != -1 \
                or json_data["Event"]["Tag"][Numbers_Of_TLP_counter]["name"].find("TLP") != -1 :
            TLP_Counter = TLP_Counter + 1
            if Number_Of_Tags -1 == Numbers_Of_TLP_counter:
                break
            else:
                Numbers_Of_TLP_counter = Numbers_Of_TLP_counter + 1
        else:
            if Number_Of_Tags - 1 == Numbers_Of_TLP_counter:
                break
            else:
                Numbers_Of_TLP_counter = Numbers_Of_TLP_counter + 1
    if TLP_Counter == 1:
        tlp_rating, tlp_highestRank = CF.TLP_used_config(1)
        Point_Counter_QoI_3 = (Point_Counter_QoI_3 + tlp_rating)
        print("#TLP used = ", TLP_Counter)
        Final_points = Final_points + tlp_highestRank
        QoI3_highestScore = QoI3_highestScore + tlp_highestRank
    elif TLP_Counter > 1:
        print("Warning! - More than 1 TLP is used - This should be avoided", TLP_Counter)
        tlp_rating, tlp_highestRank = CF.TLP_used_config(0)
        Point_Counter_QoI_3 = (Point_Counter_QoI_3 + tlp_rating)
        Final_points = Final_points + tlp_highestRank
        QoI3_highestScore = QoI3_highestScore + tlp_highestRank
    elif TLP_Counter == 0:
        tlp_rating, tlp_highestRank = CF.TLP_used_config(0)
```

Figure 33 - The data quality check for usage of TLP.

One of the more sophisticated data quality checks, and a good indicator of the data quality, is the usage of TLP (Section 2.1.2) in a threat event. The check is shown in Figure 36. The first action done, is to check whether the check has been enabled by the user. It will then, using a loop find the exact number of TLPs used in the threat event. TLP should always be used, however, only one TLP must be used for one event, therefore the exact number is checked. If the requirement is fulfilled, points are given to the confidence score, and if none or more multiple TLPs are used, no points will be given.

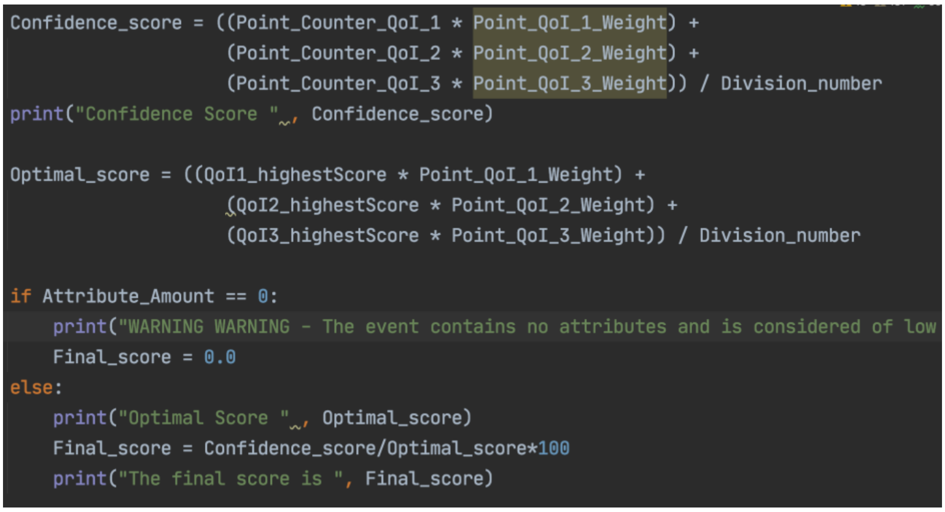
The structure of the quality checks is following the same patterns in most of the checks, with a few exceptions. The pattern is as follows:

1. It is enabled by the user.
2. Is the event feature used?
3. Fetch information about the points from the configuration file
4. Does the feature fulfill the requirement?
5. Add a correct number of points to the confidence score
6. Add the highest number of points to the optimal score

The exceptions are a false-positive check and a check if the tags are taxonomies. The false-positive check will compare the IoC to a predefined list of white-listed IPs and domains. The taxonomy check will take each tag used and compare it to prespecified lists and see if the tags is a taxonomy.

Rating Algorithm and Tag

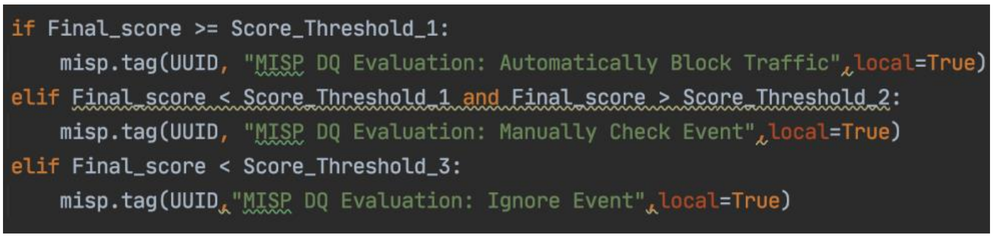
The rating algorithm is very central to the entire system. It is explained in Section 3.3.4, and in the requirements specification it is specified as ‘Must’ be implemented, due to its importance to the system. As described in the analysis, at this early stage of the project, implementing a well-functioning machine learning model is out of the scope, therefore only the mathematical calculation is implemented. This is done using by a weighted average model. Figure 37, shows how it is done. First, the confidence score is calculated. The confidence score is scored based on how many points the event got, by satisfying the previous mentioned data quality checks. The optimal score is then based on the highest possible amount of points that the event could get if all checks were fully satisfied. The percentage of how close the confidence score is to the optimal score is then calculated, and that number present the final score, that informs the user of how good the data quality of the event is.



```
Confidence_score = ((Point_Counter_QoI_1 * Point_QoI_1_Weight) +
                    (Point_Counter_QoI_2 * Point_QoI_2_Weight) +
                    (Point_Counter_QoI_3 * Point_QoI_3_Weight)) / Division_number
print("Confidence Score ", Confidence_score)

Optimal_score = ((QoI1_highestScore * Point_QoI_1_Weight) +
                 (QoI2_highestScore * Point_QoI_2_Weight) +
                 (QoI3_highestScore * Point_QoI_3_Weight)) / Division_number

if Attribute_Amount == 0:
    print("WARNING WARNING - The event contains no attributes and is considered of low
    Final_score = 0.0
else:
    print("Optimal Score ", Optimal_score)
    Final_score = Confidence_score/Optimal_score*100
    print("The final score is ", Final_score)
```

Figure 34 – The weighted average rating algorithm.

Based on the final score a tag to the event is provided. This tag is given by utilising the API client in the MISP platform, PyMISP (Section 2.1.5). The user can set a threshold for three categories of tags (Section 3.3.5). Figure 38 shows the coded mechanism. Here, the final score mentioned earlier, is compared to a threshold, specified in the configuration file by the user.

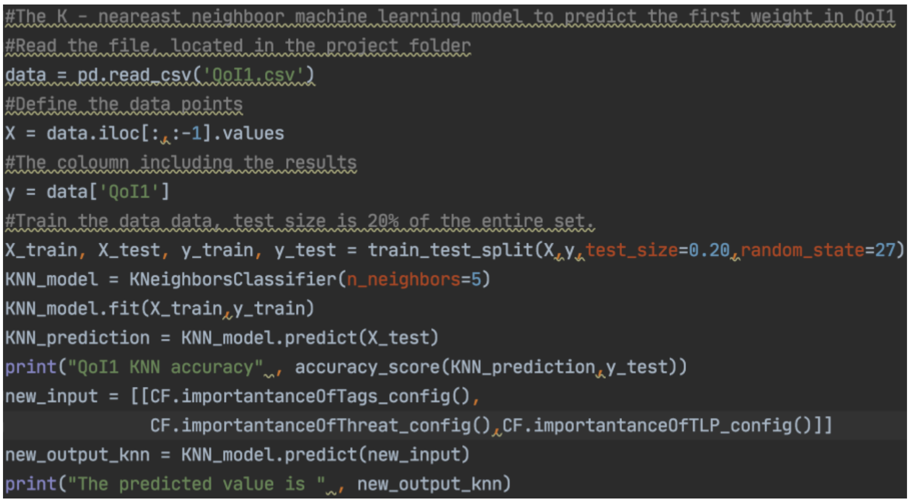


```
if Final_score >= Score_Threshold_1:
    misp.tag(UUID, "MISP DQ Evaluation: Automatically Block Traffic",local=True)
elif Final_score < Score_Threshold_1 and Final_score > Score_Threshold_2:
    misp.tag(UUID, "MISP DQ Evaluation: Manually Check Event",local=True)
elif Final_score < Score_Threshold_3:
    misp.tag(UUID,"MISP DQ Evaluation: Ignore Event",local=True)
```

Figure 35 - The tagging mechanism, and the three tag categories..

QoI Weight Determination

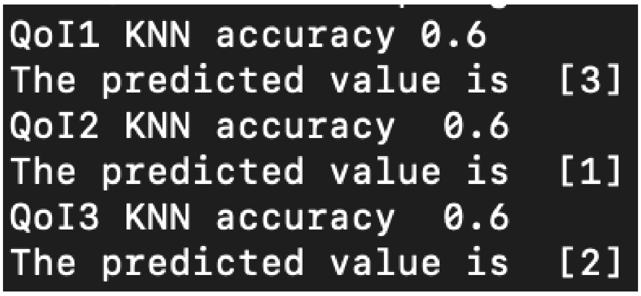
As mentioned in Section 3.3.3, it was decided that the system should allow the user to divide the data quality checks into QoI, which is three categories of importance. By doing that the user can weigh certain data quality features higher towards the final score. This option can also be turned off by the user, and all data quality checks will have the same level of importance. Setting the weights of the QoI, can be a tedious task and therefore implementing a simple guide for the user, was thought to be beneficial. The guidance would be a simple machine learning model, where the user would answer a few questions about the importance of for example tags and galaxies and based on the answers the model will propose the weights.. This simple code example of a machine learning model is the only piece of code, which is not created by the author. The code originates from [74], and then parameters are changed and modified to be used in this project.



```
#The K - neareast neighboor machine learning model to predict the first weight in QoI1
#Read the file, located in the project folder
data = pd.read_csv('QoI1.csv')
#Define the data points
X = data.iloc[:,:-1].values
#The coloumn including the results
y = data['QoI1']
#Train the data data, test size is 20% of the entire set.
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.20,random_state=27)
KNN_model = KNeighborsClassifier(n_neighbors=5)
KNN_model.fit(X_train,y_train)
KNN_prediction = KNN_model.predict(X_test)
print("QoI1 KNN accuracy", accuracy_score(KNN_prediction,y_test))
new_input = [[CF.importantanceOfTags_config(),
              CF.importantanceOfThreat_config(),CF.importantanceOfTLP_config()]]
new_output_knn = KNN_model.predict(new_input)
print("The predicted value is ", new_output_knn)
```

Figure 36 - The machine learning model, to determine weights.

Figure 39 shows the machine learning model. It was decided to use a classification KNN model (Section 2.3.6), simply because it provided the best accuracy. As discussed earlier, no datasets were available, therefore a very small and simple dataset is created by the author, to demonstrate the ML model.



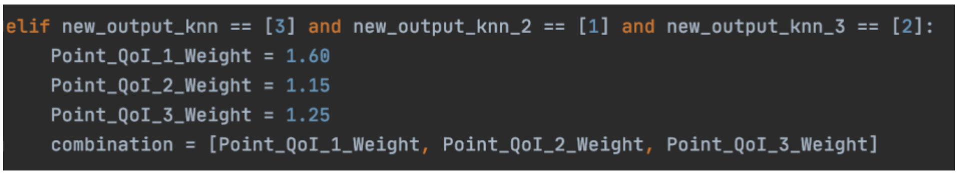
```
QoI1 KNN accuracy 0.6
The predicted value is  [3]
QoI2 KNN accuracy  0.6
The predicted value is  [1]
QoI3 KNN accuracy  0.6
The predicted value is  [2]
```

Figure 37 - Model Accuracy of QoI weight machine learning model.

After having trained the model with the dataset, new unknown data could be inserted into the model. Figure 40 shows the accuracy of the model, which is fairly consistent on every run, averaging at approximately 0.6 – 60%. The measured accuracy is relatively low. It is assumed that it is caused by the small dataset size and it is biased. A SVM (Support Vector Machine) mode was also tried, and it produced even lower accuracies.

The outcome of the ML model is a number: 1,2 or 3, representing how high the weight should be (1 – Low importance, 2 – Medium importance and 3 – High importance).

As seen in Figure 40, the model suggests, that the QoI1 should be weighted high, QoI2 should be weighted low and QoI3 should be weighted medium.



```
elif new_output_knn == [3] and new_output_knn_2 == [1] and new_output_knn_3 == [2]:
    Point_QoI_1_Weight = 1.60
    Point_QoI_2_Weight = 1.15
    Point_QoI_3_Weight = 1.25
    combination = [Point_QoI_1_Weight, Point_QoI_2_Weight, Point_QoI_3_Weight]
```

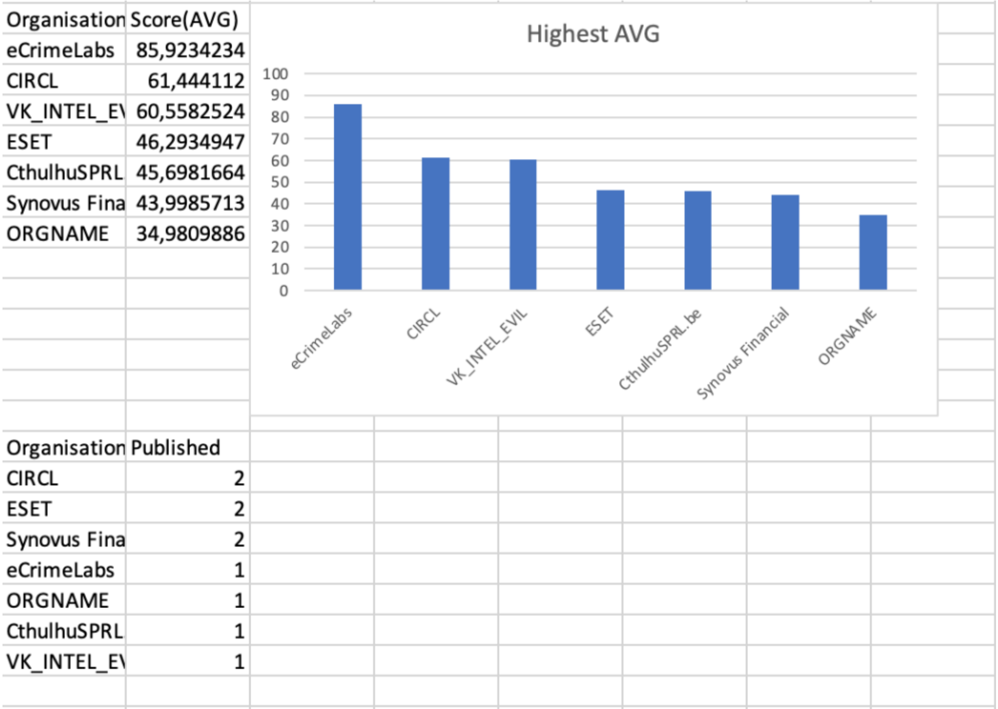
Figure 38 - The outcome of the ML model, and the weights.

The outcome is then used to set the weight, as seen in Figure 41. The code includes a list of the different combinations, and their belonging weights. This means, that if the example from Figure 40 is investigated, the weight would be as shown in Figure 41.

The Reputation system

The last part of the system, and which is one of the main contributions of this project, is the reputation system, the list of rated organisations based on their general data quality. Being placed high on the list gives a real incentive to produce high quality data, while being placed low should produce a collective sanction by avoiding using the data.

The reputation list is essentially just a list where the average of all the final scores is gathered. This could be achieved using several different approaches. The approach selected for this PoC was to use Microsoft Excel. It is implemented into the code in such a way, that every time a new event has been evaluated, the system would put the result into the Excel file. The Excel option is chosen because it is a widely known format, it is free, and its statistically analysis are very advanced.



Organisation	Score(AVG)
eCrimeLabs	85,9234234
CIRCL	61,444112
VK_INTEL_E\	60,5582524
ESET	46,2934947
CthulhuSPRL	45,6981664
Synovus Fina	43,9985713
ORGNAME	34,9809886

Organisation	Published
CIRCL	2
ESET	2
Synovus Fina	2
eCrimeLabs	1
ORGNAME	1
CthulhuSPRL	1
VK_INTEL_E\	1

Figure 39 - Reputation system and some statistical analysis.

Figure 42 shows the outcome of the evaluation - the reputation mechanism. The figure shows the rated list of feed providers and a statistical analysis of them. It is believed that if all users of the MISP, had such list, trust would be more easily obtained because it will give an insight into the previous data delivered. Users could easily navigate around feed providers that consistently delivers poor-quality data.

6 Testing

As concluded in the implementation chapter, the goal of creating a PoC is achieved. This brief subchapter will be conducting a test session, to see if the PoC can produce reliable results, and it is free of errors and bugs.

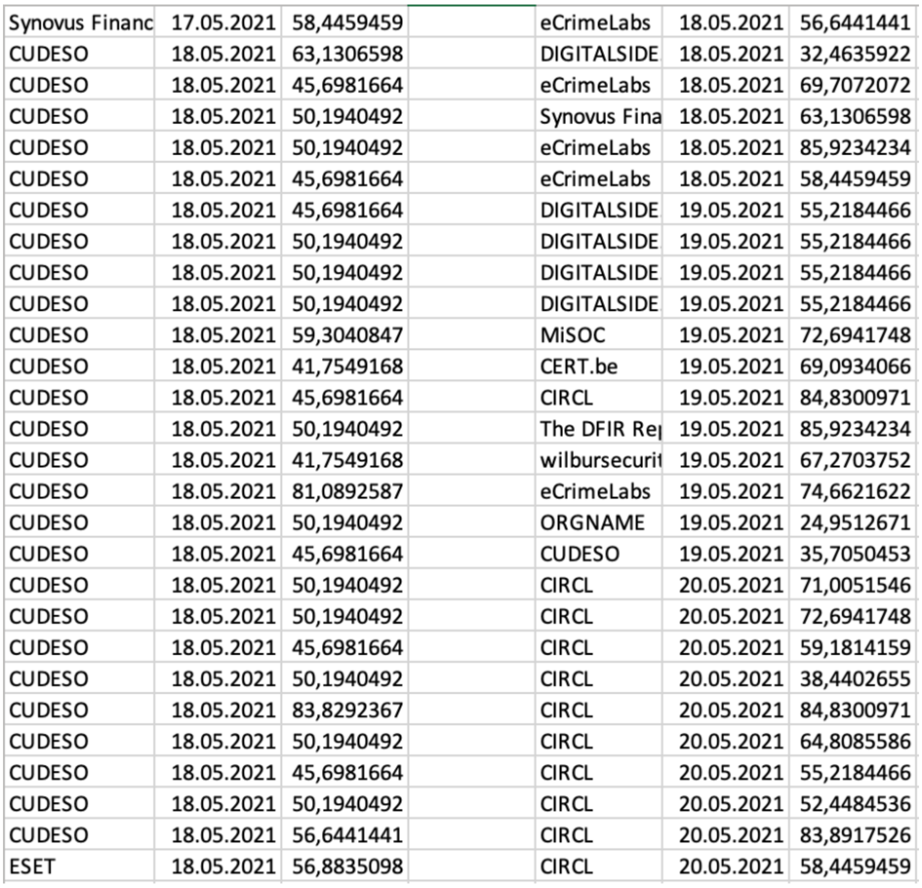
Stress Testing (three days)

The test to be conducted is a stress test. It will be conducted over the span of three days. The number of days is chosen, because it is assumed in that time span enough events will be published by the feed providers. The test will be conducted in collaboration with the MISP expert Dennis Rand, and he has granted access to his private MISP instance, so the proposed system can evaluate a list of threat events published by eCrimeLabs. The events from eCrimeLabs are considered to be of high quality, and therefore these will be compared to the result of test.

The assumption and hypothesis are that a vast number of events will be published, and the data quality of those will be very fluctuating, and to the lower side. Before starting the test, all available feed providers in the community were subscribed to, this was done to evaluate as many events as possible. Furthermore, the Python script was launched on the server as a background process so it wouldn't stop even when the terminal was closed.

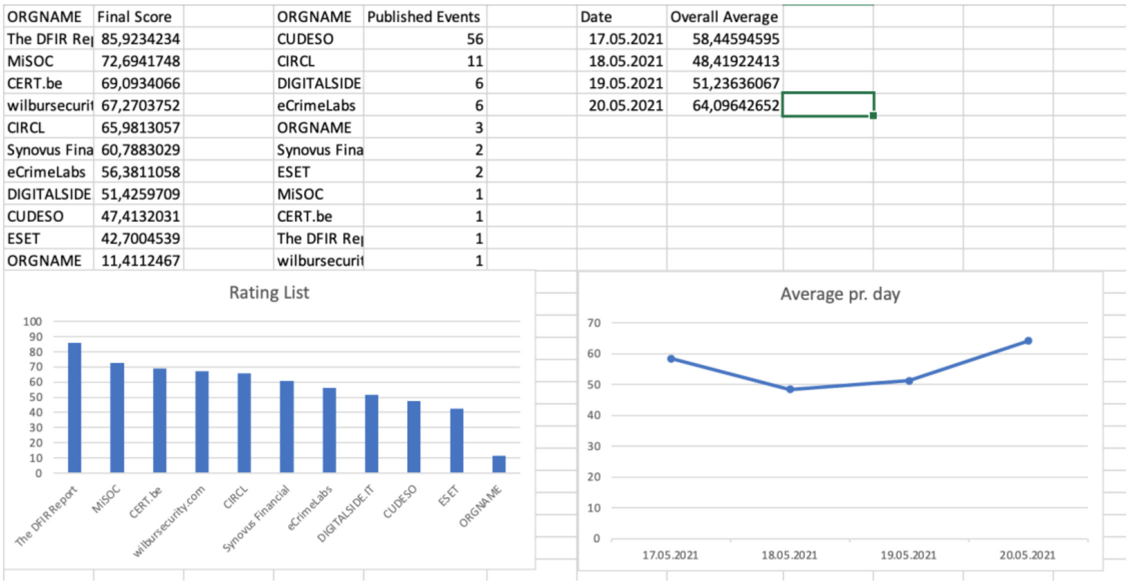
After a period of approximately three days, the background progress on the server was killed, and the script would no longer evaluate the events. As explained in Section 5.2, the events are stored in an Excel file, where further analyse will be done.

Figure 43 shows a small sample of the evaluated events (the total amount was 90 events), the left-most columns represent the feed provider name, the next column is the date for evaluation and the last is the final score. This list is then further analysed using a few statistical methods, such as rating, average per day, and a list presenting the number of events published, this can be seen in Figure 44.



Synovus Financ	17.05.2021	58,4459459		eCrimeLabs	18.05.2021	56,6441441
CUDESO	18.05.2021	63,1306598		DIGITALSIDE	18.05.2021	32,4635922
CUDESO	18.05.2021	45,6981664		eCrimeLabs	18.05.2021	69,7072072
CUDESO	18.05.2021	50,1940492		Synovus Fina	18.05.2021	63,1306598
CUDESO	18.05.2021	50,1940492		eCrimeLabs	18.05.2021	85,9234234
CUDESO	18.05.2021	45,6981664		eCrimeLabs	18.05.2021	58,4459459
CUDESO	18.05.2021	45,6981664		DIGITALSIDE	19.05.2021	55,2184466
CUDESO	18.05.2021	50,1940492		DIGITALSIDE	19.05.2021	55,2184466
CUDESO	18.05.2021	50,1940492		DIGITALSIDE	19.05.2021	55,2184466
CUDESO	18.05.2021	50,1940492		DIGITALSIDE	19.05.2021	55,2184466
CUDESO	18.05.2021	59,3040847		MiSOC	19.05.2021	72,6941748
CUDESO	18.05.2021	41,7549168		CERT.be	19.05.2021	69,0934066
CUDESO	18.05.2021	45,6981664		CIRCL	19.05.2021	84,8300971
CUDESO	18.05.2021	50,1940492		The DFIR Re	19.05.2021	85,9234234
CUDESO	18.05.2021	41,7549168		wilbursecurit	19.05.2021	67,2703752
CUDESO	18.05.2021	81,0892587		eCrimeLabs	19.05.2021	74,6621622
CUDESO	18.05.2021	50,1940492		ORGNAME	19.05.2021	24,9512671
CUDESO	18.05.2021	45,6981664		CUDESO	19.05.2021	35,7050453
CUDESO	18.05.2021	50,1940492		CIRCL	20.05.2021	71,0051546
CUDESO	18.05.2021	50,1940492		CIRCL	20.05.2021	72,6941748
CUDESO	18.05.2021	45,6981664		CIRCL	20.05.2021	59,1814159
CUDESO	18.05.2021	50,1940492		CIRCL	20.05.2021	38,4402655
CUDESO	18.05.2021	83,8292367		CIRCL	20.05.2021	84,8300971
CUDESO	18.05.2021	50,1940492		CIRCL	20.05.2021	64,8085586
CUDESO	18.05.2021	45,6981664		CIRCL	20.05.2021	55,2184466
CUDESO	18.05.2021	50,1940492		CIRCL	20.05.2021	52,4484536
CUDESO	18.05.2021	56,6441441		CIRCL	20.05.2021	83,8917526
ESET	18.05.2021	56,8835098		CIRCL	20.05.2021	58,4459459

Figure 40 - A sample of the evaluated threat events.

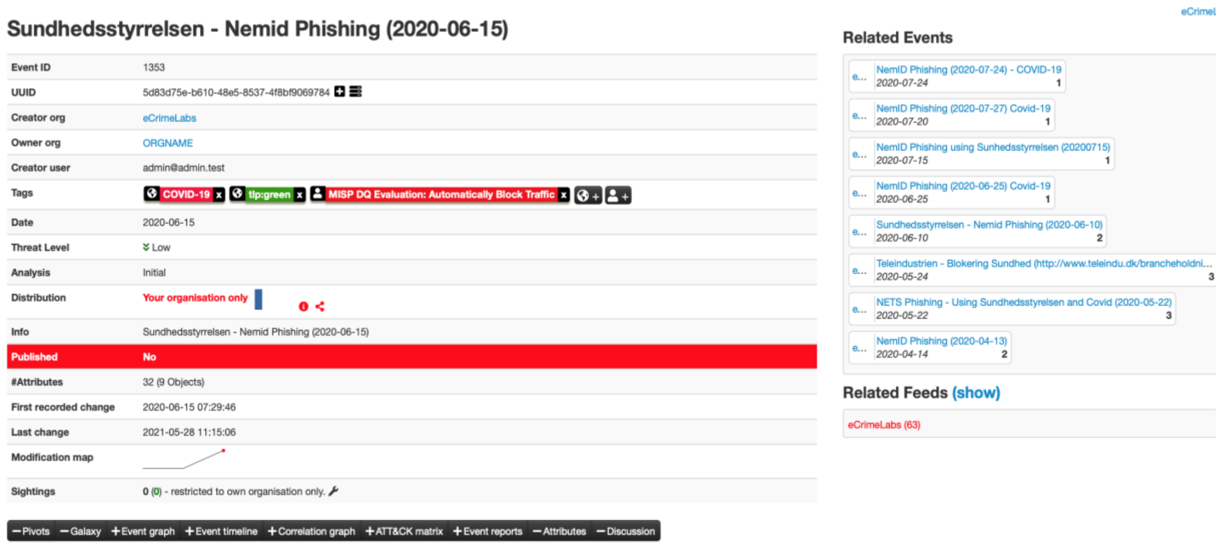


ORGNAME	Final Score		ORGNAME	Published Events		Date	Overall Average
The DFIR Re	85,9234234		CUDESO	56		17.05.2021	58,44594595
MiSOC	72,6941748		CIRCL	11		18.05.2021	48,41922413
CERT.be	69,0934066		DIGITALSIDE	6		19.05.2021	51,23636067
wilbursecurit	67,2703752		eCrimeLabs	6		20.05.2021	64,09642652
CIRCL	65,9813057		ORGNAME	3			
Synovus Fina	60,7883029		Synovus Fina	2			
eCrimeLabs	56,3811058		ESET	2			
DIGITALSIDE	51,4259709		MiSOC	1			
CUDESO	47,4132031		CERT.be	1			
ESET	42,7004539		The DFIR Re	1			
ORGNAME	11,4112467		wilbursecurit	1			

Figure 41 - Key Performance Indicators of evaluated threat events.

Validation of Test

The last part of the testing session is to validate the results of the test, this will be done by taking one threat event that score high, and one that scores low. As mentioned eCrimeLabs is considered to deliver high quality events, and therefore to find an event that score, an eCrimeLabs event is chosen. The event ID is 1353, and is about phishing email, the event is depicted in Figure 46.



Sundhedsstyrrelsen - Nemid Phishing (2020-06-15)

Event ID	1353
UUID	5d83d75e-b610-48e5-8537-4f8bf9069784
Creator org	eCrimeLabs
Owner org	ORGNAME
Creator user	admin@admin.test
Tags	COVID-19 x · tlp:green x · MISP DQ Evaluation: Automatically Block Traffic x
Date	2020-06-15
Threat Level	Low
Analysis	Initial
Distribution	Your organisation only
Info	Sundhedsstyrrelsen - Nemid Phishing (2020-06-15)
Published	No
#Attributes	32 (9 Objects)
First recorded change	2020-06-15 07:29:46
Last change	2021-05-28 11:15:06
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Related Events

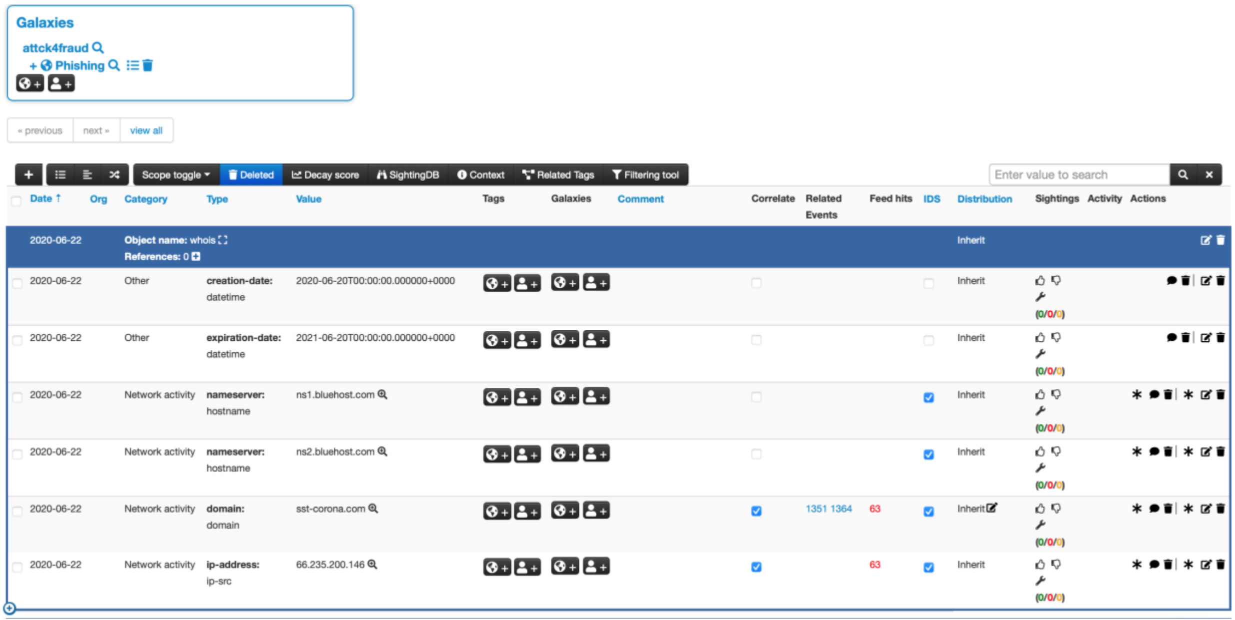
e...	NemID Phishing (2020-07-24) - COVID-19 2020-07-24	1
e...	NemID Phishing (2020-07-27) Covid-19 2020-07-20	1
e...	NemID Phishing using Sunhedsstyrrelsen (20200715) 2020-07-15	1
e...	NemID Phishing (2020-06-25) Covid-19 2020-06-25	1
e...	Sundhedsstyrrelsen - Nemid Phishing (2020-06-10) 2020-06-10	2
e...	Teleindustrien - Blokering Sundhed (http://www.teleindu.dk/brancheholdni... 2020-05-24	3
e...	NETS Phishing - Using Sundhedsstyrelsen and Covid (2020-05-22) 2020-05-22	3
e...	NemID Phishing (2020-04-13) 2020-04-14	2

Related Feeds (show)

eCrimeLabs (63)

Figure 42 - eCrimeLabs event, that has a high confidence score

The figure shows that the event has three tags, including the tag given by the PoC and only one TLP, it has many attributes, many related events, and the information field is filled. Figure 47 is the next part of the event, where the attributes are shown. It is clear that they are divided into objects, and no attributes are stand-alone. Furthermore, galaxies are used which is also good practice.



Galaxies

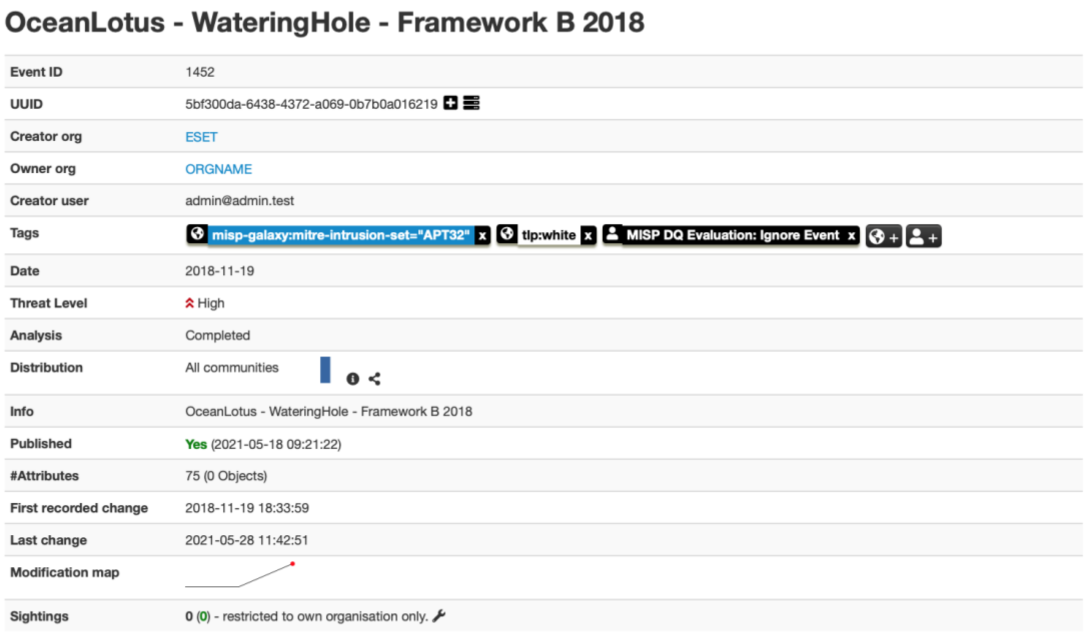
attck4fraud
+ Phishing

Date ↑	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events	Feed hits	IDS	Distribution	Sightings	Activity	Actions
2020-06-22		Object name: whois References: 0										Inherit			
2020-06-22		Other	creation-date: datetime	2020-06-20T00:00:00.000000+0000				□			□	Inherit	(0/0/0)		
2020-06-22		Other	expiration-date: datetime	2021-06-20T00:00:00.000000+0000				□			□	Inherit	(0/0/0)		
2020-06-22		Network activity	nameserver: hostname	ns1.bluehost.com				□			☑	Inherit	(0/0/0)		
2020-06-22		Network activity	nameserver: hostname	ns2.bluehost.com				□			☑	Inherit	(0/0/0)		
2020-06-22		Network activity	domain: domain	sst-corona.com				☑	1351 1364	63	☑	Inherit	(0/0/0)		
2020-06-22		Network activity	ip-address: ip-src	66.235.200.146				☑		63	☑	Inherit	(0/0/0)		

Figure 43 - eCrimeLabs event, that has a high confidence score.

The event has a confidence score of approximately 70%, which compared to the community is a high score. Should the event score even higher, features should event report, event extension, and taxonomy tags should be used.

The next event is an event that score low, and therefore must be considered of low data quality. It is an event coming from the feed provider Eset and has ID 1452. The event missed some of the important features to fill before publishing. Figure 48 shows that the event has tags and information included, but it has no event extension, event report, galaxy.

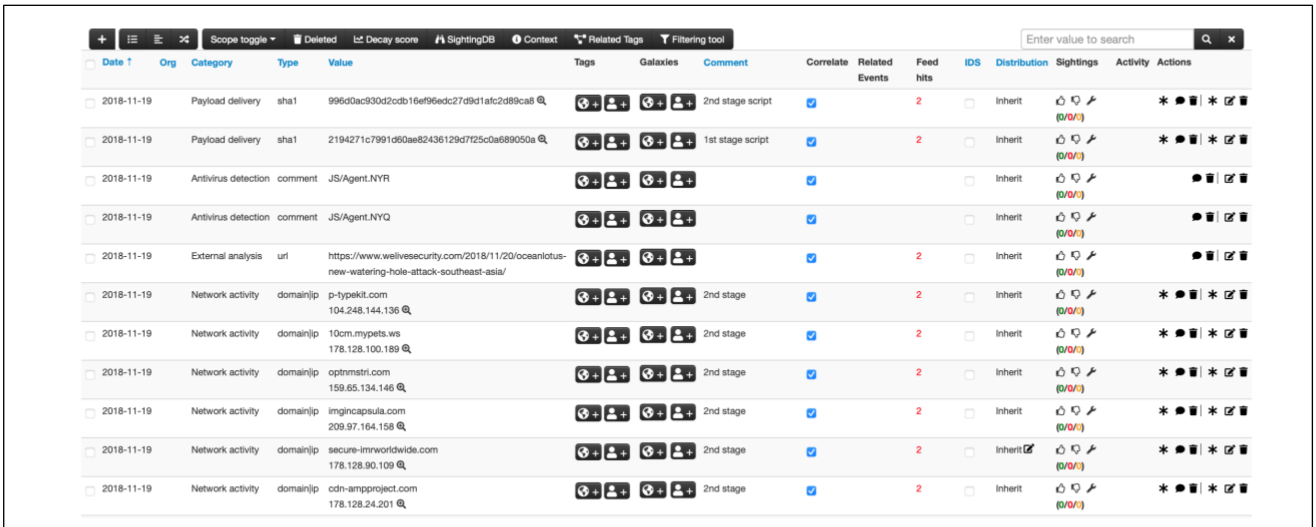


OceanLotus - WateringHole - Framework B 2018

Event ID	1452
UUID	5bf300da-6438-4372-a069-0b7b0a016219
Creator org	ESET
Owner org	ORGNAME
Creator user	admin@admin.test
Tags	misp-galaxy:mitre-intrusion-set="APT32" x tlp:white x MISP DQ Evaluation: Ignore Event x + +
Date	2018-11-19
Threat Level	High
Analysis	Completed
Distribution	All communities
Info	OceanLotus - WateringHole - Framework B 2018
Published	Yes (2021-05-18 09:21:22)
#Attributes	75 (0 Objects)
First recorded change	2018-11-19 18:33:59
Last change	2021-05-28 11:42:51
Modification map	
Sightings	0 (0) - restricted to own organisation only.

Figure 44 - ESET event that has a low confidence score.

Figure 49 shows that the even though the event has many attributes, these are not grouped into objects, and as explained in 2.1.2, having stand-alone attributes gives little or no context to the forensic. Furthermore, the event includes a false-positive, which is an indicator of low quality. The threat event only scores approximately 25%, and it is because of the essential event features which are not attached to the event.



☐	Date ↑	Org	Category	Type	Value	Tags	Galaxies	Comment	Correlate	Related Events	Feed hits	IDS	Distribution	Sightings	Activity	Actions
☐	2018-11-19		Payload delivery	sha1	996d0ac930d2cdb16ef96edc27d9d1afc2d89ca8			2nd stage script	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Payload delivery	sha1	2194271c7991d60ae82436129d7f25c0a689050a			1st stage script	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Antivirus detection	comment	JS/Agent.NYR				☑			☐	Inherit	(0/0/0)		
☐	2018-11-19		Antivirus detection	comment	JS/Agent.NYQ				☑			☐	Inherit	(0/0/0)		
☐	2018-11-19		External analysis	url	https://www.welivesecurity.com/2018/11/20/oceanlotus-new-watering-hole-attack-southeast-asia/				☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	p-typekit.com 104.248.144.136			2nd stage	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	10cm.mypets.ws 178.128.100.189			2nd stage	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	optnmstri.com 159.65.134.146			2nd stage	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	imgincapsula.com 209.97.164.158			2nd stage	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	secure-imrworldwide.com 178.128.90.109			2nd stage	☑		2	☐	Inherit	(0/0/0)		
☐	2018-11-19		Network activity	domain|ip	cdn-ampproject.com 178.128.24.201			2nd stage	☑		2	☐	Inherit	(0/0/0)		

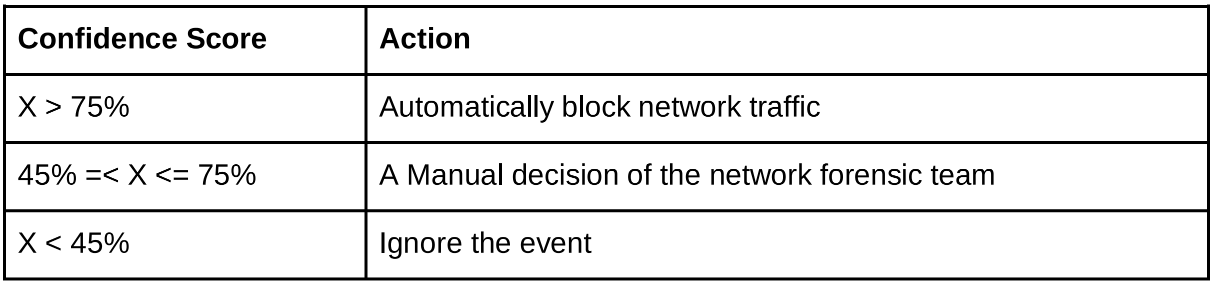
Figure 45 - ESET event that has a low confidence score.

Observations

The most prominent factor observed, is that the majority of data going through the ZMQ is not creation of new events. Feed providers are mostly updating old events, adding new attributes and features. Creation of new events is rarer than expected. This is important to notice because the current PoC is only programmed to evaluate new events, therefore this must be considered for the future.

Another observation made is the server capacity, which is not capable of handling the amount of data being transferred. The MISP platform is designed such that updated events and creation of new events are fetched based on a scheduled task. The scheduled task will run each hour of the day, and when it starts, the server will often shut down due to, too much congestion on the server, this is highly problematic and must be solved. The issue is easy solvable by buying more resources from the cloud provider, DigitalOcean.

The last observation done was the inability to measure fine-grained data quality. It is observed that many of the published events score fairly high confidence scores, which is clear indicator that the data quality is not as low as expected. Many of the feed providers score the same or higher than eCrimeLabs, who are considered to be good data quality. As discussed in 3.3.5, threshold values must be set, to give the event the correct tag, after having conducted the test and performance of the system is clear, combined with the overall data quality in the community, these thresholds can be set, there is not further rationale behind the thresholds, besides the observed performance of the algorithm. Figure 45 shows an example setup of the thresholds.



Confidence Score	Action
X > 75%	Automatically block network traffic
45% =< X <= 75%	A Manual decision of the network forensic team
X < 45%	Ignore the event

Figure 46 - The tagging mechanism, the user is able to specify the ranges.

7 Future, Challenges and Reflection

In the process of working on this project and while writing this report, some challenges were encountered. Also, some elements could not, in the timespan of this project be created or implemented, which have been intentionally left out or left for future works. This chapter will discuss these topics.

The project presents a PoC of a solution capable of evaluating data quality of JSON files containing information about threat events, and use this to create a reputation system.

7.1 A Large-Scale System

It is stated several times throughout the report that this project will only produce a proof of concept including some basic features. However, it is important to reflect how a fully functioning system would look.

The PoC included an algorithm capable of evaluating a threat event coming from the MISP instance based on different event features. The results are evaluated in a confidence score; based on the score, a tag will be assigned to the event, so later processes can acknowledge the quality of the data. An advanced system would first of all have a GUI, that would show various KPIs of important factors to further evaluate the quality of the feed providers. This GUI should very likely be presented in a webapp. After interviews with FE (Section 3.1.3), it is clear, that the system must be able to handle events sourced from multiple platforms and have a unified algorithm. The PoC measures relatively simple features, e.g., are tags used, and is the number of tags used satisfying. An improvement of such would be to implement a feature that could analyse at the specific tag and understand the context of the tag and decide, whether the tag is in the right context.

The PoC is listening on the ZMQ (Section 2.1.3), which is a somewhat simple solution implemented by the MISP project; however, this solution is not very scalable and generic, therefore another approach of extracting features must be developed. One approach is to fetch the events straight from the database instead; this would require more advanced processing, because the entire MISP database would have to be searched through to find events. However, it is believed to be a more viable solution and would allow for historical statistics. The ZMQ also suffers from the problem that, if the user is not listening for them, the events are just dropped from the queuing system, meaning that events could get past the ZMQ without being evaluated.

7.2 Data Formats

One of main challenges which have left a great impact on the system is the multiple data formats issue, briefly touched on in Section 3.3.3. To overcome the problem, several experts in programming are contacted to discuss it (these conversations were not documented). The challenge was introduced, when FE suggested that the system should be able to handle multiple input formats (Section 3.1.3), meaning inputs from different threat intelligence platforms. The problem is that each and every threat platform uses their own data format. So even though they all use the general JSON data format, each of the JSON structures is different. The actual data contained in the JSON file can be also different across sources, which means, that some platforms include tags and galaxies (like MISP), and others include different data points. To create an algorithm capable of unifying these different structures lies outside of the scope of this project, but possible solutions can be considered.

7.2.1 Different Evaluation Models

The first approach to solve the problem, which was also suggested by one of the experts is to create an evaluation algorithm for each platform. Using the same rating scale, but rate different available data points. This approach, however, would be very time consuming, and not effectively scalable. The solution would also require much maintenance, since the data often change in event reports, when new updates are released to the specific platform.

7.2.2 Unified JSON Format

The second approach considered is to create a unified JSON structure. This is because some of the different platforms use same data as tags, sighting reports and more. However, they are named differently from platform to platform. This limits the JSON analysis that must be done to fetch correct information to the evaluation. It doesn’t really solve the entire problem, because as described earlier, the platform also uses different kind of data, and this must again be solved by using different evaluation algorithms specified for the given platform.

7.3 Future Work

The future work of this project is considered very much because the intention is not to shelf this project after the deadline. The first goal is to rewrite the code, having more focus on the JSON structure, after each event have been evaluated, a new JSON file should be created, where all the features, and scores should be saved. This would allow for historical analysis of the events, and further manipulation if that is needed. Another change would be to go from Excel to a known tool like PowerBI, which has even more statidtical analysis, and allow for integrations with many different software.

One of the requirements is that the system must be generic and work for other platforms than the MISP platform. Therefore, this goal will also be pursued, this is also because an interest and challenging objective must be solved, the challenge explained above. How can the format be unified and is it even possible? If this challenge could be solved, the system could be used in many organisations, and FE has already shown an interest into that feature.

New features to measure data quality on should also be improved, the current PoC is measuring the individual features, but the system should also measure on combinations of features, e.g., if they threat is a phishing email, origin email, screenshot, and the link should be filled and given a certain amount of points, however, if only link is filled fewer points should be given to the confidence score.

7.4 Reflection

During the process of writing this thesis several thoughts and reflections have been made, these will be discussed in this subchapter. It is clear that data is highly important in today's society, and the maturity plays an important role in utilising it in the best way. It is now clear that data quality, cannot just be defined as good or bad, it is highly dependent on the user and usage, and that is why it so difficult to just make an algorithm and quantify it. However, what is learnt is this project is whole task of working with data, ETL and data linage systems. When the data has been fetched, there are so many opportunities to present, alter or measure it, this report suggest one approach.

The PoC in this report, is fully functioning capable of measuring the data quality, but the result is not as anticipated in the beginning of the project. The expected outcome was an algorithm that could measure fine-grained data quality, and this was not the end result. The PoC can measure data quality, but on a higher level. It is not suitable for fined-grained data quality, it is more suitable for high-level, meaning that threat events of very low or high quality, can be filtered out. To make it suitable for fine-grained measurement, a very deep insight into the how the events are evaluated by the forensics is needed, because it would requirement complex combinations of event features to measure.

There is one thing that should be changed if this project should be done again, or revisited, the focus on machine learning. One of the first tasks should have been to create a good data set with all the features, so a good machine learning model could be created, and see how it would perform compared to the weighted average model.

8 Conclusion

After thorough investigations of relevant concepts and framework, a conclusion can now be drawn. The thesis explores the great potential of a system, which is capable of quantifying data quality, and by that enhancing the trust in the MISP community.

Data is the essential cornerstone of many operations and organisations; therefore, data is often considered very valuable and attractive – also for cyber criminals. Organisations are spending huge amount of resources trying to avoid being successfully attacked. Threat intelligence platforms are one security measure, that are becoming more popular, and one widely used platform is the MISP platform.

The MISP platform is open-source, and anybody can join the community, this generates large amount of threat information. The information is being published to organisations that in some cases have dedicated forensics teams that evaluate this information manually, this task is tedious and time consuming.

The goal of this project was to create an algorithm capable of analysing large amount of information based on its data quality. Furthermore, it is believed that, using the information to create a reputation system the trust in the community would be increased, and as a positive consequence the forensics can automate more processes. The above mentioned were stated in the problem formulation: *How can an engine for the MISP platform be designed and developed to assist forensics in evaluating incoming events based on the data quality, to respond more precise to threats?*

The research turned out to be complicated since, the MISP platform is not heavily documented, and it is very technical, therefore interviews with MISP experts was conducted, the most prominent expert was Dennis Rand. He provided valuable information about how MISP operates, events and what to measure on when creating the algorithm. Other similar services were investigated, the research didn’t give much, since their algorithms were not exposed, even after requesting demos.

The concept of data quality was highly important to the project and therefore it was deeply explored, the important ‘fitness to use’ definition was adopted and emphasised heavily throughout the project, since user customisation of the algorithm was essential to the algorithm according to Dennis Rand. The other important concept to the project was reputation. It was found that a feeling of zero trust was existing in the MISP community, and everything must be evaluated. It is assumed if a reputation system would be implemented, more trust would be built, amongst the forensics and feed providers.

The research gave a lot of knowledge about the concepts and smaller concepts which could be attached, like collaborative sanctioning, cold start problem and rating mechanisms. This information would be analysed and put into perspective of creating a solution to solve the problem formulation. The analysis eventually ended with a requirements specification, that outlined the most important features, that must be implemented into a PoC. It, furthermore, gave an idea of the functionalities a large-scale system should include.

After the technologies, concepts and elements of the system were chosen and analysed, the system began taking shape. To visual the abstract idea, a few UML diagrams, flowcharts, architecture diagrams and a simple prototype were created. This gave a good oversight of the problems and how the components would interact. The diagrams were then used for the implementation of the PoC.

The PoC, is demonstrating the main contribution to the MISP community. This is a system capable of successfully analysing large amount of threat event information, based on the analysis give the event a confidence score. This score would produce a tag to the event for further automation. The confidence score would then be used to rate the feed provider publishing the given event, consequently making a reputation system, which main intention is to increase the trust in the MISP community.

It is the strong belief that the system has a great potential based on the result of the PoC, statements from experts in the field, and the deep research into the MISP community and other data driven systems.

9 References

[1] A. R. Haris, “INFORMATION ISSUES IN DIGITAL ERA,” Dec. 2016, Accessed: Mar. 30, 2021.[Online]. Available:https://www.researchgate.net/publication/328528038_Issues_In_Digital_Era.

[2] C. Gates and P. Matthews, “Data Is the New Currency,” *Proceedings of the 2014 workshop on New Security Paradigms Workshop - NSPW ’14*, 2014, doi: 10.1145/2683467.2683477.

[3] “currency noun - Definition, pictures, pronunciation and usage notes | Oxford Advanced Learner’s Dictionary at OxfordLearnersDictionaries.com,” *www.oxfordlearnersdictionaries.com*. https://www.oxfordlearnersdictionaries.com/definition/english/currency (accessed Mar. 30, 2021).

[4] A. Pasternack, “Here are the data brokers quietly buying and selling your personal information,” *Fast Company*, Mar. 02, 2019. https://www.fastcompany.com/90310803/here-are-the-data-brokers-quietly-buying-and-selling-your-personal-information.

[5] “Cyber Threat Intelligence 101,” *FireEye*. https://www.fireeye.com/mandiant/threat-intelligence/what-is-cyber-threat-intelligence.html (accessed Feb. 03, 2021).

[6] “The Cybersecurity Market Report covers the business of cybersecurity, including market sizing and industry forecasts, spending, notable M&A and IPO activity, and more.,” *Cybercrime Magazine*, Mar. 03, 2018. https://cybersecurityventures.com/cybersecurity-market-report-q4-2015/ (accessed Feb. 03, 2021).

[7] “Who,” *www.misp-project.org*. https://www.misp-project.org/who/ (accessed Feb. 03, 2021).

[8] “MISP Communities and MISP Feeds,” *www.misp-project.org*. https://www.misp-project.org/communities/ (accessed Feb. 14, 2021).

[9] “What is IOC in Cyber Security? - Logsign,” *www.logsign.com*. https://www.logsign.com/blog/what-is-ioc-in-cyber-security/ (accessed Mar. 30, 2021).

[10] “IOC’s turning into IOOI’s,” *SANS Internet Storm Center*. https://isc.sans.edu/forums/diary/IOCs+turning+into+IOOIs/26624/ (accessed Mar. 30, 2021).

[11] “eCrimeLabs - Helps you mitigate your cyber threats,” *eCrimeLabs - Helps you mitigate your cyber threats*. https://www.ecrimelabs.com (accessed Mar. 30, 2021).

[12] “Forsvarets Efterretningstjeneste,” *Forsvarets Efterretningstjeneste*. https://fe-ddis.dk/da/ (accessed Mar. 30, 2021).

[13] “CIRCL» MISP - Open Source Threat Intelligence Platform,” *www.circl.lu*. https://www.circl.lu/services/misp-malware-information-sharing-platform/ (accessed Mar. 30, 2021).

[14] “· User guide of MISP intelligence sharing platform,” *circl.lu*. https://circl.lu/doc/misp/GLOSSARY.html (accessed Mar. 30, 2021).

[15] “What are Message Brokers?,” *www.ibm.com*. https://www.ibm.com/cloud/learn/message-brokers (accessed Mar. 31, 2021).

[16] “Get started,” ZeroMQ. https://zeromq.org/get-started/ (accessed Mar. 31, 2021).

[17] “IBM Knowledge Center,” *www.ibm.com*. https://www.ibm.com/support/knowledgecenter/en/ssw_ibm_i_73/rzab6/howdosockets.htm (accessed Mar. 31, 2021).

[18] “Socket API,” *ZeroMQ*. https://zeromq.org/socket-api/ (accessed Mar. 31, 2021).

[19] “ZeroMQ - MISP publish-subscribe · User guide of MISP intelligence sharing platform,” *www.circl.lu*. https://www.circl.lu/doc/misp/misp-zmq/ (accessed Mar. 31, 2021).

[20] “JSON,” *www.json.org*. https://www.json.org/json-en.html (accessed Mar. 31, 2021).

[21] H. M. Yousif Al-Bayatti, A. M. S. Rahma, and H. B. Abdul Wahab, “PGP Protocol and Its Applications,” *Cryptography and Security in Computing*, Mar. 2012, doi: 10.5772/36971.

[22] “CIRCL» Report an Incident,” *www.circl.lu*. https://www.circl.lu/report/ (accessed Mar. 31, 2021).

[23] “PyMISP - Python Library to Access MISP · User guide of MISP intelligence sharing platform,” *www.circl.lu*. https://www.circl.lu/doc/misp/pymisp/ (accessed Mar. 31, 2021).

[24] “Top Threat Intelligence Platforms for 2021 | eSecurity Planet,” *eSecurityPlanet*, Dec. 19, 2020. https://www.esecurityplanet.com/products/threat-intelligence-platforms/ (accessed Mar. 31, 2021).

[25] “IBM X-Force Exchange Threat Intelligence | What is X-Force?,” *eSecurityPlanet*, Jul. 18, 2017. https://www.esecurityplanet.com/products/ibm-xforce/ (accessed Mar. 31, 2021).

[26] “IBM X-Force Exchange,” *exchange.xforce.ibmcloud.com*. https://exchange.xforce.ibmcloud.com/ (accessed Mar. 31, 2021).

[27] “IBM X-Force Exchange API Documentation,” *api.xforce.ibmcloud.com*. https://api.xforce.ibmcloud.com/doc/# (accessed Mar. 31, 2021).

[28] A. Inc, “Anomali Integrates MITRE ATT&CK Framework Across Threat Platform in Winter 2020 Product Release,” *GlobeNewswire News Room*, Feb. 21, 2020. https://www.globenewswire.com/news-release/2020/02/21/1988608/0/en/Anomali-Integrates-MITRE-ATT-CK-Framework-Across-Threat-Platform-in-Winter-2020-Product-Release.html (accessed Mar. 31, 2021).

[29] “ThreatStream - Threat Intelligence Platform,” *www.anomali.com*. https://www.anomali.com/products/threatstream (accessed Mar. 31, 2021).

[30] “AlienVault is Now AT&T Cybersecurity,” *cybersecurity.att.com*. https://cybersecurity.att.com (accessed Mar. 31, 2021).

[31] “AlienVault - Open Threat Exchange,” *AlienVault Open Threat Exchange*. https://otx.alienvault.com/dashboard/new (accessed Mar. 31, 2021).

[32] D. M. Strong, Y. W. Lee, and R. Y. Wang, “Data quality in context,” *Communications of the ACM*, vol. 40, no. 5, p. 104, May 1997, doi: 10.1145/253769.253804.

[33] L. Cai and Y. Zhu, “The Challenges of Data Quality and Data Quality Assessment in the Big Data Era,” *Data Science Journal*, vol. 14, no. 0, p. 2, May 2015, doi: 10.5334/dsj-2015-002.

[34] Cambridge Dictionary, “TRUST | meaning in the Cambridge English Dictionary,” *Cambridge.org*, 2020. https://dictionary.cambridge.org/dictionary/english/trust (accessed Mar. 31, 2021).

[35] Cambridge Dictionary, “REPUTATION | meaning in the Cambridge English Dictionary,” *Cambridge.org*, Nov. 27, 2019. https://dictionary.cambridge.org/dictionary/english/reputation (accessed Mar. 31, 2021).

[36] J. Sabater and C. Sierra, “Review on Computational Trust and Reputation Models,” *Artificial Intelligence Review*, vol. 24, no. 1, pp. 33–60, Sep. 2005, doi: 10.1007/s10462-004-0041-5.

[37] Y. Wu, Y. Qiao, Y. Ye, and B. Lee, “Towards Improved Trust in Threat Intelligence Sharing using Blockchain and Trusted Computing,” *2019 Sixth International Conference on Internet of Things: Systems, Management and Security (IOTSMS)*, Oct. 2019, doi: 10.1109/iotsms48152.2019.8939192.

[38] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The Eigentrust algorithm for reputation management in P2P networks,” *Proceedings of the twelfth international conference on World Wide Web - WWW ’03*, 2003, doi: 10.1145/775152.775242.

[39] A. Mohaisen, O. Al-Ibrahim, C. Kamhoua, K. Kwiat, and L. Njilla, “Assessing Quality of Contribution in Information Sharing for Threat Intelligence,” *2017 IEEE Symposium on Privacy-Aware Computing (PAC)*, Aug. 2017, doi: 10.1109/pac.2017.39.

[40] “Threat Intelligence Community,”, Deliverable 5.1, Horizon 2020 Programme, Oct. 2018.

[41] “Threat Intelligence Sharing: State of the Art and Requirements,”, Deliverable 5.4 , Horizon 2020 Programme, May 2017.

[42] M. Ahmed and A. N. Mahmood, “Novel Approach for Network Traffic Pattern Analysis using Clustering-based Collective Anomaly Detection,” *Annals of Data Science*, vol. 2, no. 1, pp. 111–130, Mar. 2015, doi: 10.1007/s40745-015-0035-y.

[43] A. Jøsang and J. Golbeck, “Challenges for Robust Trust and Reputation Systems,” Oct. 2009.

[44] “About Trustpilot,” *Trustpilot.com*, 2019. https://www.trustpilot.com/about.

[45] “Why do I need to fill in a reference number?,” *Trustpilot Support Center*. https://support.trustpilot.com/hc/en-us/articles/202195568-Why-do-I-need-to-fill-in-a-reference-number- (accessed Apr. 09, 2021).

[46] M.-H. Nadimi-Shahraki and M. Bahadorpour, “Cold-start Problem in Collaborative Recommender Systems: Efficient Methods Based on Ask-to-rate Technique,” *Journal of Computing and Information Technology*, vol. 22, no. 2, p. 105, 2014, doi: 10.2498/cit.1002223.

[47] “Weighted Average Definition,” *Investopedia*, 2020. https://www.investopedia.com/terms/w/weightedaverage.asp (accessed Apr. 14, 2021).

[48] “MISP Default Feeds,” *www.misp-project.org*. [Online]. Available: https://www.misp-project.org/feeds/. [Accessed: 16-Apr-2021]

[49] Agile Business Consortium, “Chapter 10: MoSCoW Prioritisation,” *Agilebusiness.org*, 2019. [Online]. Available: https://www.agilebusiness.org/page/ProjectFramework_10_MoSCoWPrioritisation. [Accessed: 16-Apr-2021]

[50] “What is FURPS+?,” *Business Analyst Training in Hyderabad - COEPD*, 05-Aug-2014. [Online]. Available: https://businessanalysttraininghyderabad.wordpress.com/2014/08/05/what-is-furps/. [Accessed: 16-Apr-2021]

[51] R. Gandhi, “Support Vector Machine — Introduction to Machine Learning Algorithms,” *Towards Data Science*, 07-Jun-2018. [Online]. Available: https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47. [Accessed: 19-Apr-2021]

[52] “What Is Threat Intelligence? Definition and Examples,” *Recorded Future*, 30-Apr-2019. [Online]. Available: https://www.recordedfuture.com/threat-intelligence-definition/. [Accessed: 19-Apr-2021]

[53] “What is a Threat Intelligence Platform (TIP)?,” *www.anomali.com*. [Online]. Available: https://www.anomali.com/resources/what-is-a-tip. [Accessed: 19-Apr-2021]

[54] A. Jøsang, R. Ismail, and C. Boyd, “A survey of trust and reputation systems for online service provision,” *Decision Support Systems*, vol. 43, no. 2, pp. 618–644, Mar. 2007, doi: 10.1016/j.dss.2005.05.019.

[55] F. Hendrikx, K. Bubendorfer, and R. Chard, “Reputation systems: A survey and taxonomy,” *Journal of Parallel and Distributed Computing*, vol. 75, pp. 184–197, Jan. 2015, doi: 10.1016/j.jpdc.2014.08.004.

[56] N. S. Hmeidat and M. Sadiq Aljumaily, “Pattern Recognition ‘Anomaly Detection Challenges,’” Dec. 2015, doi: 10.13140/RG.2.2.31798.60480.

[57] Y. Roh, G. Heo, and S. E. Whang, “A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective,” *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019, doi: 10.1109/tkde.2019.2946162. [Online]. Available: https://arxiv.org/pdf/1811.03402.pdf

[58] “eCrimeLabs - Helps you mitigate your cyber threats — eCrimeLabs Services,” *eCrimeLabs - Helps you mitigate your cyber threats*. [Online]. Available: https://www.ecrimelabs.com/services. [Accessed: 26-Apr-2021]

[59] “What is Horizon 2020?,” *European Commission*, 23-Oct-2013. [Online]. Available: https://ec.europa.eu/programmes/horizon2020/en/what-horizon-2020. [Accessed: 07-May-2021]

[60] “DAMAGE CONTROL: THE COST OF SECURITY BREACHES IT SECURITY RISKS SPECIAL REPORT SERIES Kaspersky Lab,” 2015 [Online]. Available: https://media.kaspersky.com/pdf/it-risks-survey-report-cost-of-security-breaches.pdf. [Accessed: 11-May-2021]

[61] “NINTH ANNUAL COST OF CYBERCRIME STUDY UNLOCKING THE VALUE OF IMPROVED CYBERSECURITY PROTECTION THE COST OF CYBERCRIME CONTENTS,” [Online]. Available: https://www.accenture.com/_acnmedia/pdf-96/accenture-2019-cost-of-cybercrime-study-final.pdf. [Accessed: 11-May-2021]

[62] “Traffic Light Protocol (TLP) Definitions and Usage | CISA,” *www.cisa.gov*. [Online]. Available: https://www.cisa.gov/tlp. [Accessed: 11-May-2021]

[63] “What is SOAR? Security definition,” *FireEye*. [Online]. Available: https://www.fireeye.com/products/helix/what-is-soar.html. [Accessed: 18-May-2021]

[64] M. K. Pratt, “What is Zero Trust? A model for more effective security,” *CSO Online*, 16-Jan-2018. [Online]. Available: https://www.csoonline.com/article/3247848/what-is-zero-trust-a-model-for-more-effective-security.html. [Accessed: 18-May-2021]

[65] Oasis-Open, *STIX 2.1 Examples*. https://oasis-open.github.io/cti-documentation/stix/examples.html.

[66] N. Christensen and O. Oszczanowska, “An Independent Ranking Engine for the MISP Framework,” Aalborg University, Dec. 2020.

[67] “ETL explained,” *www.ibm.com*. [Online]. Available: https://www.ibm.com/docs/en/spm/7.0.4?topic=explained-etl. [Accessed: 25-May-2021]

[68] Authors Private MISP Instance

[69] Expert Statement obtained during interview

[70] “PyMISP - Python Library to Access MISP · User guide of MISP intelligence sharing platform,” *www.circl.lu*. [Online]. Available: https://www.circl.lu/doc/misp/pymisp/. [Accessed: 31-May-2021]

[71] “IBM X-Force Exchange - Details,” *www.ibm.com*. [Online]. Available: https://www.ibm.com/products/xforce-exchange/details. [Accessed: 31-May-2021]

[72] “Cyber-security firm Anomali to create 120 Belfast jobs,” *BBC News*, 18-May-2017 [Online]. Available: https://www.bbc.com/news/uk-northern-ireland-39959091. [Accessed: 31-May-2021]

[73] “Why Is Data Quality Important? | What is Data Quality?,” *LOTAME*, 30-Apr-2019. [Online]. Available: https://www.lotame.com/why-is-data-quality-important/. [Accessed: 31-May-2021]

[74] S. Li, “Solving A Simple Classification Problem with Python — Fruits Lovers’ Edition,” *Medium*, 06-Dec-2017. [Online]. Available: https://towardsdatascience.com/solving-a-simple-classification-problem-with-python-fruits-lovers-edition-d20ab6b071d2. [Accessed: 31-May-2021]

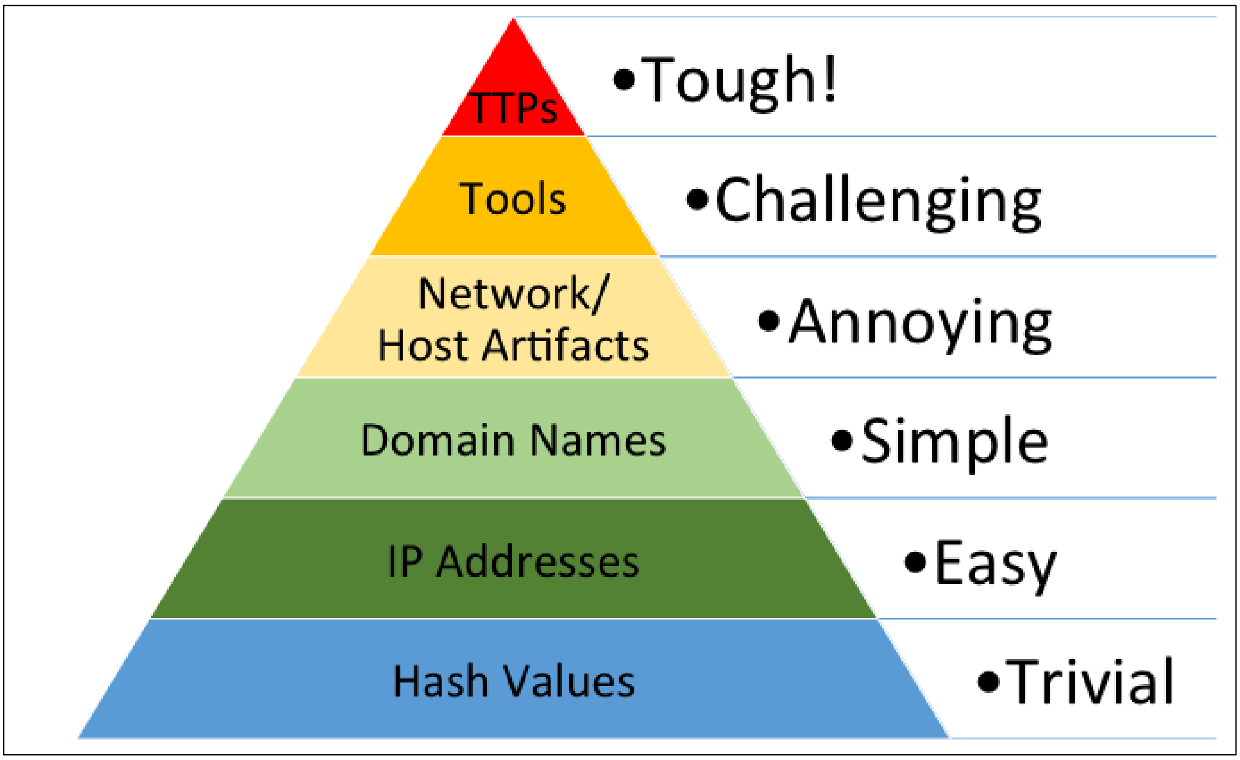
[75] “AAU - Viden for verden,” *www.aau.dk*. [Online]. Available: https://www.aau.dk. [Accessed: 02-Jun-2021]

10 Appendix

This appendix includes resources which are essential as documentation or additional knowledge to the project. It includes emails, interviews, audio files, etc.

10.1 The pyramid of IoC

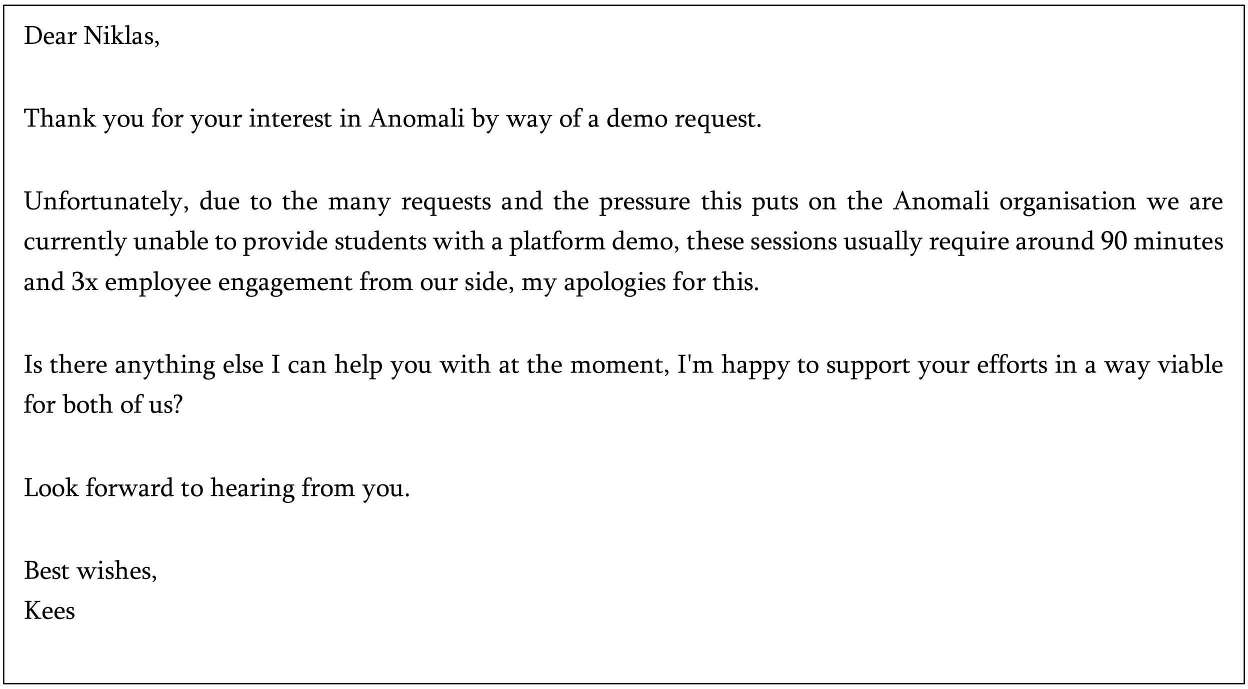
The pyramid of IoC is a figure often used to describe how difficult it is for attackers to alter their IoC. The lowest part of the pyramid is the items that can be easily changed by an attacker. The higher in the pyramid, the more difficult it is for an attacker to change the IoC, therefore network forensics should always try to gather information, in the highest part of the pyramid.



10.2 Related Emails

10.2.1 Anomali

This section shows the email correspondence between the author and an employee from the threat intelligence platform Anomali. The correspondence is about a demonstration of their platform



Dear Niklas,

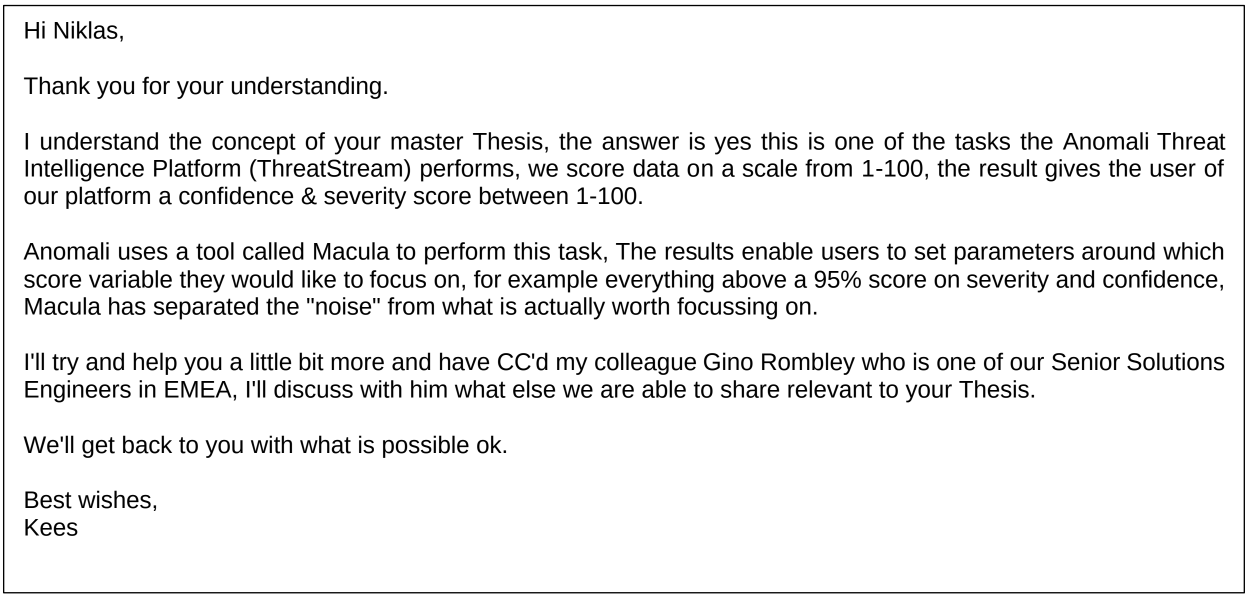
Thank you for your interest in Anomali by way of a demo request.

Unfortunately, due to the many requests and the pressure this puts on the Anomali organisation we are currently unable to provide students with a platform demo, these sessions usually require around 90 minutes and 3x employee engagement from our side, my apologies for this.

Is there anything else I can help you with at the moment, I'm happy to support your efforts in a way viable for both of us?

Look forward to hearing from you.

Best wishes,
Kees



Hi Niklas,

Thank you for your understanding.

I understand the concept of your master Thesis, the answer is yes this is one of the tasks the Anomali Threat Intelligence Platform (ThreatStream) performs, we score data on a scale from 1-100, the result gives the user of our platform a confidence & severity score between 1-100.

Anomali uses a tool called Macula to perform this task, The results enable users to set parameters around which score variable they would like to focus on, for example everything above a 95% score on severity and confidence, Macula has separated the "noise" from what is actually worth focussing on.

I'll try and help you a little bit more and have CC'd my colleague Gino Rombley who is one of our Senior Solutions Engineers in EMEA, I'll discuss with him what else we are able to share relevant to your Thesis.

We'll get back to you with what is possible ok.

Best wishes,
Kees

10.2.2 Alexandre Dulaunoy. CIRCL

The email from Alexandre Dulaunoy, that explains how useful the system could be for the MISP integration, he furthermore provides a list of parameters he thinks could be worth measuring on, in terms of data quality.

Hi Dennis, Hi Niklas,

This sounds like a very promising project. This could be indeed very useful to evaluate the events shared among a community or even within a private/local community.
I see many parameters/data point of interest which could be used for evaluating an event:

- Are objects used?
  - If yes, are those connected with relationships?
- What's the ratio of relationships on this event? versus the overall ratio of the community?
- What are the relationship type used? versus the overall ratio of the community?
- Are the comments fields used or not?
- Does the comment field contain sentences? or duplicate information from the attributes/objects?
- Are the first-last seen fields set?
- Are the type of object used common? or uncommon among the other events?
- How many tags are set? versus the median of tags used in the community?
- Are those tags for taxonomies? or not?
- Are galaxy cluster attached? Ratio/community?
- ...

We could even discuss at a later stage of the project and see together other factors which could be used to do an evaluation.

As we are interested to have such kind of feature in MISP directly, I'm wondering if we could not improve the integration of the project in the core MISP. Let us know if this could be realistic within this master thesis.

Thanks a lot for your work.

Don't hesitate to reach out to us if you need support or any specific feedback from our side.

Kind regards

Hi Niklas,

Maybe a good source would to review the ones from the OSINT feed: https://www.circl.lu/doc/misp/feed-osint/

The latest ones from CIRCL or CERT.be can be good examples.

I hope this helps.

Cheers

10.2.3 Emails from Expert, Dennis Rand

The following is an email from the MISP Dennis Rand, that explains the importance of such a system, why he thinks it will create value to the entire MISP community.

One of the biggest issues when sharing threat data amongst organisations is to in some form validate or vet the content recieved. Through the years I've seen many ways this has been attempted on the data delivered by itself, however this often fails as there are so many situations where it is not possible to vet/validate data based on as an example threat data shared can be specified to a specific organisation so if anyone else visits that IP/domain or URL they will recieve other content. Also the lack of proper validation of data can be an issue (false positives) such as organisaitons sharing none-malicious infrastructure items e.g. 8.8.8.8(Doogle DNS) or simply not giving enought context to the recieving party.

The second issue is the volume of data potentially shared as long as it is only a few events a day there is a potential that a CDC can handle and evaluate this, however any volume above often requires dedicated ressources. With the issues mentioned above a system or model is needed to be able to handle large volume as well as look at other elements than the data itself.

With this in mind the project Niklas is doing is essential to any organisation recieving arbitrary threat data from external parties, to help them assess the data recieved, as a guideline for the analyst on how potential trustworthy the data is, and with that be able to handle a larger amount and thereby reducing the threat landscape based on knowledge.