jamk

Open Cyber Threat Information Feeds for Cyber Security Education

Kati Weissenfelt

Master's thesis
April 2022
Technology
Master's Degree Programme in Information Technology
Cyber Security

jamk | Jyväskylän ammattikorkeakoulu
University of Applied Sciences

jamk

Description

Weissenfelt, Kati

Open Cyber Threat Information Feeds for Cyber Security Education

Jyväskylä: JAMK University of Applied Sciences, April 2022, 62 pages

Technology, Information Technology. Degree Programme in Information Technology. Master's thesis.

Permission for web publication: Yes

Language of publication: English

Abstract

Cyber Threat Intelligence feeds are tools to improve organization's situation awareness. To be able to protect the infrastructure, it is important to know what kind of threats and vulnerabilities there are. ENISA has published a list of topics in the Cyber Security field, that needs to be researched. The Cyber Threat Intelligence was one of the important topics.

The Food Chain Cyber Resilience project is having a Cyber Security Exercise in 2023 and the aim was to find out what feeds could be used in the exercise. Other purpose of this thesis was to act as a peer review to Juha Kuusenmäki's thesis, in which he created criteria to evaluate open Cyber Threat Intelligence feeds. On the top of Kuusenmäki's criteria, there was additional criterion to evaluate the educational use of the feeds.

The study was conducted in literature review about the Cyber Threat Intelligence feeds and case study about the assessment of the feeds.

The result was that all the evaluated feeds are suitable for educational needs. The evaluation with Kuusenmäki's criteria was done and the criteria was observed to be usable. Threat landscape is changing quickly so the results were not completely comparable.

Keywords/tags (subjects)

Cyber Threat Information, Cyber Threat Intelligence, Cyber Threat Intelligence Feeds, Cyber Threat Intelligence Tools

Miscellaneous (Confidential information)

No confidential information in this thesis.

Acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CTI	Cuber Threat Intelligence / Information
CVE	Common Vulnerabilities and Exposures
DNS	Domain Name System
ENISA	European Union Agency for Cybersecurity
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IDS	Intrusion Detection System
IOC	Indicators of Compromise
JSON	JavaScript Object Notation
OTX	Open Threat Exchange
SCO	STIX Cyber-observable Object
SDO	STIX Domain Object
SRO	STIX Relationship Object
STIX	Structured Threat Information Expression
TAXII	Trusted Automated Exchange of Intelligence Information
TLP	Traffic Light Protocol
TIP	Threat Intelligence Platform
TTP	Tactics, techniques & procedures
URL	Uniform Resource Identifier
XML	Extensible Markup Language

Contents

1 Introduction 5
2 Research Scope and Methods 6
2.1 Scope and Objectives 6
2.2 Research Questions 6
2.3 Research Methods 7
2.3.1 Qualitative Research 7
2.3.2 Literature Review 8
2.3.3 Case Study 8
3 Previous Research 9
4 Cyber Threat Information Systems 10
4.1 What Is Cyber Threat Information? 10
4.1.1 Traffic Light Protocol 12
4.2 STIX & TAXII 13
4.3 Existing Cyber Threat Information Feeds 18
4.3.1 Anomali Limo 18
4.3.2 AlienVault (AT&T) 19
4.3.3 Hail A Taxii 19
4.4 Chosen Feeds 19
4.5 Evaluation Criteria for the Feeds 22
5 The Environment and Tools For the Case Study 25
5.1 Case Study Concept 25
5.2 Test Environment 26
5.3 Target Threats 27
5.3.1 Emotet Comeback 27
5.3.2 Log4j 28
5.4 Combining the Feeds with TIP Tools 28
5.4.1 Anomali STAXX 28
5.4.2 MISP Platform 30
6 Results 31
6.1 Feed Review Results 31
6.1.1 PhishTank 31
6.1.2 DShield 33
6.1.3 EmergingThreats C&C Server 35
6.1.4 EmergingThreats – Compromised 37

6.1.5 AlienVault 39
6.1.6 Summary Of the Evalution 41
6.1.7 Comparing the Results 42
6.2 Case Study Results 43
6.2.1 PhishTank 45
6.2.2 DShield 46
6.2.3 EmergingThreats C&C Server 48
6.2.4 EmergingThreats – Compromised 49
6.2.5 AlienVault 50
6.2.6 Case Study Summary 52
6.3 Combining and Comparing the Feeds 54
7 Conclusion 55
7.1 Answers to the Research Questions 56
7.2 The Case Study Results 57
7.3 Further Research 57
References 58
Appendices 61
Appendix 1. Emotet IoC's From Cryptolaemus 61
Appendix 2. Log4j Ioc's From NCSC-NL 61
Appendix 3. PhishTank Feed Data 61
Appendix 4. DShield Feed Data 61
Appendix 5. Emerging Threats C&C Server Feed Data 61
Appendix 6. Emerging Threats – Compromised Feed Data 61
Appendix 7. AlienVault Feed Data 61
Appendix 8. Timeliness Results Between Feed Data and IOC Data 62

Figures

Figure 1 Threat Actor SDOs relations (OASIS, 2021) 15
Figure 2 Identity SDOs relations (OASIS. 2021) 16
Figure 3 TAXII Collections model (OASIS, 2021) 17
Figure 4 TAXII Channels model (OASIS, 2021) 17
Figure 5 Case Study Concept 26
Figure 6 Screenshot of Indicators from Anomali STAXX 29
Figure 7 Screenshot of Anomali STAXX dashboard 30
Figure 8 Screenshot of MISP event view 31
Figure 9 PhishTank Log4j timeliness 46
Figure 10 DShield Log4j timeliness 47
Figure 11 EmergingThreats C&C Log4j timeliness 48
Figure 12 EmergingThreats - Compromised Log4j timeliness 50
Figure 13 AlienVault Log4j timeliness 51

Tables

Table 1 Traffic Light Protocol (Paraphrased from CISA) 13
Table 2 Availability of the feeds 20
Table 3 Evaluation and scoring (Paraphrased from Kuusenmäki) 23
Table 4 Additional evalution criterion 24
Table 5 PhishTank rating 32
Table 6 DShield rating 34
Table 7 EmergingThreats C&C Server rating 36
Table 8 EmergingThreats - Compromised rating 38
Table 9 AlienVault rating 40
Table 10 Summary of the evaluation 42
Table 11 Comparing results with Kuusenmäki's thesis 43
Table 12 Threat IoC count in the feeds 44
Table 13 Case Study Summary 53
Table 14 Comparison with the feed review and the case study results 54

1 Introduction

To protect and secure an organization's infrastructure, it is important to know what kind of threats and vulnerabilities there are. Cyber Threat Information (or Cyber Threat Intelligence, CTI) feeds are feeds that provide information about the threats such as attack patterns, IP addresses, different kind of indicators, threat actors etc. Griffionen, Booij and Doerr stated in their publication *Cyber Threat Intelligence Feeds* that these feeds are only good when the data is good, meaning that the data provided should be relevant and provided in a timely manner to be considered as high-quality data (Griffionen, Booij & Doerr, 2020).

Situation awareness is a key factor in the Cyber Security. According to MITRE, the situation awareness can be divided in three categories: (a) Network awareness, which includes patching, auditing, compliance reporting and sharing the incident awareness. This is mostly under control in organizations. (b) Threat awareness, that can be improved with CTI feeds. This threat awareness is an evolving part of Cyber Security right now. (c) Mission awareness, which is needed and something to be developed more in the future (MITRE, 2021). This thesis is focused on the second category, threat awareness.

European Union Agency of Cybersecurity (ENISA) has published a list of topics in the Cyber Security field that needs to be researched. Among the topics like human dimension in Cyber Security and 5G security, the CTI related topics stands out. According to ENISA, promoting open-source CTI material helps to improve the pace of knowledge transfer and information management in organizations. The use of these open-source materials has relatively low skill requirements, and the open CTI feeds supports the possibility to importing multiple feeds on single base (ENISA, 2020).

Juha Kuusenmäki has evaluated Cyber Threat Information feeds in his thesis *Evaluation of Threat Information Feeds for a Cyber Defense Center* (2020). Kuusenmäki created criteria to evaluate the quality of the feeds, and in this thesis the same feeds and criteria are used for the evaluation of the CTI feeds. The other purpose of this thesis is to evaluate if Kuusenmäki's criteria are adequate and usable, and to assess if the selected feeds are usable in educational needs for the JYVSECTEC (Jyväskylä Security Technology Cyber Security Research, Development, and Training Center) or Jamk to use in cyber exercises or courses.

In 2023 JYVSECTEC is arranging a cyber security exercise to the Food Chain Cyber Resilience project, and the goal of this thesis is figure out a way to collect open threat feeds to a single point to be used more easily in the exercise, without having to pull all the information from the feeds individually. If there is a way to collect information from different feeds to a single source, this would save time and result in better overview about Cyber Security threats. In education use this would simplify the use of CTI feeds and save time.

2 Research Scope and Methods

2.1 Scope and Objectives

The cyber threat field is getting more complex and sophisticated over the time, so the challenges at keeping the organizations posted and up to date about the threats is getting more difficult. Collective understanding and cyber threat sharing could help organizations to make better decisions considering cyber defense. By consuming the cyber threat data from multiple sources, the existing knowledge about the threats can be enriched, and more accurate defensive and mitigation strategies can be made (NIST, 2016).

The aim of this thesis is to find out what kind of feeds there are and how these feeds could be consumed in a single point of use. This would save time and simplify the usage of the feeds. It is possible that the sensible use of this collected data would require some data-analytics, but this is not in the scope of this thesis. This could be something to research further by data analyst.

The outcome of this thesis will be a report, that describes the Cyber Threat Information feeds, their evaluation, and a way to use these feeds from a single point, if possible. The report will illustrate the significance of the Cyber Threat Intelligence data. The use of the feeds is not familiar to the author, so this will also be a learning experience.

2.2 Research Questions

The research questions are divided into three categories. First is considering the existing feeds and their content:

- What CTI feeds there are already?
- How can the feeds be consumed and used?
- Which of these feeds are suitable for the Cyber Security education use?

The second part applies to combining the feeds to a single feed:

- How is it possible to combine these feeds to a single feed?

The last question is about the used evaluation criteria:

- Are Kuusenmäki's criteria usable and valid?

2.3 Research Methods

The research methods are chosen by thinking about the desired outcome. Outcome will be a report which describes CTI feeds and their usage, and there will be a case study with evaluating and combining the feeds in a test environment.

2.3.1 Qualitative Research

According to Kananen, qualitative research has basically one question, which is: what this phenomenon is about? Qualitative research aims at the understanding of the phenomenon, and typically in the beginning of collecting the material for the study, there are no detailed questions about the research subject. Kananen points out, that if the phenomenon or subject is known, there is no need for the qualitative research (Kananen 2017, 32-34).

This thesis covers a subject that is not widely known, so the qualitative research is chosen for the research method.

2.3.2 Literature Review

Literature review as a research method is relevant especially when the subject is in a field that changes rapidly. It is important to build solid knowledge of the previous research made about the subject. It is a way to discover areas within the subject that have not yet been covered in studies (Snyder, 2019). In this research, I'm going to find out what is studied before about the Cyber Threat Information feeds, and what kind of feeds there are before I can implement the test environment.

2.3.3 Case Study

In addition to literature review, there is a need for mixed (or blended) research methods. Mixed method is combining the qualitative and quantitative research. According to Kananen (2017, 41), if there is a change involved, the design-based study is the right method to choose. It is hard to define whether the study is case study or design-based study. Kananen states that if the study is said to be case study, there should be multiple methods used with typically one case. Every research has a research subject, but every subject isn't a case. Design-based study is used to make change in for example product, method, process, organization etc. (Kananen, 2017, 48-49).

In this thesis the used method is case study, because there is no need to create new or change an existing product. I'm using existing products and existing feeds in the test environment.

The case study is done by comparing the feeds with Juha Kuusenmäki's criteria (Kuusenmäki, 2020). The criteria are described in the chapter 4.5. The results are presented in the chapter 6. Initial evaluation was conducted by analyzing the feed information provided by the feed suppliers and comparing the analysis with the Kuusenmäki's results. Criteria was compared to current Emotet and Log4j threat data.

The case study helps to determine if the feeds are usable in the Food Chain Cyber Resilience project and other educational needs. The ease of implementation criterion was added to the original criteria for this purpose.

3 Previous Research

Literature review was done by searching literature and research with key phrases “Cyber Threat Intelligence” and “Cyber Threat Information” from the JANET library, the Google Scholar and from the Internet. The literature to the study was chosen so that the subject is relevant to this thesis, published maximum ten years ago and the level of the research is at least master’s level thesis. Also, the feed providers’ information from their websites are reviewed in the results section of this thesis. The literature review is done to gain knowledge about the subject and to help to digest the information. There were five closely relevant previous publications found, which are presented below.

Evaluation of Threat Infromation Feeds for a Cyber Defence Center

There were two recent master’s theses found, that are close to this subject. The first one is the Juha Kuusemäki’s thesis *Evaluation of Threat Information Feeds for a Cyber Defense Center* (2020), from which the feed evaluation criteria to this thesis are taken. Kuusenmäki used STIX/TAXII feeds, and there were many feeds in common with my thesis. Kuusenmäki studied what STIX/TAXII feeds there are available and made a criterion, on which the feeds can be measured. Kuusenmäki stated that the feeds can be useful to increasing the situational awareness, but they also generate lots of data and can be very arduous to spot the meaningful data amongst all the noise. Kuusenmäki introduced lots of development ideas for further studies, which gave the spark for this thesis (Kuusenmäki, 2020, 64-67).

Measuring the Quality of Open Source Cyber Threat Intelligence Feeds

The other thesis is Keijo Korte’s thesis *Measuring the quality of Open Source Cyber Threat Intelligence Feeds* (2021). Korte’s thesis concentrates on the reliability of the CTI feeds. In Korte’s thesis, most of the feeds were not the same, that were chosen to this thesis, because some of them were comma separated (CSV) feeds. AlienVault and Emerging Threats were the only in common with my thesis. The outcome of Korte’s thesis was that it is not simple task to evaluate the reliability of the feeds and to compare them with each other. It is possible to compare different characteristics of the feeds, but it is not reliable. Also, the feeds are not always reliable, so the data collection period should be quite long, preferably at least a year (Korte, 2021, 48-50).

How to Define and Build an Effective Cyber Threat Intelligence Capability

Henry Dalziel wrote about what is the difference between data feeds and intelligence feeds. Dalziel states that most of the marketed feeds nowadays are data feeds, readable by machine. If the data is analysed and processed by human (or human-based) process, they can be called an intelligence feed (Dalziel, 2015, 4). Dalziel has also evaluated measures for the feeds, which are useful to compare the feeds with each other. These criteria are not very specific compared to Kuusenmäki's criteria. Dalziel is focusing on "why, what and how" one should implement a CTI system in organization.

FeedRank: A Tamper-resistant Method for the Ranking of Cyber Threat Intelligence Feeds

FeedRank is a ranking algorithm developed by Swiss research group in 2018. FeedRank is used to compare different feeds without a fear of tampering the results. Researchers discovered that feeds can be tampered so that they can come across to be better, or of higher quality, than they are. This can happen when comparing the feeds using metrics like feed's size, which can be easily tampered (Meier, Scherrer, Gugelmann, Lenders & Vanbever, 2018).

A Different Cup of TI? The Added Value of Commercial Threat Intelligence

Research groups paper *A different cup of TI? The added value of commercial threat intelligence* discusses on commercial threat intelligence products. The researchers compared two leading vendors products and their indicators on certain threats and discovered almost none overlapping between the feeds. They did the comparison to four open source feeds also, with the same results. This can mean that the feeds have serious issues in their trustworthiness - in indicators and timeliness (Bouwman, Griffionen, Egbers, Doerr, Klievink & Eeten, 2020).

4 Cyber Threat Information Systems

4.1 What Is Cyber Threat Information?

The NIST describes the CTI in the Guide to Cyber Threat Information Sharing publication like this:

> Cyber threat information is any information that can help an organization identify, assess, monitor, and respond to cyber threats. Examples of cyber threat information include indicators (system artifacts or observables associated with an attack), TTPs, security alerts, threat intelligence reports, and recommended security tool configurations. (NIST, 2016, iii)

Threat indicators are usually information about malicious IP address, suspicious DNS, URL, hash or for example subject line or content of an e-mail. Tactics, techniques & procedures (TTP's) describe the behavior of malicious actor, such as tools and malware variants they tend to use, order of their operations and delivery mechanisms. Threat intelligence reports are documents that describes for example the types of targeted systems, TTP's and actors. Tool configurations means instructions to detect and prevent the incidents with detection and removal software, and other configuration change instructions (NIST, 2016, 2).

Dalziel defines CTI as data, that has been processed by a human or human-based entity, such as artificial intelligence (AI). Feeds are usually data in some format, like JSON, XML, CSV, or they can be Application Programming Interfaces (API's), services, or other access points (Dalziel, 2015, 4). In this thesis the CTI is defined as any piece of information which includes some indicators of malicious activities.

ENISA points out that one the most important factors in Cyber Security research is topics is CTI: "CTI has been firmly established in the cybersecurity domain as an essential tool for enhancing agility and efficiency in defending cyberattacks." (ENISA, 2020, 7). This is because malicious actors are using more and more sophisticated and usually fully automated ways to find the vulnerabilities and infiltrate the systems. Usually, the attacks are automated but the means to prevent and mitigate them can be at least partially manual. Basic training in Cyber Security isn't enough, it is important to create solutions to improve security (ENISA, 2020).

Information sharing is an important part of the CTI. Information sharing creates a collective knowledge of the threat information. It provides information that usually might not be accessible: partner's experience and knowledge can help to create proactive and agile measures to protect the infrastructure from the threats. Shared and enriched knowledge is also more mature, and it

results in more effective recovery from the attacks (NIST, 2016, 3-4). This information gives more insight about the threat and helps organizations to detect and mitigate the threats more easily.

In the best scenario, the CTI indicators are timely, relevant, accurate, specific, and actionable, but this is not the situation in most cases (NIST, 2016, 22). Therefore, multiple resources and information sharing on communities and different platforms would help the enrichment of the data organization already have. Bouwman et al. discovered that, the CTI feeds are very specific and there is minimal overlapping between the feeds, and that can be a problem when trying to select the best tool to the organization (Bouwman et al., 2020).

Although information sharing is vital, it has some challenges. The trust between sharing community, sensitive and classified information, interoperability, and automation can cause some problems. Organization must have resources to handle and monitor the data, and to evaluate the quality of the information. The trust between the sharing community has to be established and organization must have guidance on how to communicate and publish the information (NIST, 2016, 4-5).

As mentioned earlier, the CTI can be presented in multiple formats. STIX 2.1 is modern, human readable and flexible way to present the data, and it uses the JSON format. STIX 1.0 presents data in XML format. All the feeds in this thesis are STIX/TAXII feeds.

4.1.1 Traffic Light Protocol

Traffic light protocol (TLP) is a tool created for sensitive information sharing. It is used to ensure that sensitive information is shared with correct audience (CISA). It is used to share any kind of sensitive information, not just information considering Cyber Security. Table 1 describes the colors and their meanings. Table is paraphrased from CISA.

Table 1 Traffic Light Protocol (Paraphrased from CISA)

Color	When used?	Sharing
TLP:RED Not for disclosure, restricted to participants only.	TLP:RED can be used when information could affect on privacy, reputation or operations if it ends up on malicious actors.	TLP:RED information cannot be shared outside the original parties in which it was originally disclosed. Usually TLP:RED information is exchanged verbally.
TLP:AMBER Limited disclosure, restricted to participants' organizations.	TLP:AMBER can be used when information cannot be shared outside the organizations involved, but requires support to be acted on.	TLP:AMBER information can be shared only on "need to know " basis with member of their own organization, clients or customers.
TLP:GREEN Limited disclosure, restricted to community.	TLP:GREEN is used when the information can be useful to all participating organizations and all the peer in community.	TLP:GREEN information can be shared with all the peers and partners, but not on public channels.
TLP:WHITE Disclosure is not limited.	TLP:WHITE information has minimal risk for the misuse.	TLP:WHITE can be distributed without restrictions. Standard copyright rules apply.

4.2 STIX & TAXII

STIX

STIX (Structured Threat Information Expression) is an open source language and a standardized format developed to exchange the CTI (OASIS, 2021). STIX was originally developed by Department of Homeland Security, MITRE and international community but was transferred under OASIS in 2015. First version of STIX was in XML format, but it required lots of customizations and had in-

teroperability issues. OASIS figured out that broader community was very supportive to JSON (JavaScript Object Notation), which has better performance, better scalability, and interoperability, but is not backwards compatible. This gave OASIS a fresh start to begin to develop STIX 2.0. The newest version of STIX is 2.1, which was published in March 2020 (Ginn, 2020).

Three main components of STIX are STIX Domain Objects (SDOs), STIX Cyber-observable Objects (SCOs) and STIX Relationship Object (SROs). STIX 2.1 objects are information which can be used to describe the threats. (OASIS, 2021) STIX 2.1 consists of 18 SDOs:

- Attack pattern
- Campaign
- Course of Action
- Grouping
- Identity
- Indicator
- Infrastructure
- Intrusion Set
- Location
- Malware
- Malware Analysis
- Note
- Observed Data
- Opinion
- Report
- Threat Actor
- Tool
- Vulnerability

Figures 1 and 2 shows examples of the relationships between the SDOs. Figures are made with Visualized SDO Relationships generator found on the OASIS website and are in courtesy of OASIS (OASIS, 2021).

Figure 1 Threat Actor SDOs relations (OASIS, 2021)

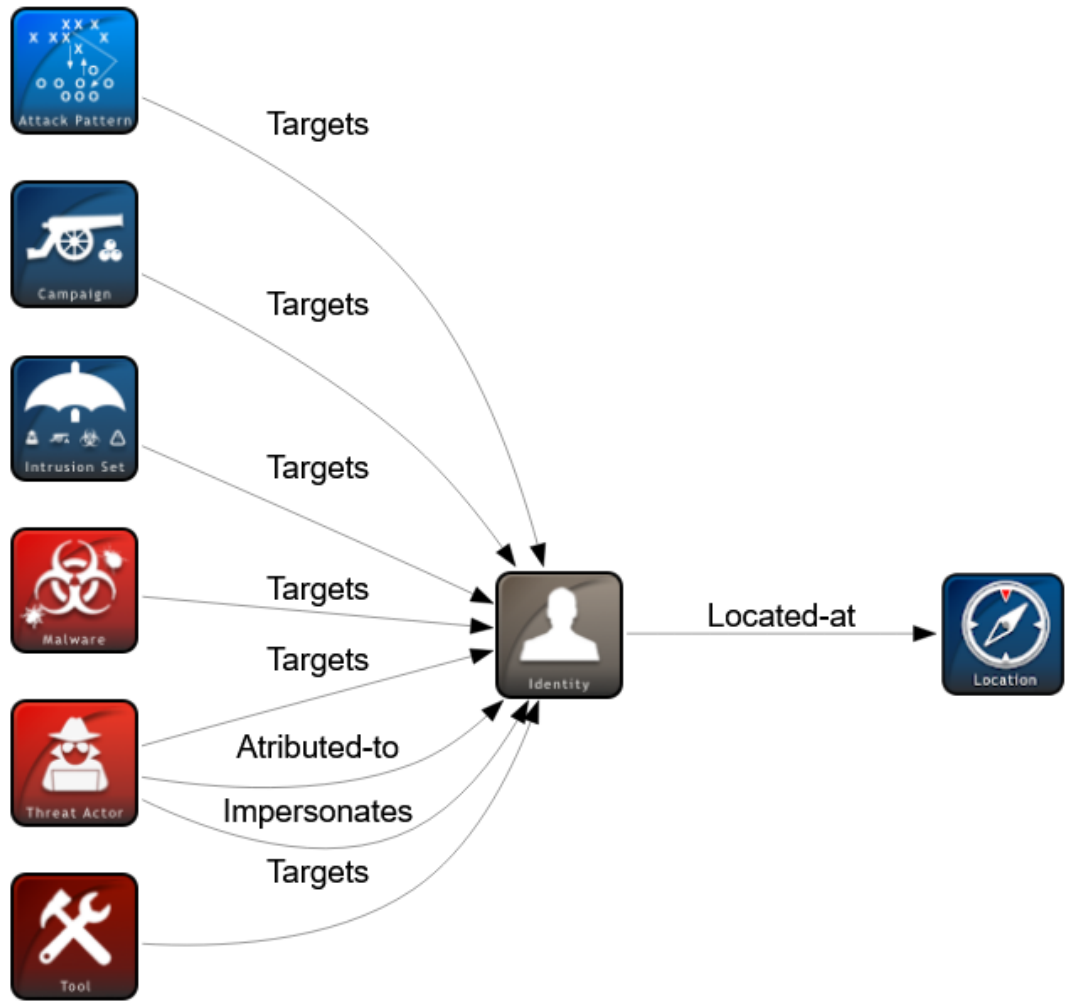

Figure 2 Identity SDOs relations (OASIS. 2021)

A common language helps CTI sharing between communities, which is a major benefit of STIX. CTI information is understandable to humans, and it can be used by machines to automate the CTI processes. It provides interoperability between different CTI tools (OASIS, 2021).

TAXII

Trusted Automated Exchange of Intelligence (TAXII) is a way to transport STIX information. It's an application protocol, which operates over Hypertext Transfer Protocol (HTTP) and Hypertext Transfer Protocol Secure (HTTPS). TAXII is not STIX specific transport protocol: STIX does not require TAXII, it can be transferred in other protocols too, and TAXII can be used to transfer other data than STIX (OASIS, 2021).

Two common sharing models of TAXII are collection and channel. Figure 3 demonstrates the collection, which is request-response model, where clients request the data and TAXII server provides it to the clients. Channel model describes a situation, where client publishes data via TAXII server to other clients to consume. This is described in the Figure 4 (OASIS, 2021). Figures are in courtesy of OASIS.

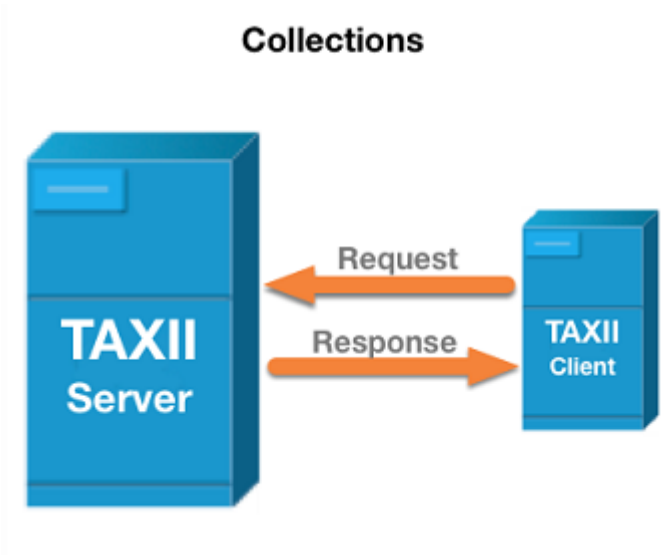

Figure 3 TAXII Collections model (OASIS, 2021)

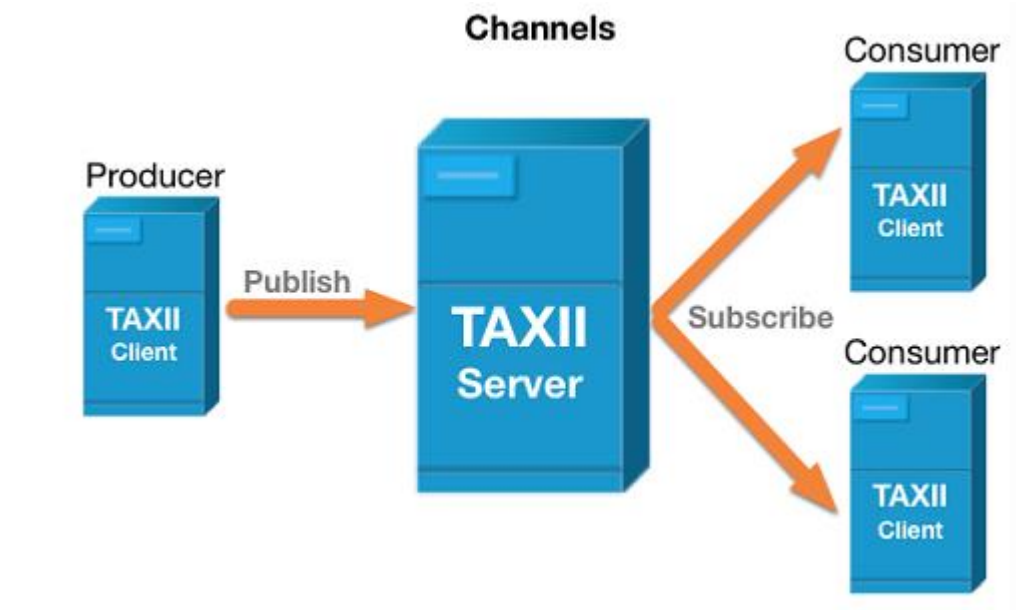

Figure 4 TAXII Channels model (OASIS, 2021)

4.3 Existing Cyber Threat Information Feeds

There are numerous CTI feeds available, but to this thesis only free of charge feeds are chosen. It is not reasonable to choose all the free feeds, so the feeds are chosen by estimating their popularity and their variance in contents. Popularity usually means that the feeds are updated regularly and are usable with typical tools. The chosen feeds are also easy to implement and use. The tools used to consume the feeds are Anomali STAXX and MISP. All the feeds are STIX/TAXII feeds.

There are repositories available, which includes multiple feeds, such as Anomali Limo and Hail A Taxii. AlienVault is service, where one can choose from multiple pulses, which can be ordered for consuming. This chapter describes these repositories and some singular feeds and their characteristics.

4.3.1 Anomali Limo

Anomali Limo is an Open Source CTI repository, which includes feeds:

- PhishTank
- Abuse.ch Ransomware IPs
- Abuse.ch Ransomware Domains
- DShield Scanning IPs
- Malware Domain List - Hotlist
- Blutmagie TOR Nodes
- Emerging Threats C&C Server
- DT COVID-19
- Lehigh Malwaredomains
- CyberCrime
- Emerging Threats - Compromised
- Anomali Weekly Threat Briefing

All these feeds are available when subscribing to Anomali Limo. The feeds contents are coming from Anomali Labs, Anomali's Honey Net, and from various open source feeds. Anomali Limo is STIX / TAXII 2.0 compatible and can be used in any TAXII client (Anomali, 2021).

4.3.2 AlienVault (AT&T)

AlienVault Open Threat Exchange (OTX) is an open threat intelligence community, which delivers threat data. It is possible to subscribe OTX pulses, that provides different indicator of compromise (IOC) that includes IP addresses, domains, email, URLs, hashes, files, CVE numbers and more. AlientVault OTX is compatible with many third party tools with DirectConnect API or with TAXII. It is also possible to download the data from the web portal for example in STIX format, which can be imported to different security tools (AT&T Cybersecurity, 2021).

Contents of the feeds (or pulses) can be chosen from lots of creators. Default pulse is AlienVault user's pulse, which is a good pulse to start investigating the indicators of compromise. AlienVault supports STIX and TAXII 1.0.

4.3.3 Hail A Taxii

Hail A Taxii is an Open Source Cyber Threat Intelligence repository, which offers the feeds in STIX format (HailATaxii, 2021). This TAXII repository offers feeds such as:

- Abuse.ch
- CyberCrime
- EmergingThreats
- Lehigh
- MalwareDomanList
- Blutmagie
- datForLast
- DShield
- PhishTank

Hail A Taxii can be used in any TAXII client that supports STIX and TAXII 1.0.

4.4 Chosen Feeds

The feeds to this thesis were chosen by their availability and by observing what feeds are reasonably active in somewhat short period of time. The range could be wider, if it would be possible to track the feeds longer than months in time, which is the case in this thesis. The chosen feeds have new data in relatively short intervals. Some of these feeds are available via Limo, Hail A Taxii and

MISP, and some are only available separately. Table 2 shows the availabilities of the chosen feeds in different repositories, and the standalone availability of the feed.

Table 2 Availability of the feeds

Feed/Availability	LIMO	Hail A Taxii	MISP	Standalone
PhishTank	X	X	X	X
DSHield	X	X		X
EmergingThreats C&C Server	X	X		X
EmergingThreats – Compromised	X		X	X
AlienVault				X

Because many of these feeds are already available in lots of repositories and tools, it is probably because these feeds are at least somewhat reliable and at least active feeds. AlienVault works a bit differently, but it is widely considered as a good CTI source. All the feeds are supervisioned by humans, either on community-basis or with volunteering Cyber Security professionals.

PhishTank

PhishTank is a free community-based anti-phishing service, which is operated by Cisco Talos Intelligence Group. Everybody can submit and review the suspected phishing sites or other malicious actors to the service. Registration is mandatory and this helps to reduce the noise a bit. It always requires more than one person to vote and verify a phish; this will depend on the voters' previous

history within the service. PhishTank data is available on the website and via their API (PhishTank, 2021). PhishTank is reliable source to retrieve CTI. They have lots of big Cyber Security companies, such as Avira, McAfee and Kaspersky lab, in their reference list.

DShield

DShield is a community-based CTI collection database behind SANS Internet Storm Center (ICS). DShield collects data across the Internet and the community uses the data to analyse, detect and predict malicious activities. DShield is sponsored by SANS Institute, and it is considered to be trust-worthy service. Registration is not required when submitting the firewall or Intrusion Detection System (IDS) logs to the service. DShield is developed originally by John Bambenek, who is still one the volunteers who investigates the logs send to DShield (DShield, 2022).

Emerging Threats C&C Server & EmergingThreats - Compromised

The free version of EmergingThreats has rulesets, which are community-based. EmergingThreats Pro Ruleset is maintained by the parent company Proofpoint (EmergingThreats, 2018). Emerg-ingThreats has different ruleset categories, and to this thesis the C&C Server and Compromises are chosen.

“Compromised rules” is a list of compromised hosts. This list is gathered from different private sources, and it is updated on daily basis. EmergingThreats does not reveal where the data is com-ing from, but they claim the sources are highly reliable. C&C is a list of confirmed Command and Control and botnet hosts, which is retrieved from shadowserver.org (EmergingThreats, 2018).

AlienVault OTX (AT&T)

AlienVault OTX (Open Threat Exchange) is AT&T’s product. AlienVault claims to be world’s first open threat intelligence community. AlienVault OTX is also a community-based service, which gives and access to wide community of Cyber Security professionals and researchers. OTX pulses are summaries of the threats, which includes all the IOCs to help in detecting the threats (AT&T Cybersecurity, 2021). This is chosen to the thesis based on its popularity and good reputation.

4.5 Evaluation Criteria for the Feeds

Kuusenmäki stated in his thesis, that there are no easily accessible research results about evaluation of different STIX/TAXII CTI feeds (Kuusenmäki, 2020, 67). When searching material to the literature review, this was found to be true. There are some other criteria, for example Dalziel has pointed out different things to evaluate, but there is very little actual research about the evaluation process and the results.

Dalziel concentrates on evaluating the vendors who offer CTI services, but the same properties apply to open CTI feeds also. According to Dalziel, there are the six criteria to evaluate: quality, quantity, uniqueness, value, ease, and the vendor itself. Quality is not described in detail, it means some way of measuring and agreeing on the type of data the feed produces. This depends on the customer's needs. Quantity is something to think about, although it is not always an indicator of value. Sometimes less data is better, but this also depends on the needs of the customer. Uniqueness means how much the feed is overlapping with other sources. Value means the price. Ease means compliance and the ease of implementation. The vendor itself refers to the reliability and reputation of the vendor (Dalziel, 2015, 26). When evaluating open CTI feeds, the value isn't typically the problem, but the system must run on some platform and have people to operate it, so even when using a free of charge feed, it usually generates some cost to the organization.

Kuusenmäki has limited his evaluation to five criteria, which are event quality, event timeliness, ease of use, event scope and cost (Kuusenmäki, 2020, 25). These comply with Dalziel's evaluation criteria, but Kuusenmäki has specified a three tier scoring system, which makes the evaluation more comparable than Dalziel's criteria. Kuusenmäki's system is also more concise, and its measures are easier to understand. In this thesis the Kuusenmäki's criteria is used on feed evaluation.

These evaluation criteria are meant to assess the feed's suitableness mostly for the business use, but in this thesis the intended use is education use. The needs are mostly the same, but in this thesis, there is no chargeable feeds taken in evaluation, so the cost is not a factor that needs to be considered. With the cost removed, the evaluation and scoring system will be reduced to four Kuusenmäki's criterion. Table 3 shows the rough scoring system, which is paraphrased from Kuusenmäki's thesis (Kuusenmäki, 2020, 26-29).

Table 3 Evaluation and scoring (Paraphrased from Kuusenmäki)

	Excellent	Good	Poor
Event quality	The feeds are from trusted sources and are human reviewed. Very little or no false positives. Can be used with automated incident respond systems.	The feeds are from trusted sources but have no filtering by human supervisor or the data is from non-trusted sources but is filtered by human supervisor. May contain some false positives. Could be used with incident respond systems, if supervised by human.	The data is from untrusted sources. High rate of false positives. Not recommended using with incident respond systems if only source of data.
Event timeliness	Events are published in almost real-time, maximum delay two hours. Can be used in incident management.	Events are published with the delay of 2 to 24 hours. Can be used in incident management.	Delay in event publication is more than 24 hours. Can be used in forensic analysis, not in incident management.
Ease of use	No complex registration process, maximum 24 hours to complete the process.	Registration takes maximum of three days to be completed. Supports 2.0 or greater STIX.	Complex registration, more than three days to complete the process. Supports only 1.x STIX version.

	Supports 2.0 or greater STIX.		
Event scope	Ten or more different type of IOCs. Can be used in incident management.	At least five different types of IOCs. Can be used in incident management in narrower scope.	Less than five different types of IOCs. Can be used in forensic analysis or in simple automated tasks as firewall rules etc.

In addition to Kuusenmäki's criteria, the evaluation is made bearing in mind the educational use. The tool to use the feeds are also in big role, because the ease of implementation and easy usage is important when there is little time to adopt the new information. The data presentation is important when there is limited time to observe the data, which is the case for example in the Cyber Security Exercises. This evaluation criterion is ease of implementation. Table 4 shows the ease of implementation criterion, which will be taking these matters in consideration with same three tier scoring system.

Table 4 Additional evalution criterion

	Excellent	**Good**	**Poor**
Ease of implementation	The data can be consumed with visual Threat Intelligence Platform (TIP), and it is easy to import to the tool.	The data can be consumed with TIP, but it takes some additional effort to import to the tool.	The data can't be consumed with TIP, the data is hard to use.

The final evaluated aspects are event quality, event timeliness, ease of use, ease of implementation and event scope. The evaluation result can be found on chapter 6.

5 The Environment and Tools For the Case Study

5.1 Case Study Concept

The case study is conducted by collecting the CTI feeds from the service providers and consuming them with Threat Intelligence Platforms Anomali STAXX and MISP. All the feeds are configured to the Anomali STAXX. In MISP there is a collection of default feeds and the PhishTank comes with the default installation. The author was not able to import the other feeds used in the study to the MISP. MISP would have required more knowledge about the tool to be used effectively in this case. All but one feeds are using TAXII 2.0 protocol to the data transfer. AlienVault data is collected with scheduled task and manual import to the Anomali STAXX.

Emotet and Log4j are used as a target threat in measuring the scoring system. Threats are presented in chapter 5.3. The threat data is stored to a Linux machine, where also the feed data is exported from the Anomali STAXX to comparison. Comparison is made on a Linux computer with a shell script using IoC lists against the exported feeds and counting matches. The case study concept is described in Figure 5. Chapter 6.2 describes the results and the sources of the threat indicators in more detail.

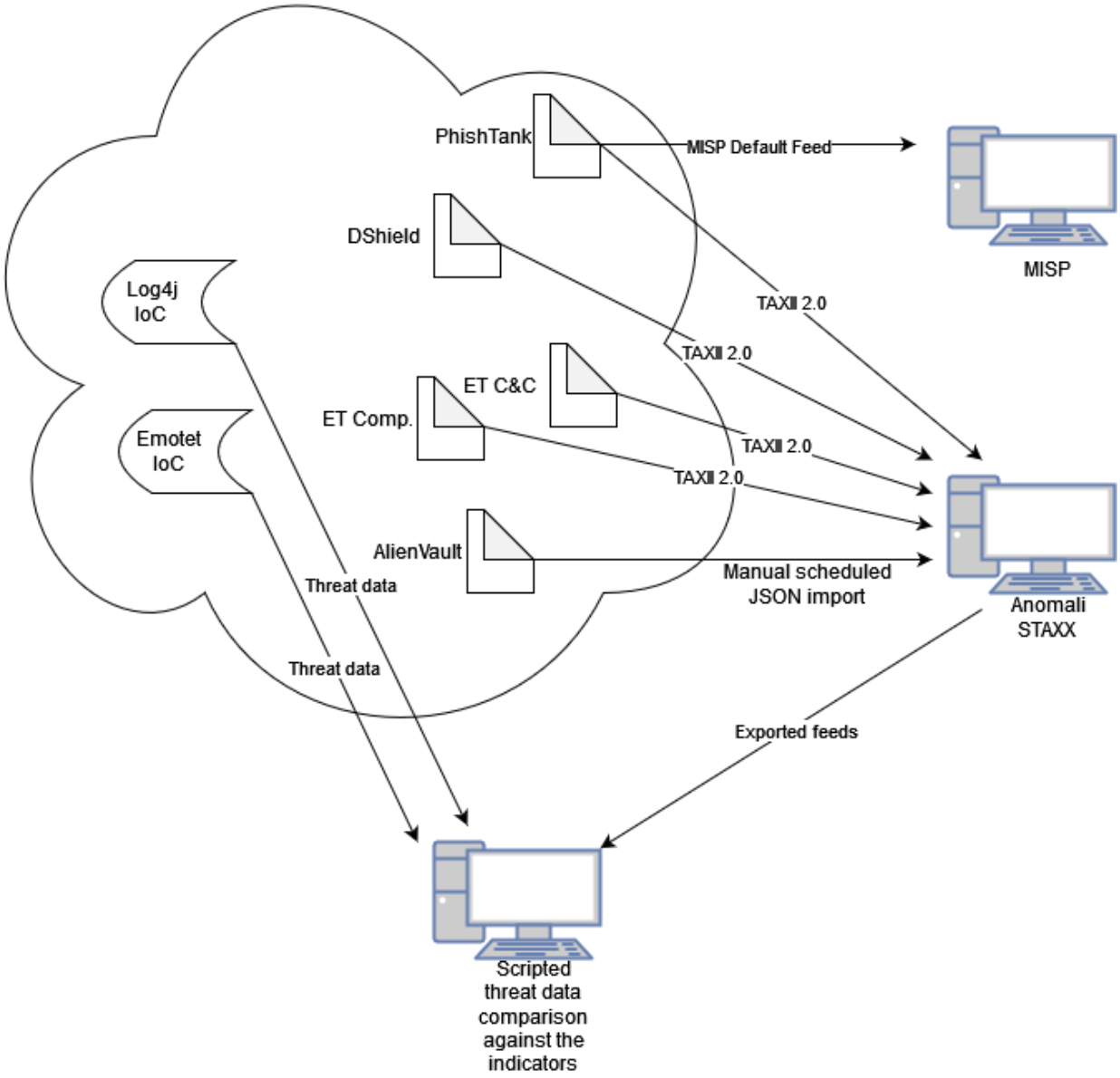

Figure 5 Case Study Concept

5.2 Test Environment

The test environment is operating on the Oracle VM VirtualBox machines on the author's laptop. Anomali STAXX is on a CentOS Linux and MISP is on an Ubuntu Linux. Both machines are installed from virtual images retrieved from the vendors. AlienVault is collected on the same machine where the Anomali STAXX is running. The setup also contains one Linux machine for the comparison of the feeds and threat data.

The data collection periods were:

- Anomali STAXX 26.10.-1.1.2022
- MISP 15.11.-1.1.2022
- AlienVault 12.11-1.1.2022

The Anomali STAXX and MISP feeds were collected once a day automatically by the software, but AlienVault feeds were scheduled to be collected once per week with command-line tool curl on the Anomali STAXX Linux machine.

5.3 Target Threats

Kuusenmäki suggested that when evaluating against his criteria, it could be good to use the same target threats to measurements. Then it would act as a peer review of his research (Kuusenmäki, 2020, 67). The target threats in Kuusenmäki's thesis were Emotet Comeback, APT29 targets COVID-19 vaccine development and WastedLocker. Because the threat landscape is evolving so quickly, this thesis covers only Emotet comeback in means of comparison. The other chosen threat is more current Log4j vulnerability.

5.3.1 Emotet Comeback

Kuusenmäki used Emotet comeback as an example threat, and it is chosen to this thesis in means of comparison. According to Finnish Transport and Communications National Cyber Security Centre (NCSC-FI), the yellow warning has been removed since 18 November 2020 (Traficom, 2020, a), but according to Trendmicro there is indicators that Emotet is might be resurging again (Trendmicro, 2021). At the time of writing the thesis and collecting the data, the Emotet was not active.

Emotet is an info stealer malware that is spread mostly through email attachments. It can be for example a PDF or Microsoft Office document with macros. It can steal any information: passwords, emails, contacts, and other data. Emotet steals the information and sends it to command-and-control servers. Emotet usually uses existing email chains which makes it seem more credible. (Traficom, 2020, b). Emotet is seen first in 2014 as a banking trojan (Trendmicro, 2021).

In this thesis the Emotet IoC's are fetched from Cryptolaemus. The fetched lists are the latest, which are from January 25 2021, and the previous weekend January 22-24 2021. In Kuusenmäki's thesis the used lists are most likely from July 2020, so the indicators have changed since then, and the results are not completely comparable. The fetched list is in the Appendix 1.

5.3.2 Log4j

According to UK's National Cyber Security Center, Log4j is a vulnerability, which was detected in the beginning of December 2021. This vulnerability is in an open source logging library, which is widely used in Apache products, and in many other products. The exploitation of this vulnerability doesn't require particularly great expertise, which makes it very dangerous. UK's National Cyber Security Center estimates this is the most severe vulnerability in years (National Cyber Security Center (UK), 2021).

NCSC-FI informed three days after the vulnerability was found, that there were already existing proof-of-concept on how to exploit the vulnerability. The vulnerability is fixable only by system administrators (Traficom, 2021, c). The IOC's to Log4j are fetched from the Netherlands' Cyber Security Center's NSCS-NL Github and are presented in the Appendix 2.

5.4 Combining the Feeds with TIP Tools

Threat Intelligence Platform (TIP) tools used in this thesis have built-in features to support multiple feeds at the same time. The simplest way to collect the data to the same platform is to choose the tools accordingly.

5.4.1 Anomali STAXX

Anomali STAXX is a free tool to consume STIX/TAXII feeds. With the default installation there is Limo feed available, but it is possible to use the Anomali STAXX to consume any STIX and TAXII feeds. In this thesis I configured the Anomali STAXX to consume TAXII repositories:

- Limo
- Hail A Taxii
- AlientVault, by manual import

Anomali Limo feed already has most of the individual feeds as seen on Table 2, so this is the preferable tool to be used on case study. Anomali STAXX is easy to use and quick to setup.

AlienVault should be compatible with the Anomali STAXX, but for some reason it did not work during the making of this thesis. The feed could be configured, but it didn't populate any data to the dashboard. AlientVault data was imported manually by collecting the data on the Linux machine with curl and importing the data to the Anomali STAXX with import option. Anomali STAXX has an import tool to import any JSON feeds.

Anomali STAXX shows the result of all the feeds in same dashboard. This gives a quick glance to the threat landscape before diving further into the data. Figure 6 shows indicators from the last 30 days and Figure 7 shows an example of the whole dashboard. In the dashboard the admin source is in this case the AlienVault data after one time import on 12 November 2021.

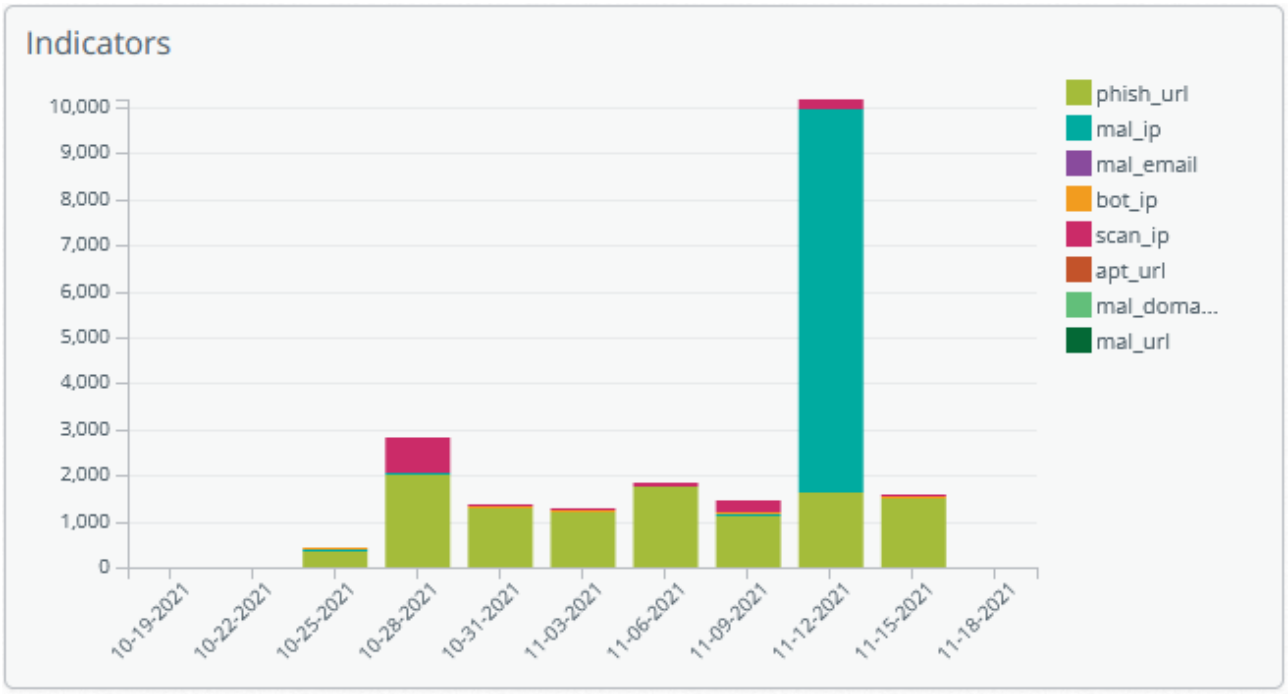

Figure 6 Screenshot of Indicators from Anomali STAXX

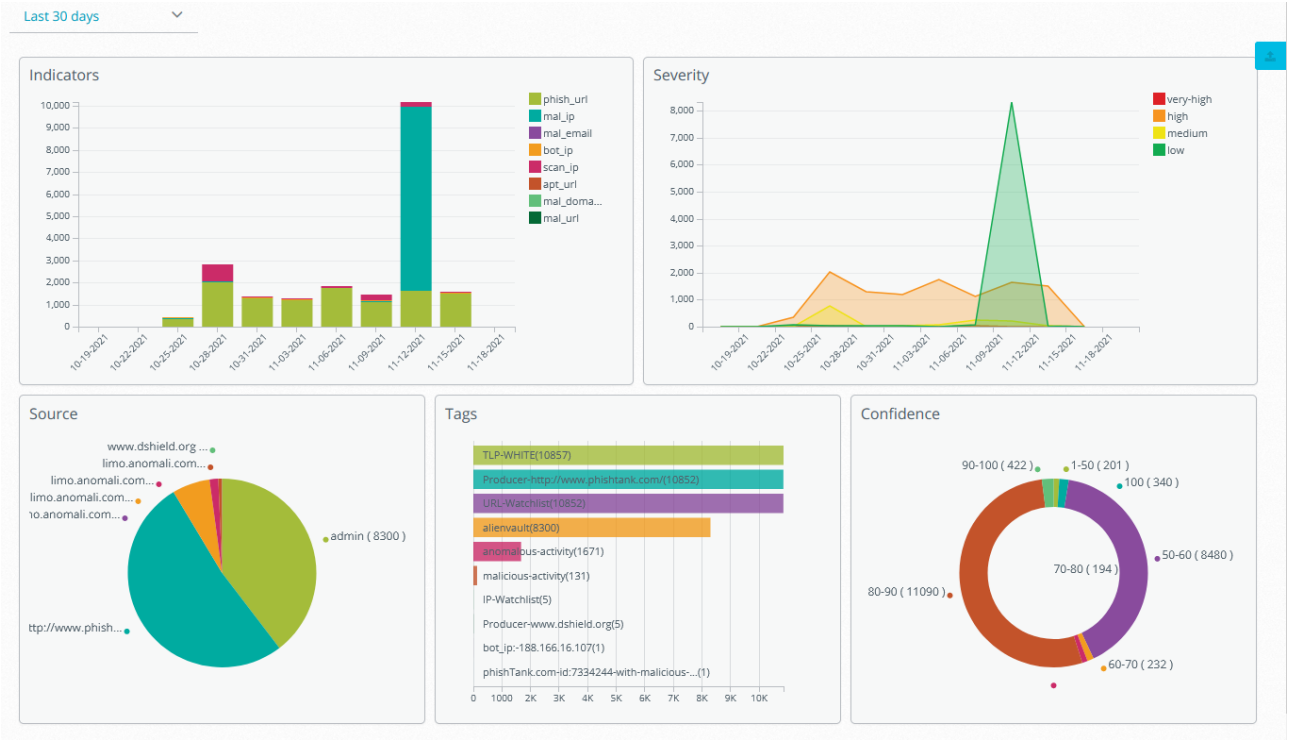

Figure 7 Screenshot of Anomali STAXX dashboard

5.4.2 MISP Platform

MISP is a versatile Open Source Threat Intelligence Platform, which is used over 6000 organizations over the world. MISP can be used to store, share, and correlate CTI, which helps in detecting and preventing threats. MISP data can be used to find relationships between IOCs, and it has built-in sharing feature (MISP).

MISP supports STIX and has API for automation and integration. In this thesis the goal is to have as easy as possible deployment for the education use, so it is possible, that without programming and scripting skills, the importing of the missing feeds may turn out to be too arduous and troublesome.

Figure 8 shows an example of MISP's event view. It is not as visual as the Anomali STAXX's dashboard, but it contains lots of data at one glance. MISP has a completely modifiable dashboard, but it is empty as default.

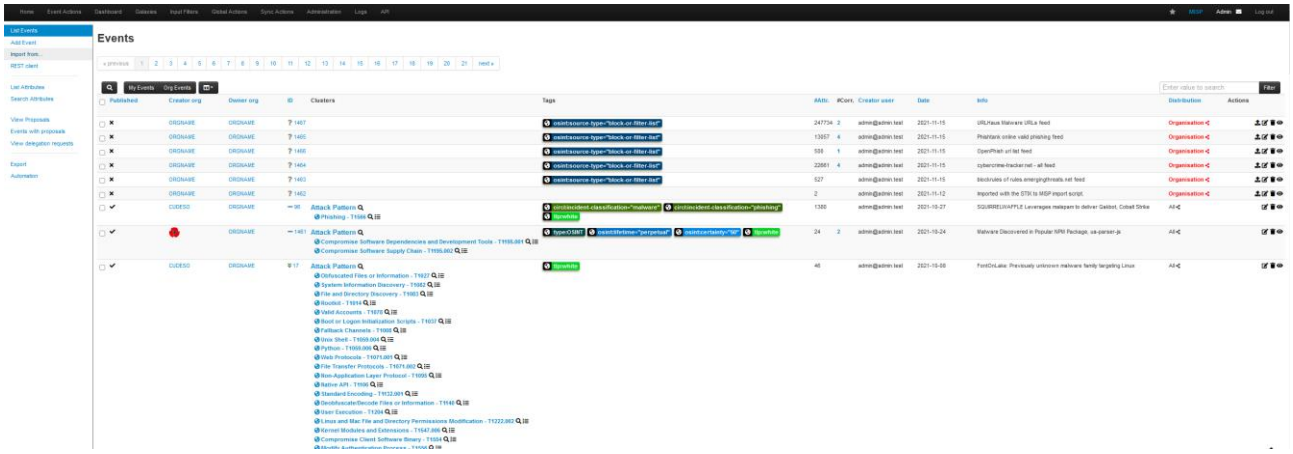

Figure 8 Screenshot of MISP event view

After numerous hours of testing and trying, the author of the thesis was not able to bring the selected feeds to the MISP, so this tool was not used on the case study very much. MISP was much more difficult and not very intuitive tool to use, but it is popular tool, so the problems were most likely due to an inexperienced user.

6 Results

6.1 Feed Review Results

Evaluation criteria, which is used to score the feeds, is presented in the chapter 4.5. The educational use is considered in the evaluation wherever it is relevant. Feeds are presented in the chapter 4.4.

6.1.1 PhishTank

Event quality. PhishTank data is submitted and supervised by humans, which makes the rating result in good.

Event timeliness. Database is updated every hour, so it is rated excellent, because the threshold to excellent is under two hours.

Event scope. PhishTank is a feed that considers only phishing URLs, so the event scope gets a rating poor.

Ease of use and ease of implementation. Ease of use and ease of implementation are both excellent: there is no complex registration process and the Limo feeds that comes with Anomali STAXX includes the PhishTank feed by default. This is easy to implement, and it could be easily used in the education, because it requires little to no expertise to get acquainted with. PhishTank feed is also one of the default feeds in MISP.

Table 5 summarizes the ratings.

Table 5 PhishTank rating

	Excellent	Good	Poor
Event quality		Data is not from trusted source but is filtered by human. Voting system to verify the data. May contain small number of false positives.	
Event timeliness	Database is updated hourly.		
Event scope			Under five different types of IoC's. The feed includes only phishing URLs.

Ease of use	No complex registration. Supports STIX / TAXII 2.0.		
Ease of implementation	Really easy to implement. Comes with Anomali STAXX Limo feed and requires no complex configuration. Can be used with TIP.		

6.1.2 DShield

Event quality. DShield is based on logs from IDS and firewall systems, which are contributed to the service by community. The data is not validated by humans, and that results in poor rating, because it can contain false positives and wrong information. There is no registration required, when sending logs to the service.

Event timeliness. Event timeliness is excellent: the log information is usable when someone sends it to the service.

Event scope. Dshield provides only IP addresses, which makes the event scope rating poor.

Ease of use and ease of implementation. There is no registration process needed to consume the feeds, which makes the feed easy to use and implement. DShield comes with Anomali Limo feeds, and it is easy to use with Anomali STAXX and other TIPs without previous experience, so it can be considered suitable for educational needs. The result is excellent in both criteria.

Table 6 summarizes the ratings.

Table 6 DShield rating

	Excellent	Good	Poor
Event quality			Can contain false positives, not supervised by humans.
Event timeliness	Data is usable under two hours threshold.		
Event scope			Under five different types of IoC's. The feed includes only scanning IP's.
Ease of use	No complex registration. Supports STIX / TAXII 2.0.		
Ease of implementation	Really easy to implement. Comes with Anomali STAXX Limo feed and requires no complex configuration. Can be used with TIP.		

6.1.3 EmergingThreats C&C Server

Event quality. EmergingThreats C&C provides data from known command-and-control servers. The feed is supervised by humans, and it comes from reliable sources, such as Abuse.ch. Event quality get a grade excellent.

Event timeliness. The feeds are updated daily, which gives this feed good grade.

Event scope. Scope is narrow because the feed contains only the IP addresses of the command-and-control servers. This results in poor grade.

Ease of use and ease of implementation. Ease of use and ease of implementation are both excellent – there is no complex registration process, and the feed comes with Anomali Limo and is easy to use with Anomali STAXX. This feed is suitable for educational needs: it requires no previous knowledge of the CTI feeds to get acquainted with.

Table 7 summarizes the ratings.

Table 7 EmergingThreats C&C Server rating

	Excellent	Good	Poor
Event quality	Feeds are reviewed by humans and the data is from the trusted source.		
Event timeliness		Events are published once per day.	
Event scope			Under five different types of IoC's. The feed includes only IP's of the C&C servers.
Ease of use	No complex registration. Supports STIX / TAXII 2.0.		
Ease of implementation	Really easy to implement. Comes with Anomali STAXX Limo feed and requires no complex configuration. Can be used with TIP.		

6.1.4 EmergingThreats – Compromised

Event quality. EmergingThreats – Compromised feed data is not verified by human entity, so the event quality is good. The data is from reliable sources.

Event timeliness. Events are published once per day, which makes the event timeliness is rated good.

Event scope. Event scope results in poor because the feed contains only compromised IP addresses.

Ease of use and ease of implementation. Ease of use and ease of implementation are both excellent: no complex registration process and the feed comes with Anomali Limo. Easy to use with TIP. Like other Anomali LIMO feeds, this is easy to adapt, and it is good feed to get one familiarized with CTI feeds.

Table 8 summarizes the ratings.

Table 8 EmergingThreats - Compromised rating

	Excellent	Good	Poor
Event quality		Data is not supervised by humans and can contain small number of false positives. Data is from the reliable sources.	
Event timeliness		Events are published once per day.	
Event scope			Under five different types of IoC's. The feed includes only compromised IP addresses.
Ease of use	No complex registration. Supports STIX / TAXII 2.0.		
Ease of implementation	Really easy to implement. Comes with Anomali STAXX Limo feed and requires no		

	complex configuration. Can be used with TIP.		

6.1.5 AlienVault

Event quality. AlienVault is community based, but the data is supervised and sanitized by Alien Lab before it is published. The worldwide community of security professionals and over 100,000 participants provide huge amount of vulnerability data in daily basis (AT&T Cybersecurity, 2021). The feed quality is rated good because the data is from non-trusted sources but is filtered or supervised by human.

Event timeliness. Event timeliness is excellent. The events are published under two hours threshold.

Event scope. The pulses include different types of IoC's, so the event scope rate is excellent.

Ease of use. There is registration process needed to order OTX pulses, but it is not complex. This would get a good rate in the ease of use if the feed was STIX/TAXII 2.0 compliant, but because it is only STIX/TAXII 1.0 compatible, the final rating is poor.

Ease of implementation. The ease of implementation is good: the feed data can be presented in TIP, but at least at this moment it requires additional work. Anomali STAXX can present the data, but the data needs to be collected first in other means and be imported to the Anomali STAXX as JSON. In this thesis the data is collected with Linux machine with curl and imported to the Anomali STAXX manually. The feed is still usable in education but requires some additional expertise to collect the data. JSON is human readable, so the data itself is easy to understand.

Table 9 summarizes the AlienVault ratings.

Table 9 AlienVault rating

	Excellent	Good	Poor
Event quality		Data is not from trusted source but is filtered by human. May contain small number of false positives.	
Event timeliness	Data is usable under two hours threshold.		
Event scope	Ten or more different type of IOCs. Can be used in incident management.		
Ease of use			Feed is not STIX / TAXII 2.0 or newer compliant. Registration process is needed, but it is not complex.
Ease of implementation		The data can be consumed with TPI, but it takes some additional	

		effort to import to the tool.	

6.1.6 Summary Of the Evalution

Table 10 shows the summary of the grading of the feeds. None of the feeds graded excellent in every criterion. This gives the impression that when using CTI feeds, it is justifiable to use more than one feed to get as good as possible overview of the threats and increase the situational awareness. In business use, the event quality, timeliness and scope are more important than the ease of use and implementation, but in educational needs these might be more important aspects. When using TIP like Anomali STAXX and its Limo feed, the implementation is almost automatic, and user gets to use quality feeds with very little effort.

Table 10 Summary of the evaluation

Feed	Event quality	Event timeliness	Event scope	Ease of use	Ease of implementation
PhishTank	Good	Excellent	Poor	Excellent	Excellent
DShield	Poor	Excellent	Poor	Excellent	Excellent
EmergingThreats C&C	Excellent	Good	Poor	Excellent	Excellent
EmergingThreats - Compromised	Good	Good	Poor	Excellent	Excellent
AlienVault	Good	Excellent	Excellent	Poor	Good

6.1.7 Comparing the Results

Comparing the result to Kuusenmäki's results on those feeds, that were same in both theses. In this thesis the cost is not relevant, because only free of charge feeds are chosen, and therefore the cost criteria are not considered. Also, the ease of implementation is not present in Kuusenmäki's thesis, but is important in educational use, so that is also exception in this comparison.

All the results correlate in both theses. Kuusenmäki's criteria was easy to understand and use. The evaluation of the chosen feeds was simple and straightforward. This resulted in exactly same results in both theses. This indicates that the criteria and research design are repeatable.

Table 11 shows the results of feed review comparison between Kuusenmäki's and this thesis. If results were correlated in both theses, it is marked with X.

Table 11 Comparing results with Kuusenmäki's thesis

	Event quality	Event timeliness	Event scope	Ease of use	Ease of implementation	Cost
PhishTank	X	X	X	X	N/A	N/A
DShield	X	X	X	X	N/A	N/A
Emerging Threats C&C server	X	X	X	X	N/A	N/A
Emerging Threats - Compromised	X	X	X	X	N/A	N/A
AlientVault	X	X	X	X	N/A	N/A

6.2 Case Study Results

In the case study there were two example threats: one was mutual with Kuusenmäki's thesis, and the other was more recent big vulnerability. The mutual threat was the Emotet comeback and the recent was the Log4j vulnerability that was found in December 2021.

Emotet indicators were retrieved from Cryptolaemus' Pastedump (Cryptolaemus, 2021). Cryptolaemus is a group of cyber security researchers and administrators who are actively fought against the Emotet. (Cimpanu, 2020) The latest dump from January 25, 2021, were used in this thesis. Appendix 1 shows all the 17041 retrieved IoC's of Emotet.

Log4j indicators were retrieved from Netherlands's National Cyber Security Center's Github. (NCSC-NL, 2020). There were lots of different sites collected under the Github – too many to choose them all. Appendix 2 shows all the 31514 retrieved indicators and their sources.

Table 12 shows the summary of the Log4j and Emotet IoC's compared to the total IoC count of the feed. The IoC comparison were made with Linux grep command. TIP tools did not provide tools to do this. More information about the data comparison is found on the chapter 6.3.

Table 12 Threat IoC count in the feeds

Feed / IoC count	Total IoC count of the feed	Emotet IoC count	Log4j IoC count
PhishTank	20000	22	3
DShield	3856	0	70
Emerging Threats C&C server	354	0	235
Emerging Threats – Compromised	738	0	41
AlientVault	22195	0	587

The PhishTank was the only feed that contained data from both threats. Emotet was not active on the data collection period, which explains this. Log4j was active and this shows in the results.

In the case study the evaluated sections are event scope, event timeliness, ease of use and ease of implementation.

6.2.1 PhishTank

PhishTank was the only feed that contained data from both Emotet and Log4j threats.

Event Scope

On the data collection period, the PhishTank had 20,000 indicators of compromise. These IoC's were all phishing URL's, which makes the scope narrow. From these 20,000 indicators, there were 22 Emotet indicators and 3 Log4j indicators. Event scope aligns with the feed review results.

Event Timeliness

All the Emotet findings were from the Cryptolaemus's domain bucket list, which was last updated on January 25, 2021. On the domain bucket list, there was no information when the individual domain was listed, so the timeliness could not be resolved for the Emotet indicators.

The Log4j findings were collected from Netherland's National Cyber Security Center's Github. Three matches were found, and the timeliness was excellent on one and poor on two. One match was on PhishTank six days before on the Github source. Two other findings were on the PhishTank two days later than on the Github source. The event timeliness did not entirely match the feed review results.

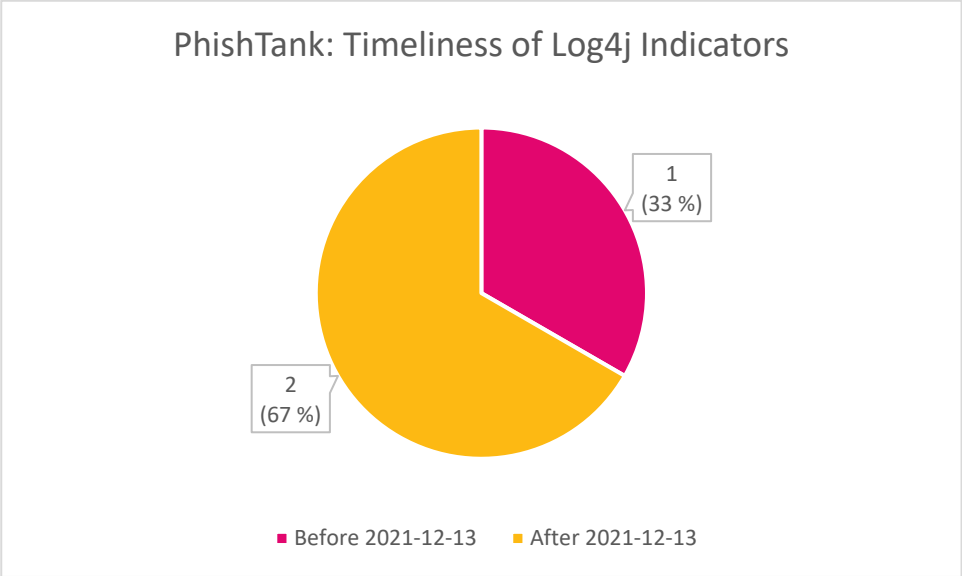

Figure 9 PhishTank Log4j timeliness

Ease of Use

No complex registration needed, and PhishTank supports STIX/TAXII. PhishTank has API and can be used with any TAXII client. Ease of use aligns with feed review results.

Ease of Implementation

PhishTank comes with Anomali Limo and MISP, and the implementation is easy. Both TIP's are easy and quick to setup. Suitable for educational use for the ease of implementation point of view. Ease of implementation aligns with the feed review results.

6.2.2 DShield

DShield feed contained only Log4j vulnerability indicators.

Event Scope

On the data collection period, there were 3856 indicators of compromise. All the indicators were IP addresses, which makes the feed rating poor and correlates with the feed review results. From the 3856 indicators 70 were Log4j indicators.

Event Timeliness

From the Log4j indicators retrieved from the NCSC-NL's Github, there were variance in the timeliness. DShield data should be usable in under two hours threshold, but in this setup this didn't hold true. There were delays for days or even weeks in most of the indicators, and it is unclear was this because they were published late to the feed or is this due to the Anomali STAXX tool. At least with this setup the result did not align with the feed review results. With this setup the rating results in poor.

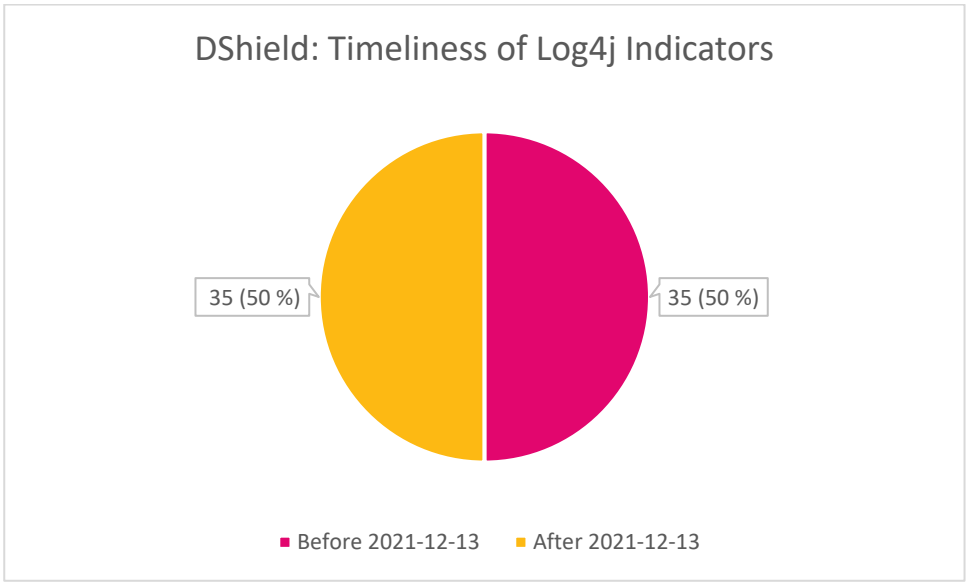

Figure 10 DShield Log4j timeliness

Ease of Use

No registration process needed, and the feed is STIX/TAXII compliant. Ease of use aligns with the feed review results.

Ease of Implementation

DShield comes with Anomali Limo and the implementation is easy. DShield was easy to use with TIP and is suitable for educational use if the ease of implementation is important factor. This result aligns with the feed review results.

6.2.3 EmergingThreats C&C Server

EmergingThreats C&C Server feed contained only Log4j vulnerability indicators.

Event Scope

On the data collection period the EmergingThreats C&C Server contained 354 indicators, from which the Log4j indicators were 235. All the indicators were IP addresses and this correlates with the feed review result, which is poor.

Event Timeliness

EmergingThreats C&C Server timeliness were better than the feed review rating implies. The indicators retrieved from NCSC-NL were almost every time on the feed before the indicators were published. Some were even months earlier. This might suggest that the compromised IPs were already suspicious or used in other malicious activities. Indicators are published once per day to the feed and the Log4j indicators were usable in time or as said much earlier. This result did not correlate with the feed review with all the indicators, but the result was better than expected. 94 % of the indicators were published in time, that equals excellent rate. Because the rate was over 90 %, it could be safe to rate this as excellent.

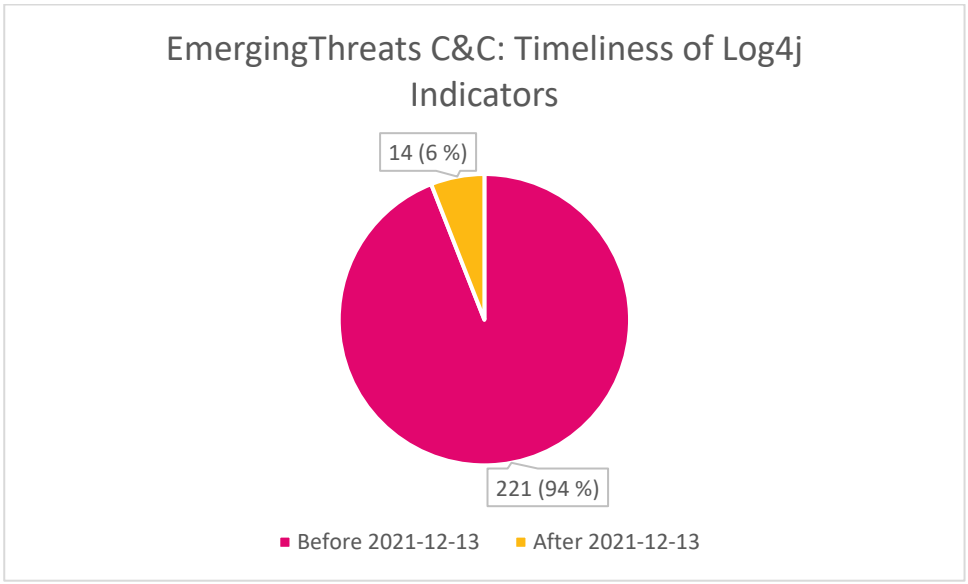

Figure 11 EmergingThreats C&C Log4j timeliness

Ease of Use

No registration process needed, and the feed is STIX/TAXII compliant. Ease of use aligns with the feed review results. The result aligns with the feed review, the rating is poor in both.

Ease of Implementation

Feed comes with Anomali LIMO and is easy to setup. The feed is usable with TIP and setup is easy, so this is also ideal feed to use in education. This results in excellent, which aligns with the feed review result.

6.2.4 EmergingThreats – Compromised

EmergingThreats – Compromised feed contained only Log4j vulnerability indicators.

Event Scope

On the data collection period the feed contained 738 indicators of compromise, and 40 of these were Log4j indicators. Feed contains only IPs, and this aligns with the feed review results. Under five different indicators results in poor grade.

Event Timeliness

With Log4j indicators, the timeliness varied from early to late publish. Some IOCs were on the feed weeks earlier, some on time and some weeks later than the NCSC-NL publish. 20 % of the indicators were published before or on time, and 80 % after, so it could be safe to evaluate this as poor grade.

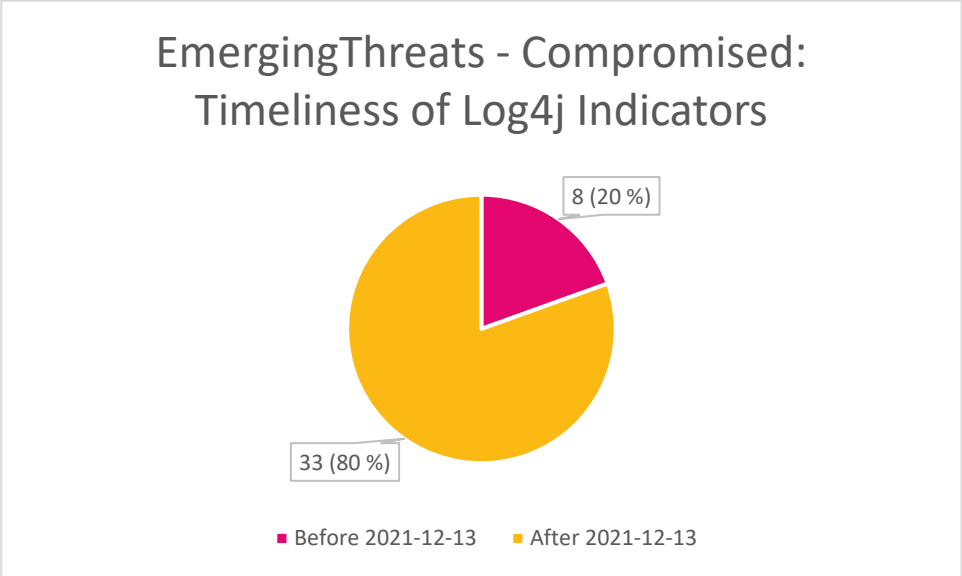

Figure 12 EmergingThreats - Compromised Log4j timeliness

Ease of Use

No registration process needed, and the feed is STIX/TAXII compliant. Ease of use aligns with the feed review results.

Ease of Implementation

This feed is also easy to implement and suitable for educational needs. The feed is usable with TIP, and it comes with Anomali LIMO. This result aligns with the feed review result.

6.2.5 AlienVault

AlienVault OTX feed contained only Log4j vulnerability indicators.

Event Scope

On the data collection period, the AlienVault contained total of 22195 indicators, from which 587 indicators were Log4j indicators. AlienVault should include over 10 different types of IoCs, but on this case, there were only three different types of IoCs: malicious IP addresses, geolocation URLs and malicious domains. This results in poor grade.

Event Timeliness

Most of the Log4j data was on the feed before the IoCs were published by NCSC-NL. Like in EmergingThreats C&C Server feed, this might suggest that the compromised IPs were already suspicious or used in other malicious activities. There were 519 indicators published in AlienVault before they were published by CERT AGID. 68 indicators were published later.

Timeliness result is difficult to evaluate accurately because there is dispersion in the results, but the majority of the indicators were published within the excellent grade.

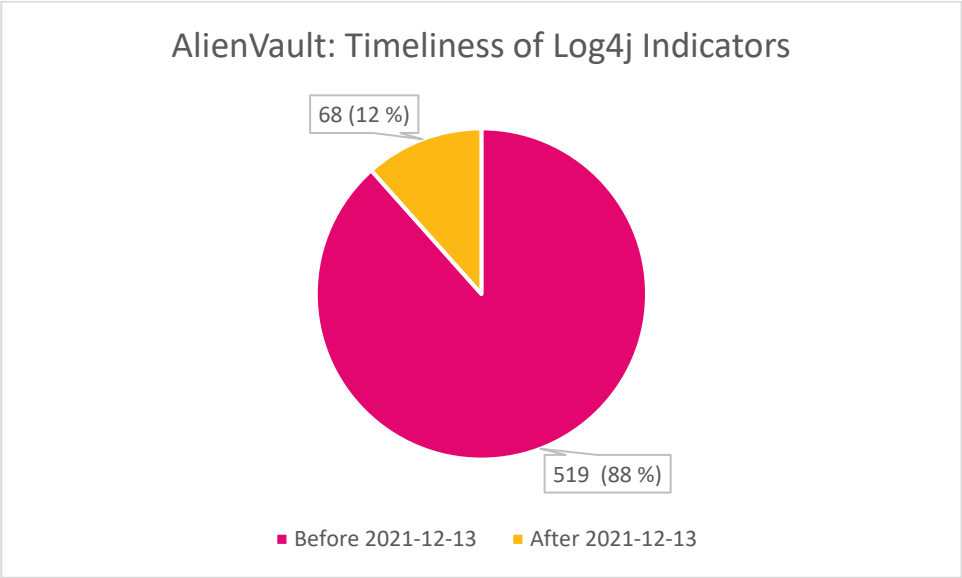

Figure 13 AlienVault Log4j timeliness

Ease of Use

AlienVault is not STIX / TAXII 2.0 compliant. Registration process is needed, but it is fast and straightforward. This results in poor grade and is aligned with the feed review results.

Ease of Implementation

AlienVault requires some additional effort to be consumed with TIP, but it is possible. This results in good grade, and this aligns with the feed review results. AlienVault pulse selection can be a little hard, there is so many pulses from which to choose. This is not necessarily the best tool to use in educational use at least in the beginner's level. This depends on the needs: if there is a need to a

big variance of different kind of data, AlienVault is a good choice, but this can also be a downside if there is not lots of time to get one familiarized with the different feeds.

6.2.6 Case Study Summary

Most of the feed criterion were matching in the feed review and case study. Event timeliness did not match on most the feeds or was difficult to evaluate. Timeliness had lots of variances in almost every feed, and with this data and tools, it is not possible to undoubtedly tell if the timeliness is accurate or not.

Almost every feed had a poor grade from the event scope. AlienVault was the only exception with this criterion. This might be because there was only one current and active threat chosen to comparison. With this data it is not possible to undoubtedly tell if this result is accurate. Good thing is, that there were no feeds with no matches – every feed included indicators from Log4j vulnerability.

Table 13 shows the summary of the case study results and Table 14 compares the results in the case study and the feed review. The matching results are marked as “yes” and where there was no match or just partly match, the result is marked as “no”.

Table 13 Case Study Summary

Feed	Event timeliness	Event scope	Ease of use	Ease of implementation
PhishTank	Poor	Poor	Excellent	Excellent
DShield	Poor	Poor	Excellent	Excellent
EmergingThreats C&C	Excellent	Poor	Excellent	Excellent
EmergingThreats - Compromised	Poor	Poor	Excellent	Excellent
AlienVault	Excellent	Poor	Poor	Good

Table 14 Comparison with the feed review and the case study results

Feed	Event timeliness	Event Scope	Ease of Use	Ease of Implementation
PhishTank	No	Yes	Yes	Yes
DShield	No	Yes	Yes	Yes
EmergingThreats C&C	No	Yes	Yes	Yes
EmergingThreats – Compromised	No	Yes	Yes	Yes
AlienVault	Yes	No	Yes	Yes

6.3 Combining and Comparing the Feeds

The case study was executed by collecting the data from all the feeds and then comparing the known IoC's of the threats with the collected data. The TIP tools did not work well with comparing multiple IoC's with the feed data. Anomali STAXX does not support a bulk import of IoC's, so all the indicators would be impossible to check one by one. MISP supports bulk import, but at least for the beginners, the tool is hard to use, and it does not contain all the evaluated feeds. Like Anomali STAXX, one indicator at a time it is easy to find the matches between indicators and the feed data, but at least this thesis author did not manage to do the bulk comparison even though the bulk import of the indicators were done.

In educational use the MISP seems difficult and arduous to learn, so if the goal is to learn to use the feeds and not the tool itself, it might be too time consuming to use in the beginner's level. If

there is time to learn to use the tool first, I believe MISP would be more versatile than Anomali STAXX, which is simple and hasn't got many features.

The feed data itself is in JSON format, so with data analytics and programming skills there would be more means to do more effective data merge with the raw data from the feeds. With the TIP tools the data can be presented in one place, but the bulk data processing can be laborious.

7 Conclusion

The purpose of this study was to find out what STIX/TAXII feeds there are, how the feeds can be consumed with a TIP and which of the feeds could be suitable for Cyber Security exercises and other educational uses. It turned out, that there aren't many feeds that are both active and STIX/TAXII compatible. All the selected feeds were usable with TIP and all the feeds could bring some insight in Cyber Security education.

Using the feeds does not require much expertise, and at least one of the used tools was easy to use and implement: Anomali STAXX is effortless to get familiar with and if there is not much time to use in setup and implementation, that would be recommendable tool to use. On the other hand, MISP is widely used and has lots of features, but it requires a lot more expertise and is more complicated to setup and use. The author of this thesis was not able to bring to all the selected feeds to MISP, but it has lots of good selection of default feeds. These feeds were unfortunately outside the scope of this thesis, because most of them were not STIX/TAXII feeds. In advanced educational use, the MISP is good tool to learn, but if the purpose is to quickly inspect the feeds and what kind of information they contain, the Anomali STAXX could be the more reasonable tool to use.

Other purpose of this thesis was to evaluate Juha Kuusenmäki's criteria, which he produced in his thesis *Evaluation of Threat Information Feeds for a Cyber Defence Center*. In addition to Kuusenmäki's criteria, there were new criteria introduced to evaluate the ease of implementation. Also, the evaluation was made bearing in mind the educational use of the feeds. Kuusenmäki suggested that the possible peer review could be made with the same feeds and same threats, but the threat landscape is evolving so quickly that 1.5 years after the Kuusenmäki's thesis, most of the feeds didn't contain data from the example threat "Emotet Comeback". Kuusenmäki's criteria

were easy to use, and the feed review results were the same in both theses. Case study results varied, but as mentioned, the threat landscape is changing quickly and the results with different example threats would not be comparable.

Research progressed otherwise without major problems, but some technical issues were encountered. AlienVault should be compatible with Anomali STAXX, but at least with this version and this data collection period, the data didn't populate automatically. Also, MISP turned out to be too difficult to use without any previous knowledge.

7.1 Answers to the Research Questions

Four of five research questions got answered. In hindsight, there were too many research questions, and some of them got dropped out during the research. It was anticipated that it is possible some questions could be left unanswered.

The first research question was "What CTI feeds there are already?" This got answered, but only inside the scope of this thesis: only STIX/TAXII feeds were covered. This thesis did not cover all the STIX/TAXII feeds, but the most active and popular ones were presented and evaluated.

The second question was "How can the feeds be consumed and used?" This was also a question that got answered. All the feeds can be consumed with TIP, or just simply pulling the JSON data to a computer. JSON is human readable, but the data is hard to interpret without any tools.

The third research question was "Which of these feeds are suitable for the Cyber Security education use?" All the feeds were suitable for educational use, although the amount of effort to get the feed presented in TIP varies. With Anomali STAXX the implementation is almost automatic, and with pre-set LIMO feed user gets to use quality feeds with very little previous knowledge.

The fourth question, "How is it possible to combine these feeds to a single feed?", did not get answered in a satisfactory level. It is possible to combine the feeds with TIP tools, but it still requires the feeds to be inspected one by one to find the possible matches with threat indicators. TIP tools

give a good perception for the feeds but is not actually combining the feed data. To properly combine the feed data, data-analysis and, possibly, coding skills could be helpful. This could be something to study further.

The fifth and the final research question was “Are Kuusenmäki’s criteria ausable and valid?” Kuusenmäki’s criteria was easy to use, and the results seems to be valid with the chosen feeds.

7.2 The Case Study Results

The case study results were as anticipated: Emotet was not active during the data collection period, so there were little to none hits in the feeds with the Emotet IOC’s. Log4j vulnerability was recently surfaced during the data collection period, so it was expected to be found on all the feeds. Feed review results were also as expected, because there was fairly short time between Kuusenmäki’s thesis and this one. It could have been good to have some different feeds for evaluation, but there were not any reasonable STIX/TAXII feeds that were not already evaluated in Kuusenmäki’s work. The evaluation was still a bit different, because the ease of implementation was a new step in the evaluation in this thesis.

7.3 Further Research

A further research suggestion would be the merging of the feed data to a single JSON feed. That was out of the scope of this thesis. Other further research suggestion would be to find out and evaluate other tools to be used in consuming the CTI feeds.

Also, it would be interesting to evaluate more STIX/TAXII feeds with the same criteria. This would require new active STIX/TAXII feeds to be available. During the writing of this thesis, the usable feeds were all the same than with the Kuusenmäki’s thesis, so the results were very similar. Maybe in few years there will be more reasonable feeds to evaluate.

References

Anomali (2021, November 12). *Anomali Limo.* https://www.anomali.com/resources/limo

AT&T Cybersecurity (2021, November 16). *AT&T Alien Labs Open Threat Exchange.* https://cybersecurity.att.com/open-threat-exchange

Bouwman, X., Griffioen, H., Egbers, J., Doerr, C., Klievink, B., Eeten, M. (2020). *A different cup of TI? The added value of commercial threat intelligence.* Conference on 29th USENIX Security Symposium (pp. 433-450). USENIX Association. https://www.usenix.org/system/files/sec20-bouwman.pdf

Cimpanu, C. (2022, January 3). *Meet the white-hat group fighting Emotet, the world's most dangerous malware.* ZDNet.com. https://www.zdnet.com/article/meet-the-white-hat-group-fighting-emotet-the-worlds-most-dangerous-malware/

CISA. (2021, November 12) *Traffic light protocol (TLP) definitions and usage.* https://www.cisa.gov/tlp

Crytolaemus (2022, January 3). *Cryptolaemus Pastedump.* https://paste.cryptolaemus.com/

Dalziel, H. (2015). *How to define and build an effective Cyber Threat Intelligence capability.* Syngress. https://janet.finna.fi/Record/jamk.993620817606251

DShield (2021, November 16). *ISC History and Overview.* https://www.dshield.org/about.html

EmergingThreats (2021, November 16). *Emerging Threats FAQ.* https://doc.emergingthreats.net/bin/view/Main/EmergingFAQ.

ENISA (2021, November 4). *ENISA Threat Landscape 2020 - Research topics.* https://www.enisa.europa.eu/publications/enisa-threat-landscape-2020-research-topics

Ginn, Jane (2021, November 10). *STIX 2.1 CS01: Specification overview. CTI TC Webinar Series 6 April 2020.* https://www.youtube.com/watch?v=SCWLssIKGwg&t=1s

Griffionen, H., Boiij T., Doerr C. (2020). *Quality Evaluation of Cyber Threat Intelligence Feeds.* ACNS '20, October 19-22, 2020, Rome, Italy. https://www.researchgate.net/profile/Christian-Doerr/publication/343954119_Quality_Evaluation_of_Cyber_Threat_Intelligence_Feeds/links/5f72f35192851c14bc9d0393/Quality-Evaluation-of-Cyber-Threat-Intelligence-Feeds.pdf.

Kananen, J. (2017). *Laadullinen tutkimus pro graduna ja opinnäytetyönä.* Jyväskylän ammattikorkeakoulu. https://janet.finna.fi/Record/jamk.993276444806251

Korte, K. (2021). *Measuring the quality of Open Source Cyber Threat Intelligence Feeds.* [Master's thesis, JAMK University of Applied Sciences]. https://urn.fi/URN:NBN:fi:amk-202105178967

Kuusenmäki, J. (2020). *Evaluation of Threat Information Feeds for a Cyber Defence Center. A Case Study for Company Netcloud AG.* [Master's thesis, JAMK University of Applied Sciences]. https://urn.fi/URN:NBN:fi:amk-2020121528323

Lauf, F., Kuziemsky, C. (2017). *Handbook of eHealth Evaluation: An Evidence-based Approach.* University of Victoria. https://www.ncbi.nlm.nih.gov/books/NBK481583/

Meier, R., Scherrer, C., Gugelmann, D., Lenders, V., Vanbever, L. (2018). *FeedRank: A tamper-resistant method for the ranking of cyber threat intelligence feeds.* 2018 10th International Conference on Cyber Conflict (CyCon) (pp. 321-344). IEEE. https://ieeexplore.ieee.org/document/8405024.

MISP (2021, November 17). *Features of MISP, the open source threat sharing platform.* https://www.misp-project.org/features.html.

MITRE (2021, November 2). *Situation awareness.* https://www.mitre.org/capabilities/cybersecurity/situation-awareness

National Cyber Security Center (UK) (2022, January 3). *Log4j vulnerability - what everyone needs to know.* https://www.ncsc.gov.uk/information/log4j-vulnerability-what-everyone-needs-to-know

NCSC-NL (2022, January 3). *Log4j overview IoCs.* https://github.com/NCSC-NL/log4shell/blob/main/iocs/README.md

NIST. (2016). *NIST Special Publication 800-150.* https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-150.pdf

OASIS (2021. November 10). *CTI Documentation.* https://oasis-open.github.io/cti-documentation/

PhishTank (2021, November 16). *FAQ.* https://phishtank.org/faq.php

Snyder, H. (2019). *Literature review as a research methodology: An overview and guidelines.* Journal of business research, 104 (pp. 333-339). https://www.sciencedirect.com/science/article/pii/S0148296319304564

Traficom (2022, January 3 a) *Yellow warning concerning the Emotet malware strain no longer in force.* https://www.kyberturvallisuuskeskus.fi/en/news/yellow-warning-concerning-emotet-malware-strain-no-longer-force

Traficom (2022, January 3 b) *Emotet malware actively spread in Finland.* https://www.kyberturvallisuuskeskus.fi/en/emotet-malware-actively-spread-finland

Traficom (2022, January 3 c) *Severe alert issued on an actively exploited Log4j vulnerability*. https://www.kyberturvallisuuskeskus.fi/en/varo_ttn_5/2021

Trendmicro (2022, January 3). *Malware Awareness - EMOTET resurges with new detections.* https://success.trendmicro.com/solution/1118391-malware-awareness-emotet-resurgence

Appendices

Appendix 1. Emotet IoC's From Cryptolaemus

https://github.com/katiwei/thesis/blob/main/APPENDIX%201%20Emotet%20IOCs%20from%20Cryptolaemus.xlsx

Appendix 2. Log4j Ioc's From NCSC-NL

https://github.com/katiwei/thesis/blob/main/APPENDIX%202%20Log4j%20IOCs%20from%20NCSC-NL.xlsx

Appendix 3. PhishTank Feed Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%203%20PhishTank%20Feed%20Data.xlsx

Appendix 4. DShield Feed Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%204%20DShield%20Feed%20Data.xlsx

Appendix 5. Emerging Threats C&C Server Feed Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%205%20EmergingThreats%20C%26C%20Server%20Feed%20Data.xlsx

Appendix 6. Emerging Threats – Compromised Feed Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%206%20EmergingThreats%20-%20Compromised%20Feed%20Data.xlsx

Appendix 7. AlienVault Feed Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%207%20AlienVault%20Feed%20Data.xlsx

Appendix 8. Timeliness Results Between Feed Data and IOC Data

https://github.com/katiwei/thesis/blob/main/APPENDIX%208%20Timeliness%20Results%20Between%20Feed%20Data%20and%20IOC%20Data.xlsx